DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.
**REPORT DOCUMENTATION PAGE**

1. **REPORT SECURITY CLASSIFICATION**: UNCLASSIFIED/UNLIMITED
2. **SECURITY CLASSIFICATION AUTHORITY**: AFOSR
3. **DECLASSIFICATION/RECLASSIFICATION INFORMATION**: Approved for public release; distribution unlimited.
4. **PERFORMING ORGANIZATION REPORT NUMBER(S)**: N/A
5. **MONITORING ORGANIZATION REPORT NUMBER(S)**: N/A
6. **NAME OF PERFORMING ORGANIZATION**: University of Pennsylvania
7. **ADDRESS (City, State, and ZIP Code)**: Office of Research Administration, Philadelphia, PA 19104
8. **NAME OF FUNDING/SPONSORING ORGANIZATION**: AFOSR - Life Science Directorate
9. **ADDRESS (City, State, and ZIP Code)**: Bolling AFB, DC 20332
10. **PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER**: AFOSR 91-0082
11. **TITLE (Include Security Classification)**: Dimensional Signal Coding in the Visual System
12. **PERSONAL AUTHOR(S)**: Buchsbaum, Gershon
13. **TYPE OF REPORT**: Final
14. **DATE OF REPORT (Yr, Mo., Day)**: 95/3/17
15. **PAGE COUNT**: 19
16. **SUPPLEMENTARY NOTATION**: N/A
17. **COSATI CODES**
18. **SUBJECT TERMS**: N/A
19. **ABSTRACT**: N/A
20. **DISTRIBUTION/AVAILABILITY OF ABSTRACT**: UNCLASSIFIED/UNLIMITED
21. **ABSTRACT SECURITY CLASSIFICATION**: N/A
22. **NAME OF RESPONSIBLE INDIVIDUAL**: Dr. John F. Tangney
23. **TELEPHONE NUMBER**: 908-737-5041
24. **OFFICE SYMBOL**: N/A
The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements either expressed or implied, of the Air Force Office of Scientific Research or the US Government.
MULTIDIMENSIONAL SIGNAL CODING IN THE VISUAL SYSTEM

The purpose of the research was to investigate the significance of natural images in understanding early image coding in the visual system. The main objectives were:

1. To study the image processing capabilities in early vision and how they modify the space, time and color properties of an image and the efficiency of these image processing operations.

2. To identify key attributes of natural image signals which are sensitive to modification and filtering in early vision and how these attributes are transformed for encoding beyond early vision.

The research addressed a number of issues in early vision including the relationship between color, spatial and temporal properties of images and concentrated in three main thrusts.

1. The match between neural pathways in early vision and their underlying retinal architecture and spatial and temporal properties of images was investigated. Visual tracking was incorporated into the study as contributing important modifications to the incoming spectrum of the image. The results indicate that the combination of visual tracking together with specialized neural pathways in the retina makes the coding of the spatial and temporal features of images efficient. Bio-encoding of images in the early visual system is optimal in the sense that the retinal neural pathways are tuned to intrinsic properties of natural image sequences.

2. The multilayered retinal architecture and how it mediates signal propagation and prevents distortions as the signal propagates through retinal cell layers was investigated. The research incorporated anatomical and functional details of retinal architecture including different cell densities in different retinal cell layers, cell-to-cell variations, and how cell arrays sample and propagate the image. These properties were incorporated into a multi-stage signal processing model. Understanding of the hierarchical multi-layered signal processing strategy of the retina revealed the role of various components of retinal anatomical architecture. Together with the optics of the eye, retinal architecture provides a means to preserve the image and prevent distortions in it.

3. Color constancy, or the ability of the visual system to perceive color independently of the ambient illumination, was investigated in the context of a biologically-based neural network. In particular, the role of retinal adaptation and higher level visual operations in mediating color constancy was investigated. The study incorporated properties of individual cells and how they combine to make complex color and spatial operations. The neural network simulations indicate how early visual stages complement each other to compensate and maintain relatively constant color perception under conditions of varying illumination and spatial context in the image.

The research was reported in papers published in scientific journals which are included as appendices A-H. The following pages include a list of the publications as well as other activities, and the personnel involved.
PUBLICATIONS

PAPERS IN SCIENTIFIC JOURNALS:


BOOK CHAPTERS:


CONFERENCES, PROCEEDINGS, SYMPOSIA, WORKSHOPS:


PARTICIPATING PERSONNEL:

Faculty:

Buchsbaum, Gershon
Finkel, Leif, H.

Graduate students and Ph.D. thesis titles


Levitan, Bennett, S., (M.D./Ph.D.), Image Propagation through Multiple Retinal Cell Layers: Multirate Filter-Based Modeling and Analysis in the Cat Retina (1994)
APPENDICES:

The appendices are arranged in three groups corresponding to the research thrusts.

**Group 1: Spatiotemporal visual image coding**


**Group 2: Signal propagation in the retina**


**Group 3: Color constancy and interactions of space and color**


Appendix H: A Multi-Stage neural network for Color Constancy and Color Induction, IEEE Transactions on Neural Networks, in press, 1995
Appendices
Group 1: Spatiotemporal visual image coding


Efficient coding of natural time varying images in the early visual system

MICHAEL P. ECKERT AND GERSHON BUCHSBAUM

Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, 220 S. 33rd Street, Philadelphia, Pennsylvania 19104-6315, U.S.A.

SUMMARY

We investigate the hypothesis that the early visual system efficiently codes natural time varying images, first by tracking part of the image, then by matching the spatiotemporal properties of the neural pathway to those of the tracked image. A representation for the time varying image is formulated which consists of two spatiotemporal components, a velocity field component and a stationary component. We show, using digitized sequences of natural images, that the spatiotemporal spectrum and other attributes of the image markedly differ before and after tracking. The temporal frequency bandwidth and velocity distribution of the velocity field component are diminished in the region of tracking and broaden with increasing eccentricity from this region. On the other hand, the spectrum of the stationary component is unaffected by tracking. Comparison of the properties of the tracked image to those of the M and P pathways suggests that each pathway transmits different attributes of the tracked image. A retinal architecture which varies with eccentricity also matches the properties of the tracked image.

1. INTRODUCTION

Natural images contain spatiotemporal information comprised of motion and other time varying details such as flicker. Motion in the retinal image includes object motion in the visual scene, observer motion, and eye motion. Image motion presents a significant problem for efficient coding and representation of images in the visual system. The visual system must code and interpret the visual scene while accounting for objects moving at velocities which may exceed the temporal limitations of visual system processing.

The visual system confronts this complex time varying signal with two spatiotemporal mechanisms: eye movements and the spatiotemporal filter arrays known as the M and P pathways. In this paper, we examine how eye movements and the M and P pathways conjoin to make an efficient coder of the time varying image. Eye movements limit the temporal bandwidth of images by reducing the range of velocities reaching the fovea. M and P pathways efficiently carry image components, modified by eye movements, for analysis at cortical levels. The spatiotemporal properties of ganglion and lateral geniculate cells which form the M and P pathways have been extensively investigated in recent years (Kaplan & Shapley 1982; Hicks et al. 1982; Derrington & Lennie 1984; Blakemore & Vital-Durand 1986; Crook et al. 1988; Lee et al. 1989a; Purpura et al. 1990), especially the role of these pathways in coding spatiotemporal image components (Shapley & Perry 1986; Merigan 1986; Merigan 1989; Merigan & Maunsell 1990; Merigan et al. 1991; and Schiller et al. 1990). Generally, the M pathway is associated with fast temporal changes and the P pathway with high spatial acuity and colour, although there is considerable overlap across spatial and temporal frequencies.

The idea that the visual system efficiently codes the visual scene is not a new one (Barlow 1961, 1981; Snyder et al. 1977; Srinivasan et al. 1982; Buchsbaum & Gottschalk 1983; Laughlin 1983; Field 1987; Tsukamoto et al. 1990; Watson 1990; Derrico & Buchsbaum 1991). Under the efficient coding hypothesis, the purpose of retinal processing is to transmit visual information as effectively as possible to higher visual centers. This means that the visual system optimizes its coding strategy, given the physiological constraints of limited dynamic range of nerves, noise, and limited spatial and temporal bandwidths.

A general block diagram of the coding system under investigation is presented in figure 1. The coder is comprised of two components, the pre-retinal eye movements, modelled as a linear time variant filter, and the retinal spatiotemporal pathways, modelled as linear time invariant filters. Investigation of efficiency and other properties of the coder requires an understanding of the signal environment in which it operates. For the visual system, the environment is an observer freely viewing natural images. We begin by investigating the spatiotemporal spectrum of the time varying image (the input, and the effects of tracking on the spatiotemporal spectrum and other properties of the image (the signal at the retina, , in figure 1). A representation of natural images...
The velocity distribution is formed by sampling the velocity field component at the point of tracking (Jain & Jain 1981; Girod 1987), but the temporal frequency bandwidth and velocity distribution broaden with increasing eccentricity from the point of tracking. The spectrum of the stationary component is not affected by tracking.

We discuss the properties of the M and P pathways and how they match the spatiotemporal components of images after tracking. We give special attention to calculating the effect of tracking in the region of the retina. Spatial invariance filters are a linear filter which removes the velocity field component. The residual spatiotemporal intensity variations, the stationary component, will consist of flicker of the illuminant, the photometric effects of motion, and the occlusion and disocclusion at the edges of moving objects. These spatiotemporal effects are biologically relevant. For example, photometric motion provides depth and three-dimensional structure information about the image (Pentland 1991), and occlusion effects provide information about the location of object edges and relative depth. While these effects are caused by object motion, they cannot be removed by translational shifts of image intensity, and thus cannot be incorporated into the velocity field component.

The space and time variant spatiotemporal spectrum derived from the model of equation (1) is

$$S(u, t, k, f) = S(k)\delta (f - \nu(u, t) \cdot k) + S_1(k, f),$$

where $I(u, t)$ is the intensity at spatial point $u$ and time $t$, $I(u, 0)$ specifies the initial image intensity, $\nu(u, t)$ is the velocity field assigning a velocity vector, $\nu = (\nu_x, \nu_y)$, to each point of space and time, and $S_1(u, t)$ is the stationary component.

The spectral component can be formed by removing local translational motion. Conceptually, this operation is analogous to filtering the image with a space-time variant filter which removes the velocity field component. The residual spatiotemporal intensity variations, the stationary component, will consist of flicker of the illuminant, the photometric effects of motion, and the occlusion and disocclusion at the edges of moving objects. These spatiotemporal effects are biologically relevant. For example, photometric motion provides depth and three-dimensional structure information about the image (Pentland 1991), and occlusion effects provide information about the location of object edges and relative depth. While these effects are caused by object motion, they cannot be removed by translational shifts of image intensity, and thus cannot be incorporated into the velocity field component.

The space and time variant spatiotemporal spectrum derived from the model of equation (1) is

$$S(u, t, k, f) = S(k)\delta (f - \nu(u, t) \cdot k) + S_1(k, f),$$

where $\delta$ is the Dirac delta function, $k = (k_x, k_y)$ is a two-dimensional spatial frequency vector, $f$ is the temporal frequency, $S(k)$ is the spatial power spectrum of $I(u, 0)$, and $S_1(k, f)$ is the spatiotemporal spectrum of the stationary component. By definition, the velocity field and stationary component are uncorrelated, so the spectrum is the sum of the two components. In local spatiotemporal regions, the energy of the velocity field component exists on a plane in the three dimensions of frequency space (2 dimensions spatial, 1 dimension temporal), where the local velocity determines the orientation of the plane (Watson 1983; Watson & Ahumada 1985). A highly ordered structure does not exist for the stationary component which is distributed throughout spatiotemporal frequency space. Figure 2 illustrates the differences between the spectra of these components in a local spatiotemporal neighborhood.

The spatiotemporal spectrum (equation 2) can also be defined in terms of a velocity distribution, which is a probability distribution (histogram) of velocities. The velocity distribution is formed by sampling the velocity field through time at every spatial point in
Eye movements have the effect of shifting all velocities in the image by the eye velocity. In the case of tracking, eye velocity is set equal to the velocity at spatial point, \( u_0 \).

\[ v_e(t) = v(t, u_0). \]  

Tracking has the effect of minimizing the spread of the velocity distribution, \( h_e(u, v) \), which decreases the temporal bandwidth of the signal in the neighborhood of the tracked point, \( u_0 \). For perfect tracking, the velocity distribution becomes a delta function at point, \( u_0 \), and we have the spatiotemporal spectrum.

\[ S(u_0, k, f) = S(k) \delta(f) + S(k, f). \]  

Because tracking compensates for the velocity field component in the region of \( u_0 \), the temporal variations are contributed by the stationary component, \( S(k, f) \).

For points away from \( u_0 \), the spectrum is weighted by the velocity distribution.

\[ S(u - u_0, k, f) = h_e(u - u_0, v) S(k) + S(k, f), \]  

where \( h_e(u - u_0, v) \) is a velocity distribution which broadens with increasing eccentricity, \( |u - u_0| \). In the region around \( u_0 \), tracking narrows the velocity distribution, thereby reducing the variability of the spatiotemporal spectrum and the temporal frequency bandwidth. With increasing eccentricity from the point of tracking, the spatiotemporal spectrum will have a broader velocity distribution and larger temporal frequency bandwidth. The degree to which tracking narrows the velocity distribution away from the point \( u_0 \) depends on the spatial correlation of the velocity field, which is a measure of the change of the velocity field across space. For a highly correlated velocity field which changes slowly through space, tracking can reduce the velocity distribution at relatively large distances from the point of tracking.

Implementation of tracking invariably requires a feedback loop which estimates position and velocity from a time delayed input and past expectations (Stark et al. 1962; Lisberger et al. 1987; Steinman et al. 1990). This means that eye velocity can be set only to an estimated value of image velocity, rather than the true image velocity. As a result, the velocity distribution in the tracked region, \( h_e(0, v) \), will have a spread related to the effectiveness of tracking. Tracking effectiveness is a signal related phenomenon, so that highly predictable motion will be more effectively tracked than unpredictable motion (Stark et al. 1962; Barnes & Lawson 1989). However, tracking of 'real world' motion can be quite accurate (Steinman et al. 1990), usually maintaining a foveal velocity of less than 1–2 deg s\(^{-1}\).

3. THE EFFECT OF TRACKING ON DIGITIZED IMAGE SEQUENCES

We calculated the velocity field, velocity distribution, and spatiotemporal spectrum of four real world image sequences before and after tracking objects in the sequences. The sequences: 256 pixels x 256 pixels x 64 frames at 8 bits per pixel, 30 frames per second with no scene cuts. were taken from a video disk which...
Table 1. Description of sequences

<table>
<thead>
<tr>
<th>sequence number</th>
<th>sequence description</th>
</tr>
</thead>
<tbody>
<tr>
<td>U10833</td>
<td>Jungle scene with some three-dimensional object motion and a small amount of camera motion</td>
</tr>
<tr>
<td>IJ12426</td>
<td>Man walking. Some camera motion to keep man centred in visual field. The result is significant amounts of background motion</td>
</tr>
<tr>
<td>IJ01300</td>
<td>Man talking while moving head occasionally. No camera motion. Some slight motion in the background</td>
</tr>
<tr>
<td>IJ04454</td>
<td>Storm scene. No camera motion, but large amounts of non-rigid three-dimensional motion from waves. Intensity changes from lightning</td>
</tr>
</tbody>
</table>

contained scenes from movies. The frame rate of 30 frames per second limits the maximum estimated temporal frequency bandwidth of the images to 15 Hz. However, image energy drops off quickly with temporal frequency, and we found that signal energy is concentrated below 10 Hz. This suggests that any aliasing introduced by the sampling rate has little effect on the estimated spectrum. Each sequence was selected to contain varying levels of motion activity to form the broadest possible ensemble of images with the small sample size (see table 1). The velocity field for each frame was estimated by minimizing the squared difference between 24 pixel × 24 pixel blocks in two sequential frames of the sequence (Jain & Jain 1981). The same method was used to track selected regions in the sequence (see figure 3). While this (minimization of sum of squared differences) algorithm is unlikely to be the method used by the visual system.

Figure 3. A typical image sequence over which the spatiotemporal statistics were analysed (sequence IJ12426). The black box indicates a region which was selected for tracking. The graph illustrates the magnitude of the velocity of the tracked region as a function of time.
it tracks a region of the image and keeps it centred, as does smooth pursuit.

We examined the effect of tracking on the velocity distribution in the region of tracking and at increasing eccentricities from that region. The velocity distribution of each 64 frame sequence was calculated from the velocity field as the frequency of occurrence of the velocity magnitude. To calculate the change of the velocity distribution with eccentricity, a region of interest was selected and tracked by shifting the entire image for each frame to maintain the tracked region in the same spatial location (figure 3). The velocity distribution and average velocity for the tracked image were then computed as a function of eccentricity from the point of tracking. The results are presented in figures 4 and 5. Before tracking, the velocity distribution and average velocity vary across the spatial extent of the image, but we found no trend from sequence to sequence. This is expected, as there is no reason why one part of the scene should experience more motion than any other part. When tracking is performed, the velocity distribution at the point of tracking is a delta function (for perfect tracking), and broadens with eccentricity. Higher velocities occur more frequently with increasing eccentricity from the region of tracking. The increase in average velocity with eccentricity was accompanied by a corresponding increase in the standard deviation, reflecting the fact that the variability of the velocity distribution increases with eccentricity. This result was consistent across all four sequences examined, although the degree to which tracking reduced the velocity distribution away from the point of tracking varied depending on the spatial correlation of the velocity field in that particular sequence. At the largest eccentricities, the velocity distribution after tracking can exceed the distribution before tracking. This can occur because the velocity distribution at large eccentricities is the vector sum of two uncorrelated velocity components, the image velocity and the velocity of eye movements.

The changes in the velocity distribution as a result of tracking have corresponding effects in the spatiotemporal frequency spectrum (equation 7). In figure 6 we compare the spatiotemporal spectrum of images before and after tracking. The spectrum was computed in spatially and temporally localized blocks of size 32 pixels x 32 pixels x 16 frames at the point of tracking for a viewing distance of four screen heights (1 screen height = 256 pixels). For purposes of comparison, 1 pixel = 1/15 deg and 1 pixel per frame = 2 deg s^{-1} on a 256 pixel x 256 pixel image at a standard viewing distance of 4 screen heights with a frame rate of 30 frame per second. Most of the energy is concentrated below 10 Hz and 4 cycles per degree and diminishes quickly above these frequencies. After tracking, the temporal bandwidth of the spatiotemporal spectrum is greatly reduced. This is consistent with the changes in the velocity distribution after tracking (figure 4b). A reduction in the velocity distribution decreases the temporal bandwidth of the spatiotemporal spectrum.

Figure 3. The average velocity in sequence IJ01300 before and after tracking (solid and dashed line respectively) as a function of eccentricity from the tracked point. The average was computed over the full 64 frames of the sequence. After tracking, the average velocity in the region of tracking drops to zero, and there is a regular increase in average velocity with eccentricity from the point of tracking. At large eccentricities, the average velocity after tracking can be larger than the average velocity before tracking.

Figure 4. The velocity distribution of sequence IJ12426 as a function of eccentricity from the point of tracking. The distribution was computed from a 64 frame sequence. The curves represent the distribution at eccentricities of 0, 16, 32, 48, 64, 80, 96 pixels from the point of tracking. (a) Before tracking, the velocity distribution does not depend on eccentricity. The standard deviation for the curves is about 1.5 pixels per frame at all eccentricities. (b) After tracking, the distribution varies with eccentricity, from a delta function at the point of tracking, to a broad distribution at the largest eccentricity. The standard deviation increases with eccentricity and is 0, 0.51, 0.75, 1.6, 1.7, 1.71, 1.8 pixels per frame, respectively, for the eccentricities shown.
Figure 6. The spatiotemporal spectrum of the image before and after tracking computed for a $32 \times 32 \times 16$ spatiotemporal block from sequence IJI2426. The contour plots have lines at 0.01, 0.05, 0.1, 0.5 of the maximum values. The bottom figures are surface plots of the corresponding spectra. a Before tracking, the image has a large temporal bandwidth due to occasional large velocities. b After tracking, spatiotemporal energy is concentrated in lower temporal frequencies. The units of cycles per degree for spatial frequency axis were determined by using a viewing distance of 4 scene heights from the image. The range of spatial frequencies will vary for different viewing distances, but the range of temporal frequencies will not vary with viewing distance.

Figure 7 shows how tracking modifies the instantaneous temporal frequency bandwidth in the region of tracking. The instantaneous temporal frequency bandwidth was computed from the frame to frame correlation, $p_T$, using a correlation model of the form $p_T = e^{-\alpha t}$ (Jayant & Noll 1984; Eckert et al. 1992). For a frame rate of 30 frames per second, the temporal bandwidth can be computed as $\Delta f = -30 \log p_T$. Before tracking, the signal has a highly variable temporal frequency bandwidth, due to variable object velocity. After tracking, the temporal bandwidth is small and less variable. The temporal bandwidth is not zero after tracking, however, since tracking does not remove all time variations, only those which result from translational motion at the tracked point. Figure 7 highlights the two ways tracking affects temporal frequency bandwidth: (i) tracking greatly reduces the temporal frequency bandwidth during high velocity motion, and (ii) before tracking, the temporal frequency bandwidth fluctuates widely, depending on the velocity, but after tracking, the temporal frequency bandwidth in the region of tracking has only small fluctuations.

Tracking does not remove the velocity field component, but only shifts it to lower temporal frequencies. This shift will increase the energy share of the stationary component at high temporal frequencies. The relative share of the two components will vary from sequence to sequence and for different regions of tracking, depending on the amount of motion, and the degree to which time variations are represented by the velocity field component or by the stationary component. For the first three sequences in table 1, tracking of an object in the foreground reduces the average temporal bandwidth in the tracked region by 59%, 92%, and 56%, respectively. For these sequences, the large decrease of the temporal frequency bandwidth after tracking signifies that the velocity field component accounts for much of the signal energy at high temporal frequencies of the untracked image. The average temporal bandwidth of the last sequence...
storm scene IJ04454 was reduced by only 22%. As tracking had little effect on the temporal bandwidth of this sequence, most time variations can be attributed to the stationary component. Large temporal intensity changes in this scene were due primarily to changing illuminant (lightning) and changing reflectance of light off ocean waves.

4. CODING BY THE VISUAL SYSTEM IN THE CONTEXT OF TRACKING

(a) Advantages of tracking

The velocity field of natural time varying images is signal dependent and variable. A scene may contain objects moving at high velocities, low velocities, or both. The corresponding spatiotemporal spectrum is also signal dependent and variable, with a large temporal bandwidth in the spatial regions which move at high velocities, and a small temporal bandwidth in slowly moving regions. A basic premise of coding theory is that a signal with a small bandwidth can be more efficiently coded than a signal with a large bandwidth (Javant & Noll 1984). The coding efficiency is also affected by signal variability, because time-invariant coders (such as the retinal pathways) can be optimized only for a particular spectrum. Kassam & Poor 1983, 1985. The most efficiently coded signal is one with a small bandwidth and little or no variability. This corresponds to an image with little or no velocity, and thus a small temporal frequency bandwidth. Tracking with eye movements compensates for motion by matching eye velocity to the expected value of the image velocity in a region around the fovea. After tracking, the signal which actually reaches the retina (at the fovea) has a narrow velocity distribution and, therefore, a reduced temporal frequency bandwidth. A direct corollary of minimizing the temporal frequency bandwidth is a reduction in blur due to motion when the image is coded by fixed bandwidth time invariant channels. The role of eye movements in reducing blur as suggested before (Miller & Ludvigh 1962; Murphy 1978; Flipse et al. 1988) and follows from their role in the context of efficient coding.

Field (1987) showed that the spatial spectrum of natural images is scale invariant. This enables the visual system to use fixed, scene invariant, spatial filters to efficiently code a scene regardless of scale. The temporal spectrum and velocity distribution of natural time varying images are not scale invariant, and depend on the distance of moving objects from the observer. However, tracking maps the tracked region into the same temporal frequency and velocity distribution range regardless of the velocity (or scale) of the tracked region. Therefore, tracking provides a region of the retina with a virtually scale invariant signal in time and the coding advantages that accrue from the invariance.

In addition to increased coding efficiency, tracking accentuates the importance of the stationary component in the temporal frequency domain in the region of tracking. Before tracking, this component is difficult to detect and isolate because it cannot easily be separated from the velocity field component. After tracking, the velocity field component is dominated along the spatial frequency axis, and does not contribute to temporal variations, so the remaining temporal variations belong to the stationary term. Thus, perceptually important information associated with this component, such as flicker, photometric effects of motion, and motion edge effects, can be extracted more easily because of the removal of the velocity field component from high temporal frequencies. This argument does not hold in the periphery, however, because tracking only reduces the velocity field component of the region of tracking.

(b) Retinal pathways and eccentricity dependent architecture are matched to the tracked image

The second stage of the coder (figure 1) are the M and P pathways which operate on the tracked image, i_s(t). These pathways and the underlying single cell units from which they are made have received considerable attention. Because of their significance in the present context, they are briefly reviewed here. The spatiotemporal filter properties of the M and P pathways are based on single cell properties of phasic and tonic cells from the retina and M and P cells from the LGN (Marrocco et al. 1982; Kaplan & Shapley 1982; Hicks et al. 1982; Derrington & Lennie 1984; Blakemore & Vital-Durand 1986; Crook et al. 1988; Lee et al. 1989b; Purpura et al. 1990). P (tonic) cells respond well to low temporal frequencies (below 5 Hz), whereas M (phasic) cells attenuate these frequencies. The spatial resolution of the P pathway is about three times higher than the M pathway at all eccentricities. This is due to receptive field center size and spatial sampling rates of the respective arrays (Merigan 1989; Merigan et al. 1991). The main characteristics of the pathways are summarized in table 2. The numbers in table 2 represent averages within the respective pathways rather than the response of any particular cell since there are large deviations among cells even in the same pathway (Hicks et al. 1982; Marrocco et al. 1982; Derrington & Lennie 1984).

Figure 8 illustrates the spatiotemporal transfer function of the M and P pathways inferred from the specifications in table 2. To obtain these responses, we fitted a frequency transfer function with the form of a spatial and temporal difference of Gaussians (Rohaly & Buchsbaum 1988; Rohaly 1988) and selected constants so as to meet the spatial and temporal frequency slopes and peaks in table 2.

\[
RF(|k|, f, e) = C_1 e^{-\left(\frac{r_c|k|}{T_c}\right)^2} - S_1 e^{-\left(\frac{r_s|k|}{T_s}\right)^2} + C_2 e^{-\left(\frac{r_c|k|}{T_c}\right)^2} - S_2 e^{-\left(\frac{r_s|k|}{T_s}\right)^2},
\]

where \(e\) is eccentricity, \(r_c(e), r_s(e)\) are the centre and surround sizes for receptive fields, and \(k\) and \(f\) are the spatial and temporal frequencies, respectively. \(T_c, T_s\) are temporal constants selected so as to provide peak temporal response at a specified frequency, and \(C_1, C_2, S_1, S_2\) were selected so as to provide a specified response at low temporal and spatial frequencies. The
Table 2. Spatial and temporal characteristics of the M and P pathways

<table>
<thead>
<tr>
<th></th>
<th>M pathway</th>
<th>P pathway</th>
</tr>
</thead>
<tbody>
<tr>
<td>spatial structure</td>
<td>centre-surround (relatively powerful surround)</td>
<td>centre surround (surrounds often have little power)</td>
</tr>
<tr>
<td>spatial resolution</td>
<td>one-third that of P cells (decreases with eccentricity)</td>
<td>three times that of M cells (decreases with eccentricity)</td>
</tr>
<tr>
<td>foveal spatial resolution</td>
<td>13 cycles per degree</td>
<td>40 cycles per degree</td>
</tr>
<tr>
<td>numbers of cells</td>
<td>10% of cells</td>
<td>80% of cells</td>
</tr>
<tr>
<td>spatial sampling rate</td>
<td>one-third that of P cells</td>
<td>three times that of M cells (80 samples per degree at fovea)</td>
</tr>
<tr>
<td>temporal frequency-peak</td>
<td>20 Hz; large variance between individual cells</td>
<td>10 Hz; large variance between individual cells</td>
</tr>
<tr>
<td>response at low frequencies</td>
<td>highly attenuated; phasic response to a step increase in light intensity</td>
<td>partially attenuated; tonic or sustained response to a step increase in light intensity</td>
</tr>
<tr>
<td>high frequency cutoff</td>
<td>up to 80 Hz</td>
<td>20–40 Hz</td>
</tr>
<tr>
<td>contrast sensitivity</td>
<td>high; eight times higher than P cells</td>
<td>low; eight times lower than M cells</td>
</tr>
<tr>
<td>speed (latency of response to visual stimulation)</td>
<td>fast; latency is about 24 ms at LGN for visual stimulation. Large variance</td>
<td>slow; latency is about 28 ms at LGN for visual stimulation. Large variance</td>
</tr>
<tr>
<td>suggested roles</td>
<td>carries information about quickly moving images with a low degree of spatial detail, such as flicker</td>
<td>carries slowly moving images with a high degree of spatial detail</td>
</tr>
</tbody>
</table>

centre and surround sizes for the receptive fields, \( r_c(e) \) and \( r_s(e) \), are assumed to increase in size with the inverse of the cortical magnification factor (Sakitt & Barlow 1982).

\[
\begin{align*}
    r_c(e) &= r_c(0)(1 + 0.33e), \\
    r_s(e) &= r_s(0)(1 + 0.33e).
\end{align*}
\]

The spatiotemporal frequency the velocity responses of M and P pathways can now be discussed in the context of the properties of images before and after tracking. Before tracking, the image spatiotemporal

Figure 8. The spatiotemporal transfer function of the M and P pathways calculated from equation (8) with constants chosen to match details in table 2 (solid lines, M pathway; dashed lines, P pathway). Contours are at 0.1, 0.5, and 0.9 of the maximum value in the respective pathway. The P pathway is tuned to higher spatial frequencies and lower temporal frequencies than the M pathway, though there is considerable overlap.

Figure 9. (a) The simulated response of M and P pathways at the fovea to a translating white noise stimulus as a function of velocity (solid line, M pathway; dashed line, P pathway). This is equivalent to integrating the frequency response over lines of constant velocity. The normalized peak response of the M and P pathways (arrows) is about 7 deg s\(^{-1}\) and 1 deg s\(^{-1}\), respectively. (b) Peak velocity of the two pathways as a function of eccentricity. The peak velocity of the M pathway increases with eccentricity at a greater rate than the P pathway.
The P pathway is thought to carry information about slowly moving images with a high degree of spatial detail (Schiller et al. 1990; Merigan 1989; Merigan et al. 1991). Figure 9a illustrates that the P pathway will respond better to low velocity images. The P pathway matches the properties of the velocity field component in the tracked region, and will carry the maximum amount of spatiotemporal information about this component. The M pathway is thought to carry information about quickly moving images with a low degree of spatial detail (Schiller et al. 1990; Merigan 1989; Merigan & Maunsell 1990). Figure 9a illustrates that the M pathway is tuned to higher velocities than the P pathway. However, large image velocities will only rarely arise in the tracked region. When large velocities do arise, it is during tracking errors which occur for unpredictable motion, and for cases such as transparent motion when there are two velocity field components in the same spatial region. Because tracking is generally quite accurate for motion of ‘real world’ stimuli (Steinman et al. 1990), the M pathway can be expected to carry only a small fraction of the velocity field component of image information in the region of tracking (the fovea). However, the stationary component is broadly distributed across spatiotemporal frequency space (figure 2b), and contains a significant amount of energy in the region covered by the M pathway. Therefore, in addition to carrying (infrequent) high velocity images, another role for the M pathway at the fovea could be to carry the stationary component of time varying images.

The change of the velocity tuning of the M and P pathways with eccentricity (figure 9b) is consistent with the change in the velocity distribution of tracked images with eccentricity. The peak velocity of the two pathways increases with eccentricity, though at different rates, so a larger average velocity and larger range of velocities is covered with increasing eccentricity. This can be compared with the velocity distribution after tracking figure 4b. At the fixation point fovea, the image has a narrowly distributed velocity distribution and a small temporal bandwidth. With increasing eccentricity, the velocity distribution broadens (figure 4b), the average velocity reaching the retina increases (figure 5), and the range of velocities increases. The broader image velocity distribution in the periphery means that information is lost due to temporal blur because of the limited temporal frequency bandwidth of retinal pathways. This decreases the average spatial frequency limit of the peripheral retinal image. Because of this, larger receptive fields can be utilized in the periphery without significant loss of information.

Psychophysical evidence also shows a gradual change in motion perception between the fovea and the periphery of the visual field. The fovea is sensitive to a lower range of velocities than the periphery and essentially becomes blind when this velocity range is exceeded (van de Grind et al. 1986; Baker & Braddick 1985). As eccentricity increases, the visual system is better able to discriminate images with a higher average velocity, and over a larger range of velocities (McKee & Nakayama 1984). This is consistent with the change in velocity distribution and average velocity with eccentricity (see figures 4 and 5) which results from tracking.

Hughes (1977) argues that receptor packing matches the change in velocity across the retina for the case of an observer moving through a scene (egomotion). In some way, this paper can be viewed as a generalization of Hughes (1977) original arguments, by showing that an eccentricity dependent velocity distribution results for any scene rather than the special case of ego-motion, as long as the observer continually tracks with eye movements. This paper diverges from Hughes by matching the retinal velocity distribution to the velocity sensitivity of the M and P pathways, rather than to the change in receptor packing. However, receptive field size and velocity sensitivity are linked so both arguments are complementary.

5. CONCLUSION

We examined the spatiotemporal spectrum and other attributes of natural time varying images in the context of efficient coding in the early visual system. The image is modeled as a combination of a velocity field component and a stationary component which have markedly different spatiotemporal spectra. Tracking, as implemented with smooth pursuit eye movements, decreases the average velocity and the variability of velocities reaching the fovea (tracked region). The result is a spectrum with minimal temporal bandwidth and variability in the tracked region, but which broadens with increasing eccentricity. Tracking does not affect the stationary component, which remains broadly distributed across temporal frequency space.

An efficient coding strategy will be influenced by tracking because it changes the image spectrum. In the tracked region, the spectrum has minimal temporal bandwidth and variability. This enables efficient coding of the image with fixed time invariant path-
ways as found in the retina. The reduction in temporal bandwidth ensures that minimal information will be lost due to motion blur in the tracked region. The stationary component of time varying images is emphasized in the tracked region, enabling temporal information not attributed to translational motion to be analyzed effectively. Finally, since the average velocity of the image increases with eccentricity from the tracked region, an efficient coding strategy should reflect this change with a corresponding change in velocity tuning with eccentricity.

The results suggest that the M and P pathways are matched to the tracked image. Both the M and P pathways are tuned to low image velocities at the fovea, where the image has consistently low velocities because of tracking. However, the M pathway, with the broader temporal frequency response, will respond better to the temporal changes of the stationary component. The M and P pathways are tuned to higher velocities and a broader range of velocities with increasing eccentricity from the fovea. This is matched to the change of image velocity after tracking, in which both the average velocity and range of velocities increase.

In conclusion, the visual system combines smooth pursuit tracking with specialized pathways and an eccentricity dependent retinal architecture to efficiently code time varying images.

We thank Horace Barlow and Peter Sterling for their many comments and suggestions, and Andrew B. Watson for his aid in collecting the image sequences. This work was supported by AFOSR grant 91-0082 and the NASA graduate student fellowship program.

REFERENCES


Received 4 March 1992; accepted 19 October 1992
Effect of tracking strategies on the velocity structure of two-dimensional image sequences

Michael P. Eckert and Gershon Buchsbaum
University of Pennsylvania, School of Engineering and Applied Science, Department of Bioengineering, 220 South 33rd Street, Philadelphia, Pennsylvania 19104-6315

Received November 23, 1992; accepted February 10, 1993

We investigate the effect of different tracking strategies, such as local and full-field tracking, on the mean and variance of the image velocity field. We show that while local tracking reduces the velocity variability in an eccentricity-dependent manner, full-field tracking reduces velocity variability equally across the image. We test our predictions with digitized image sequences.

INTRODUCTION

It is well known that tracking actively modifies the spatio-temporal structure of the visual scene; it shifts image velocity and position, thus making the tracked image markedly different from the original image. In this Communication we examine and formalize how different tracking strategies affect the local and global velocity structures of image sequences. To describe the local velocity structure, we use the velocity field, which assigns a velocity vector to the image at each point in space and time and is specific to each scene. For a more global description of image velocity, we calculate the variance of the velocity field, sampled through time. We focus on two strategies of tracking: local tracking, in which a small spatial region in the original image is pursued, and full-field tracking, in which the average image velocity is tracked. These two tracking strategies reduce the variance of the velocity distribution in markedly different ways. Local tracking greatly reduces the variance at the point of tracking, but the variance increases with eccentricity away from the point of tracking in a manner dependent on the spatial correlation of the velocity field. Full-field tracking, on the other hand, reduces the variance equally across the entire spatial extent of the image. We illustrate the analysis with examples by using image sequences.

VELOCITY FIELD OF THE TRACKED IMAGE

The velocity field assigns a velocity vector to each point of the image plane to represent intensity variations resulting from geometric motion in the world. Types of motion that are adequately described by the velocity field include affine transformations such as translation, rotation, and dilation that are commonly found with perspective projections of three-dimensional motion.

We write the velocity field as the sum of two components,

\[ v(u, t) = v_d(u, t) + v_m(u, t), \]

where \( v_m(t) \) denotes the time-varying mean velocity and \( v_d(u, t) \) denotes differential image velocities that vary with both space and time. Each of these components can be associated with common physical sources of motion. Changes in the time-varying mean velocity occur during camera pans (or head and body rotation in the biological case). \( v_m(u, t) \) represents differential velocities across the field of view, such as motion of objects or linear motion of the camera through the world (ego motion).

Tracking introduces a single time-varying vector-velocity term to the velocity field. After tracking, the velocity field is simply the difference between the velocity field of the original image and the velocity of tracking:

\[ v_t(u, t) = v(u, t) - v_t(t), \]
where \( \mathbf{v}(u, t) \) is the tracked velocity field. This relationship is valid for any type of tracking strategy, including both full-field and local tracking.

We consider the effect that the two cases of tracking have on the velocity field. Local tracking, in the limit, consists of tracking a single point, \( \mathbf{v}(t) = \mathbf{v}(u_0, t) \). And using Eq. (2) gives

\[
\mathbf{v}(t) = \mathbf{v}(u_0, t) + \mathbf{v}_n(t).
\]

Substituting Eqs. (2) and (4) into Eq. (3) shows that local tracking removes the mean velocity, leaving a tracked velocity field of

\[
\mathbf{v}(u, t) = \mathbf{v}(u_0, t) - \mathbf{v}_n(t).
\]

Full-field tracking, with the definition from Eq. (1), amounts to setting the velocity of the tracked field to the time-varying mean image velocity, i.e., \( \mathbf{v}(t) = \mathbf{v}_m(t) \). Thus the velocity field after full-field tracking is simply that of the differential velocity term,

\[
\mathbf{v}(u, t) = \mathbf{v}_d(u, t).
\]

**VARIANCE OF THE VELOCITY FIELD AFTER TRACKING**

The velocity field provides an instantaneous and locally specific measure of image velocity. A more global measure is produced by calculating the mean and variance of the original and tracked velocity fields. For this purpose, we consider the velocity field of an image sequence to be a realization of a three-dimensional (two-dimensional space, one-dimensional time) vector random field. The mean and variance are then calculated by sampling the velocity field across space, through time, or both. In this Communication, velocity samples are collected through time at every point in space. Conceptually, these samples form a velocity distribution or density function and represent a time-averaged measure of image velocity at that point in space. Since velocity is a vector quantity, the mean and variance of the horizontal and vertical velocity components are calculated separately and are considered independent and uncorrelated. We assume that \( \mathbf{v}_n(u) \) and \( \mathbf{v}_m(t) \) have a zero temporal mean. The basis for this assumption is that, on average, there will be no long-term fixed mean-velocity bias in image sequences.

After local tracking, the mean velocity will be zero, but the velocity variance will depend on distance from the point of tracking:

\[
\sigma^2(u - u_0)^2 = \sigma_v^2(u) + \sigma_d^2(u_0) - 2B_d(u - u_0),
\]

where \( \sigma_v^2(u - u_0)^2 \) is the variance velocity of the locally tracked image as a function of distance from the point of tracking; \( \sigma_d^2(u) \) is the variance of the differential velocity term; \( \sigma_d^2(u_0) \) is the variance of velocity at point \( u_0 \); and \( B_d(u - u_0) \) is the spatial autocorrelation function of the differential velocity term. The key point to recognize here is that the variance approaches zero at the point of tracking and will increase regularly with distance from the tracked point for a monotonically decreasing velocity-field autocorrelation function. At the largest eccentricities, the velocity-field autocorrelation approaches zero, so the variance is simply the sum of the first two terms.

Full-field tracking removes the time-varying mean from the original image. As a result, the full-field tracked image will have a zero mean and a variance equal to the variance of the differential velocity term:

\[
\sigma^2(u) = \sigma_d^2(u).
\]

The variance in Eq. (8) depends on the object velocity and is not a function of the spatial correlation of the velocity field.

Figure 1 illustrates the effect of local and full-field tracking, assuming constant variance for the differential velocity term and a spatial correlation of the velocity field that falls off exponentially with space. The variance of the velocity field after local tracking is greatly reduced in the region of tracking but increases with larger eccentricities. The reason for this is that the velocity field of the image is correlated with tracking velocity at the point of tracking but is almost completely uncorrelated at large distances from that point. Far from the point of tracking, \( |u| \gg 0 \), the variance of the velocity approaches twice the variance of the object-velocity term in the original image. Between these endpoints, the variance of the velocity field depends on the spatial correlation of the velocity field. Full-field tracking removes the mean velocity, \( \mathbf{v}_m(t) \), causing a downward shift of the variance from \( \sigma_v^2 \) for the original image to \( \sigma_d^2 \) for the tracked image. That is, the reduction in variance for full-field tracking is not dependent on eccentricity and amounts to removal of the variance contributed by the mean-velocity term, \( \mathbf{v}_m(t) \).

An implicit assumption is that the correlation between the differential velocity term and the time-varying mean velocity is small. In limiting cases, such as when a single object moves across a fixed background, the two terms are obviously correlated. However, the examples below, with the use of natural image sequences, suggest that the as-

---

**Fig. 1.** Predicted effect of local and full-field tracking on the variance of the velocity field as a function of eccentricity from the point of tracking. We use a spatial autocorrelation function for the differential velocity field of the form \( B_d(u) = \exp(-\|u\|/\xi_d) \). The dashed curves represent the variance of local tracking for different spatial autocorrelation functions of the velocity field when \( \alpha = (1/4, 2, 4) \). The bottom curve \( (\alpha = 4) \) illustrates the case in which the velocity field is highly correlated across space. The top curve \( (\alpha = 1/4) \) represents the case in which the velocity field is highly correlated across space. The endpoint of the top curve asymptotes at twice the variance of the full-field tracked image. Full-field tracking reduces the variance from \( \sigma_v^2 \) to \( \sigma_d^2 \) and is independent of the correlation of the velocity field.
Table 1. Description of Image Sequences

<table>
<thead>
<tr>
<th>Sequence Number</th>
<th>Sequence Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IJ10833</td>
<td>Jungle scene with some three-dimensional object motion and camera motion.</td>
</tr>
<tr>
<td>IJ12426</td>
<td>Man walking. Some camera motion to keep man centered in visual field. Result is considerable background motion.</td>
</tr>
<tr>
<td>IJ01300</td>
<td>Man talking while moving head occasionally. No camera motion. Some motion in the background.</td>
</tr>
</tbody>
</table>

EFFECT OF TRACKING ON IMAGE SEQUENCES

We simulated local and full-field tracking on three digitized image sequences and calculated the velocity field of the original image, the image after local tracking, and the image after full-field tracking. The sequences (256 × 256 × 64 pixels at 8 bits/pixel, 30 frames/s with no scene cuts) were taken from a video disk that contained scenes from movies (see Table 1). Local tracking was initialized by selecting a 24 × 24 pixel block in the first frame of the sequence. This region was then tracked by minimizing the squared difference between 24 × 24 pixel blocks in each pair of sequential frames in the sequence. The frames of the sequence were shifted so as to maintain the tracked object in the same spatial location for the entire sequence. This algorithm was effective because the scenes generally consisted of rigid, moving objects. Full-field tracking was implemented by shifting each frame in the sequence by the mean velocity of the entire image rather than by the velocity in a small region. With full-field tracking, no region was guaranteed to remain static throughout the sequence. The velocity field was calculated in a sparse array (every 16 pixels) by finding the displacement that minimized the squared difference between blocks in two sequential frames of the sequence. Thus the local velocity structure of each sequence was described by a 16 × 16 × 64 cube of velocity vectors. The original velocity field was calculated from the original image, and the local and the full-field tracked velocity fields were calculated from the scene after tracking. The variance of the original, the full-field, and the local tracked velocity fields were formed by sampling through time at each spatial location. The variance of each velocity distribution was then calculated at selected vertical and horizontal eccentricities from the tracked point.

Figure 2 illustrates the variance of the velocity field for the original image (solid curves), the local tracked image (dashed curves), and the full-field tracked image (dotted curves). The variance represents samples collected from a horizontal or a vertical eccentric point from the point of tracking. The different magnitudes of the variance from sequence to sequence reflect the different levels of motion activity in each sequence. Figure 2 also shows that the variance of the original velocity field can vary significantly across space. This is not surprising since a short, 2-s sequence will usually have different levels of motion activity in different areas of the scene. However, there was no systematic bias in motion activity across the scene in the different sequences.

Local tracking removes the variance of image velocity in the region of tracking but at the expense of increasing the variance at large eccentricities. This feature is consistent for all sequences, independent of the level of motion activity in the sequence. The rate at which the variance changes with eccentricity depends on the spatial correlation of the velocity field in each image. At large eccentricities, the variance of the local tracked image is greater than the variance of the full-field tracked image. This is expected because of the diminishing correlation of the velocity field between the tracked point and points at large eccentricities. For another limiting case, that of purely full-field motion, in which there is only a single translational velocity in the image, the differential velocity term, \( v_i(u, t) \), is zero, and local tracking decreases the variance at all eccentricities. For this case full-field and local tracking produce the same effect.
eccentricities. As expected, the reduction in variance after full-field tracking is well described by a downward shift of the variance of the original image. Differences among the variance curves that are not accounted for by the downward shift can be attributed to the correlation between the differential velocity field and the time-varying mean velocity and are small for all sequences.

DISCUSSION

Local or full-field tracking almost always reduces the variance of the velocity field. Local tracking ensures that one area of the visual field will possess low velocities but at the expense of increasing image velocity at large eccentricities from the point of tracking. Full-field tracking reduces velocities across the entire field of view, but no area is assured of consistently low velocities. If velocity magnitude of the tracked image is averaged across the extent of the image, with equal weight given to all areas, then the full-field tracked image will have a lower space-averaged variability than the local tracked image. This effect can be seen by inspection of Fig. 2. However, if the primary constraint is to reduce image velocity maximally in a small area, then local tracking is obviously more suitable. There are also other advantages of local tracking, which include regularizing high-level visual tasks such as structure from motion, wayfinding, and shape from shading. It remains to be seen whether full-field tracking can provide similar benefits.

An original motivation of this study was the significance of biological tracking strategies as implemented by various species using head and eye movements. Local tracking is a reasonable approximation of smooth-pursuit eye movements that are found in some vertebrate species, and full-field tracking is a good approximation of opticokinetic and vestibulo-ocular eye movements that are found in most, if not all, vertebrate species. As illustrated in this communication, different tracking strategies result in different degrees of velocity variability across the visual field. This finding suggests that biological velocity sensitivity at various eccentricities in a given species should reflect the particular tracking strategy most common to that species. Determining whether this is so requires further comparative research into tracking capabilities and eccentricity-dependent velocity sensitivity in different species.

ACKNOWLEDGMENT

This research was supported by U.S. Air Force Office of Scientific Research grant 91-0082.

REFERENCES AND NOTES

11. These digitized images are available from the authors.
Separability of Spatiotemporal Spectra of Image Sequences

Michael P. Eckert, Gershon Buchsbaum, and Andrew B. Watson

Abstract—We calculated the spatiotemporal power spectrum of 14 image sequences in order to determine the degree to which the spectra are separable in space and time and to assess the validity of the commonly used exponential correlation model found in the literature. We expand the spectrum by a singular value decomposition into a sum of separable terms and define an index of spatiotemporal separability as the fraction of the signal energy that can be represented by the first (largest) separable term. All spectra were found to be highly separable with an index of separability above 0.98. The power spectra of the sequences were well fit by a separable model of the form

$$P(k, f) = \frac{ab/(4\pi^3)}{((a/2\pi)^2 + k^2)^{3/2}((b/2\pi)^2 + f^2)}$$

where $k$ is radial spatial frequency, $f$ is temporal frequency, and $a, b$ are spatial and temporal model parameters that determine the effective spatiotemporal bandwidth of the signal. This power spectrum model corresponds to a product of exponential autocorrelation functions separable in space and time.

I. INTRODUCTION

The statistics of images and image sequences have been extensively studied for image coding and compression applications [1], [2] as well as for the development of models of biological image processing [3], [4]. An exponential autocorrelation function has been shown to be a good model for temporal frame-to-frame correlations of image sequences, e.g., [5]–[8], and for spatial correlations within each frame, e.g., [2], [3], [9].

This paper focuses on the separability of the spatiotemporal statistics of image sequences and on the validity of using a separable exponential autocorrelation model for the spatiotemporal statistics. The autocorrelation function is uniquely related to the power spectrum via a Fourier transform, and either is valid as a description of the statistics.

Manuscript received November 7, 1990; revised March 6, 1992. This work was supported by the NASA Graduate Fellowship Program and by grants NSF 8351637 and AFOSR 91-0082. Recommended for acceptance by Associate Editor N. Ahuja.

M. P. Eckert and G. Buchsbaum are with the Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA 19104-6392.

A. B. Watson is with the NASA Ames Research Center, Moffet Field, CA 94035-1000.

IEEE Log Number 9204244.

The spectra of 14 image sequences were calculated. The sequences represented a small ensemble of possible motion activity. The sequences were selected for a range of motion activity. For example, a fast camera pan represents the maximum image motion activity, and a small moving object with a static background represents the least activity. Sequences with motion activity between these extremes had slight camera motion and some object motion.

II. CALCULATION OF IMAGE STATISTICS

We collected 14 image sequences (256 × 256 × 64 @ 8 b/pixel, 30 frames/s with no scene cuts) from a video disc that contained scenes from a broadcast TV source. Each frame was originally sampled at 512 × 512 pixels/screen, but adjacent pixels were averaged, and the image was subsampled to 256 × 256 pixels/screen. The sample mean of each sequence was removed to reduce low-frequency bias in the calculations.

The sample power spectrum $P(k_1, k_2, f)$ of each sequence $x(n_1, n_2, t)$ is the squared magnitude of the discrete Fourier transform calculated as

$$P(k_1, k_2, f) = \frac{1}{256 \cdot 256 \cdot 64} \left| \sum_{n_1=0}^{255} \sum_{n_2=0}^{255} \sum_{t=0}^{63} x(n_1, n_2, t) e^{-j2\pi(k_1 n_1 + k_2 n_2 + f t)} \right|^2$$

where $k_1, k_2$ are spatial frequencies, $f$ is temporal frequency, $n_1, n_2$ are spatial locations, and $t$ is time measured in frame number.

We converted the two spatial frequency dimensions $k_1$ and $k_2$ into one radial frequency dimension $k$ by averaging in 32 annuli around the spatial frequency origin as illustrated in Fig. 1. In this manner, the spatial frequency range of 0–127 cycles/screen of $k_1$ and $k_2$ is represented by 32 annuli in bands of 4 cycles/screen. Averaging the spectral spectra in annuli is equivalent to assuming a circularly symmetric spatial autocorrelation function. This autocorrelation function is not separable in the two spatial dimensions but is considered a better fit than the corresponding separable autocorrelation function for most images [9].

The average magnitude of the power spectrum in each annulus can be obtained by summing over the power spectrum $P(k_1, k_2, f)$ in the annulus indexed by $k$ and normalizing by the number of sample
A space-time separable spectrum is modeled as the product of a spatial and temporal spectrum (as in (8)). In this section, we define an index of separability for an arbitrary spectrum $P(k, f)$ based on a singular value decomposition.

Any $m \times n$ matrix $D$ with $m \geq n$ may be expanded into a sum of terms by a singular value decomposition [10], [11]

$$D = \sum_{i=1}^{n} \sqrt{\gamma_i} u_i v_i^T,$$  \hspace{1cm} (9)

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ are the real nonnegative eigenvalues of the $n$th-order symmetric matrix $S = D^T D$, $u_1, u_2, \ldots, u_n$ are normalized, orthogonal row eigenvectors associated with the corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ of $S$, $v_1, v_2, \ldots, v_n$ are normalized, orthogonal column eigenvectors associated with the corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n$ of the $m$th-order symmetric matrix $Q = D D^T$, where $Q$ can have a maximum of $n$ nonzero eigenvalues that are the same as those of $S$. In the case of duplicate eigenvalues, an orthonormal combination of eigenvalues can be selected.

Approximating $D$ by the first term of the decomposition

$$D' = \sqrt{\gamma_1} u_1 v_1^T,$$  \hspace{1cm} (10)

gives the minimum mean squared error separable approximation to $D$, where the mean squared error is

$$e = \sum_{i=1}^{n} \sum_{j=1}^{m} (d_{ij} - d_{ij}')^2,$$  \hspace{1cm} (11)

where $d_{ij}$ and $d_{ij}'$ are the elements of $D$ and $D'$, respectively. Noting that

$$\sum_{i=1}^{n} \sum_{j=1}^{m} d_{ij}^2 = \sum_{i=1}^{n} \gamma_i,$$  \hspace{1cm} (12)

and

$$\sum_{i=1}^{n} \sum_{j=1}^{m} d_{ij}'^2 = \gamma_1,$$

the mean square error between the approximate matrix $D'$ and the true matrix $D$ is determined by the eigenvalues as

$$e = \gamma_2 + \gamma_3 + \ldots \gamma_n.$$  \hspace{1cm} (13)

We define an index of separability $\alpha$ as the relative energy share of $D'$

$$\alpha = \frac{\gamma_1}{\gamma_1 + \gamma_2 + \ldots \gamma_n}.$$  \hspace{1cm} (14)

Since $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n \geq 0$, $\alpha$ will range from $1/n$ for the most inseparable spectrum to 1 for a completely separable spectrum. The eigenvalues represent the energy carried by each term of the expansion in (9). The index of separability $\alpha$ is simply the fraction of the total energy carried by the first and largest term in the expansion, which is the term that constitutes the best separable approximation.

We applied the singular value decomposition to the spatiotemporal spectra by considering each spectrum as a matrix $P$ of dimension $33 \times 32$. As shown in (9), $P$ can be expanded as

$$P = \sum_{i=1}^{32} \sqrt{\gamma_i} t_i s_i,$$  \hspace{1cm} (15)

where $s_i$ are new orthonormal row vectors representing spatial spectra, and $t_i$ are orthonormal column vectors representing temporal
spectra in each term of the sum. A separable approximation of the form

$$P' = \sqrt{\sum_t s_t^\alpha s_i^\beta}$$  \hspace{1cm} (16)$$

exists where $s_t$ and $s_i$ represent the spatial and temporal components of the separable approximation. The normalized energy share of this term is $\alpha$, which is the index of separability. Examination of $\alpha$ for the spatiotemporal spectra of the 14 image sequences (Table I) shows that for 13 out of the 14 sequences, $\alpha > 0.993$, which constitutes a high degree of separability [10]. Although the separability was low for one sequence, $\alpha = 0.982$. This suggests that a spacetime separable model such as (8) may adequately describe the spatiotemporal spectrum of image sequences since the assumption of separability is valid. The extraction of nearly all the energy with the separable term is also significant for perceptual reasons since small fractions of image energy can markedly affect the perception of some images [12].

V. Calculation of Model Parameters

Since the spatiotemporal spectra of the image sequence $P$ are all highly separable, we need only determine whether the model of (8) adequately characterizes the frequency distribution of the spectra and find the spatial and temporal parameters $a$ and $b$. This will determine whether the commonly used model defined by a separable exponential autocorrelation in space and time is satisfactory.

We find the model parameters $a$ and $b$ by minimizing the mean squared error between the actual signal spectra $P$ of (2) and the analytical separable model of (8).

$$\min \left[ (P - P(k, f))^2 \right]$$ \hspace{1cm} (17)$$

The optimal parameters $a, b$ for each of the sequences were calculated using the Nelder-Mead simplex algorithm [13]. The mean squared error between the analytical separable model (8) and the true spectrum, which was expressed as a percentage of the average squared power of the spectrum, is small (0.03% < mSE < 4.7%) and is given in Table I. The parameters $a$ and $b$ determine the effective bandwidth for the spatiotemporal power spectrum. Fig. 2 illustrates the relationship between the parameters $a$ and $b$ for all 14 sequences, and thus, the simultaneous spatial and temporal bandwidths. All of the pairs of $a$ and $b$ are located within a well-defined range for this ensemble such that no sequence contains both high spatial and high temporal frequencies.

The separable kernel in the model of (8) is based on theoretical considerations, mainly, statistical properties of Markov processes as models for image signals. It is interesting to investigate how this theoretical separable model captures the functional shape of the spectra in spatial and temporal frequency compared with the empirically derived separable kernels derived by the singular value decomposition. The empirically derived kernels are not constrained by a predetermined functional shape as is the theoretical model. We compare the spatial and temporal components of the analytical separable model to the corresponding components of the separable approximation (16). Four examples are shown in Figs. 3 and 4. The model provides a good fit for the sample signal spectra in all frequency ranges. (Note that the ordinate scale is logarithmic, and therefore, the contribution to the mean squared error is small at high frequencies.) This finding is consistent with the applicability of the models of (6) and (7) in earlier studies of spatial and temporal statistics [2], [5], [7]-[9].
VI. DISCUSSION

We calculated the spatiotemporal power spectra of 14 image sequences to investigate whether these spectra are separable in space and time. Using a normalized index of separability, we show that a separable approximation for the spectra derived from the singular value decomposition extracts over 98% of the signal energy (Table I). We also investigated whether the space-time separable exponential model commonly used in the literature provides a reasonable description of the statistics of image sequences. This exponential model is equivalent to the space-time separable power spectrum model of (8). We show that this model provides a good analytical description of the spectrum of image sequences.

For this ensemble of image sequences, no sequence possessed both high spatial and high temporal frequencies (Fig. 2). This property may be a result of spatial blurring caused by motion. If so, it is not an inherent property of the image sequence but rather is caused by the low-pass temporal filtering of the camera. The visual system also temporally low-pass filters images (mainly due to photoreceptor integration time); therefore, this property holds true for a signal perceived by the visual system as well. This limitation on signal spatiotemporal bandwidth may be useful for perceptually based image coding and processing applications [14].

Applications of the model to image processing accrues both the advantages and limitations of using autocorrelation and power spectrum methods. As descriptions of images, the autocorrelation and power spectra are global in the sense that they represent a calculation averaged over the entire image or image sequence. This averaging does not retain the phase spectrum of images and removes local nonstationarities and, hence, specific local details of images. In addition, the separable model may not apply to local sections of image sequences even though the global spectrum of the sequence is separable. In those cases where the autocorrelation and power spectrum methods are applicable, the assumption of separability enables considerable mathematical simplicity. Any methods of image processing developed for spatial-only or temporal-only processing using (6) and (7) can be extended in a straightforward manner to spatiotemporal processing with (8).

REFERENCES

Group 2: Signal propagation in the retina


Signal sampling and propagation through multiple cell layers in the retina: modeling and analysis with multirate filtering

Bennett Levitan and Gershon Buchsbaum

Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, 220 South 33rd Street, Philadelphia, Pennsylvania 19104-6392

Received May 27, 1992; revised manuscript received January 5, 1993; accepted January 11, 1993

The retina is a multilayered structure. Each layer consists of one or more classes of cell, each at its own density and with its own anatomic and physiologic properties. Signals converge from many cells in one layer onto single cells in another layer, and a signal from a single cell diverges to many cells in the next layer. In this methods paper we develop a general approach to retinal analysis and modeling that incorporates multiple cell classes, their densities, and related anatomic properties. The method is based on multirate filtering, a branch of signal processing in which signals of different sampling rates are manipulated. By drawing a correspondence between cell density and signal sampling rate, we define multirate models that incorporate different cell densities, convergence, divergence, variation in dendritic field shape, cell-to-cell variation in synaptic weights, and other anatomic features. We develop the multirate approach and apply it to the cat cone \( \Rightarrow \) cone bipolar CBb, \( \Rightarrow \) on-\( \beta \) ganglion cell pathway as an example. We calculate the spatial frequency responses of the CBb and on-\( \beta \) cells based on the cone spatial frequency response and find that the attenuation of high frequencies in the cones prevents aliasing that would otherwise occur in CBb and on-\( \beta \) cells. We compare the calculations with cat psychophysics. We show that the optics of the cat eye are insufficient in themselves for the prevention of aliasing in these cells; additional attenuation by the cone–cone gap junctions and the cone aperture is necessary. By including this postreceptoral filtering, we demonstrate that the highest spatial frequency that can be passed by the retina without aliasing is determined not only by the densities of cones, bipolar cells, and ganglion cells but also by the synaptic and the dendritic weighting between these cells.

1. INTRODUCTION

The retina is a multilayered structure. Each layer consists of one or more classes of cell, each at its own density and with its own anatomic and physiologic properties. Signals converge from many cells in one layer onto single cells in another layer, and a signal from a single cell diverges to many cells in the next layer (Fig. 1). Recent studies of the cat retina have measured the detailed anatomic properties that are necessary to model this information flow for several classes of cells. Wässle et al. measured the convergences and the divergences between photoreceptors and type A and B horizontal cells.\(^1\)\(^2\) Cohen and Sterling distinguished and modeled several classes of cone bipolar, their densities, and their convergences and divergences to cones and on-\( \beta \) ganglion cells.\(^3\)\(^4\) Other applications that model retinal processing and coding and use actual convergences, divergences, and number of synapses can be found in Refs. 7–9. If explicit attention is paid to these and other properties, these models could accurately incorporate many anatomic details such as dendritic field shape and the number of synapses between cells. All these properties are highly dependent on the different densities of cell classes.

The purpose of this paper is to describe a method of retinal modeling that generalizes the multiple cell layer approach taken in these studies of the cat retina. Our motivation is to incorporate detailed anatomic properties into multilayered retinal models that can be easily analyzed. The method is based on a branch of signal processing known as multirate filtering. Multirate filtering concerns the manipulation, the filtering, and the analysis of signals in systems whose signals are not all at the same sampling rate. By drawing a correspondence between the density of an array of cells and the sampling rate of a signal, we can incorporate into retinal models different cell densities, convergence, divergence, variation in dendritic field shape, cell-to-cell variation in synaptic weights, and other anatomic features. Multirate models are easily manipulated analytically in both space and frequency domains, have computationally efficient implementations, and allow for closed-form solutions.\(^10\)\(^12\) Multirate filtering permits an examination of the artifacts that result from a change in a signal's sampling rate. For this reason, it is particularly useful for the analysis of effects such as aliasing in a multilayered structure with different densities, such as the retina, and for the modeling of postreceptoral filtering. Some of the retinal properties that can be derived from this approach, such as the trade-offs in properties among parallel, hierarchic, and hybrid architectures, are reported in an early form in Refs. 13–15. Among the benefits of the multirate approach used here is that the models display some of the variation and the irregularity seen in retinal anatomy.

In Subsection 2.A of this paper we develop the method of modeling with multirate filtering. In Subsection 2.B we apply this technique to the cat cone \( \Rightarrow \) cone bipolar CBb, \( \Rightarrow \) on-\( \beta \) ganglion cell pathway and demonstrate some of the types of analysis permitted by multirate filtering. We show how spatial aliasing is prevented in this path-
corresponds to a different B cell (Fig. 2b). The image
signals have different spatial variables because the \( n_{th} \)
sample in \( x \) does not correspond to the same location
as that of the \( n_{th} \) sample in \( y \), except at the origin
(Fig. 2b). We address the issue of irregular sampling in
the retina in Section 3.

To an excellent approximation, at a single adaptive
state, many retinal cells sum their inputs linearly over
both space and time.\(^{16-19}\) For these cells the voltage pro-
duced in a B cell can be viewed as a weighted sum of the
voltages in the presynaptic A cells, where the weights are
determined by the number and the location of synapses,
the details of the dendritic tree, the types of neurotrans-
mittor, and the types of receptor. Since the cells sum
linearly, one can simulate the operation of the synapses by
convolving \( x(n) \) with a filter of the appropriate weights.
This approach is taken in many signal-processing-based
models. However, since \( x(n) \) and \( y(m) \) are at different
sampling rates, standard convolution is inadequate, as
standard convolution assumes the same sampling rates in
all the layers. Convolution between signals with different
sampling rates requires multirate filtering. The use of
multirate filtering allows for different sampling rates
(densities) in different layers and hence arbitrary conver-
gence/divergence ratios (see Subsection 2A.1b). We now
summarize the relevant basics of multirate filtering and
use them to develop the model.

The basic operations of multirate filtering are upsam-
pling and downsampling.\(^{10-12}\) Operating on a signal \( x(n) \),
an \( L \)-fold upsampler inserts \( L - 1 \) zero-valued samples be-
 tween adjacent samples in \( x(n) \) and decreases the sam-
ping period of \( x(n) \) by a factor of \( L \). Combined with the ap-
propriate low-pass filtering, the upsampler increases the sam-
ping rate of \( x(n) \) by a factor of \( L \) while maintaining the
signal’s form (Fig. 3(a)). Operating on \( x(n) \), an \( M \)-fold
downsampler removes \( M - 1 \) of every \( M \) samples in \( x(n) \)
and increases the sampling period of \( x(n) \) by a factor of \( M \).
Combined with the appropriate low-pass filtering, a downsam-
pler decreases the sampling rate of \( x(n) \) by a factor of \( M \)
while maintaining the exact form of the signal (Fig. 3(b)).
Of importance is that, without the appropriate filtering,
either resampler may induce aliasing (technically called
imaging for the upsampler).\(^{10-12}\) The insets in Fig. 3 de-
 monstrate the anatomic situations that correspond to these
resampling operations.

Technically, one can avoid the resampling operations
in the model by setting both \( T_1 \) and \( T_2 \) to the greatest
common divisor of \( n_{nd}A \) and \( n_{nd}B \). This small sampling
period ensures that both \( x(n) \) and \( y(m) \) have samples at the
locations that correspond to every cell of classes A and B.
However, this choice also results in artifactual filler

### Table 1. Anatomic Parameters and Their Analogs in a Multirate Model

<table>
<thead>
<tr>
<th>Anatomic Parameter</th>
<th>Multirate Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nearest-neighbor distance ( n_{nd} )</td>
<td>Sampling period ( T )</td>
</tr>
<tr>
<td>Convergence ( C )</td>
<td>Model convergence ( C_x )</td>
</tr>
<tr>
<td>Divergence ( D )</td>
<td>Model divergence ( D_x )</td>
</tr>
<tr>
<td>Changes in cell densities</td>
<td>Downsampling/upsampling by ( M/L )</td>
</tr>
</tbody>
</table>
| Synaptic/dendritic weighting func-
  tions                            | Direct-form filter \( h(n) \)  |
| Dendritic field radius             | Circular filter radius \( R_{cc} \)  |
| Dendritic fields and weights       | Space-varying filters \( g_{in} \)  |
| Neurotransmitter/receptor gain      | Filter gain \( K \)            |
and $L$ that satisfy Eq. (4). There is often flexibility in the densities for the one-dimensional case. The actual values show in Subsection 2.A.2 that such a combined operation three steps can be combined into a single-step operation in terms of the anatomic Equation (4) gives the ratio $M/L$.

To derive $M$ and $L$ from anatomic parameters, we combine Eqs. (1)-(3) to obtain

$$\frac{M}{L} = \frac{T_i}{T_y} = \frac{\text{nnd}_a}{\text{nnd}_b} = \frac{d_a}{d_b}. \tag{4}$$

Equation (4) gives the ratio $M/L$ in terms of the anatomic densities for the one-dimensional case. The actual values for $M$ and $L$ are given by the smallest pair of integers $M$ and $L$ that satisfy Eq. (4). There is often flexibility in the exact choice of $M$ and $L$, since Eq. (4) can be exactly satisfied only when $d_a/d_b$ is a rational number. Provided that $M$ and $L$ are relatively prime, increasing $M$ and $L$ allows $M/L$ to approximate a given $d_a/d_b$ better and causes the sampling rates $T_i$ and $T_y$ to match nda and ndn more closely. We discuss other effects of an increase in $M$ and $L$ when we consider the two-dimensional case in Subsection 2.A.3.

In Fig. 3(c) $h(u)$ is a discrete spatial filter with $N_x$ samples that emulates the synaptic and dendritic weighting. In deriving $h(u)$ from anatomic parameters, one must consider three aspects of the filter: (a) functional form, (b) spatial extent, and (c) gain. We now discuss the relevant anatomy and its correlate in the model for each aspect.

a. Functional Form In Fig. 2(a) the number of offshoots on the terminal branches of an A cell indicates the number of synapses that it provides to the B cells beneath it. The number of these synapses is characterized by the synaptic weighting function (swf). Calculated from retinal data, $\text{swf}(x, y)$ gives the average number of synapses between a B cell and an A cell that synapses at location $(x, y)$ relative to the center of the B cell's dendritic field. Figure 4(a) shows the $\text{swf}$ for the one-dimensional array of cells in Fig. 2(a). In this example $\text{swf}(r) = 4 - 0.5|r|$, where $r$ is in micrometers. The change in voltage induced in a B cell by the A cells presynaptic to it is characterized by the dendritic weighting function (dwf). $\text{dwf}(x, y)$ gives the average change in voltage in the soma of a B cell induced by a unit injection of current from a presynaptic A cell that synapses at location $(x, y)$ relative to the center of the B cell's dendritic field. The $\text{dwf}$ reflects branching in the dendrites, dendritic diameters, and membrane and cytoplasmic characteristics. Though the $\text{swf}$ and the $\text{d wf}$ may be of any form, in practice they are often assumed to be on average circularly symmetric, such that $\text{swf}(x, y)$ and $\text{d wf}(x, y)$ are functions of the radius $r = (x^2 + y^2)^{1/2}$. With this assumption the weight between a presynaptic and a postsynaptic cell is a function of the distance between them.

Assuming that all synapses of the same type contribute equally to the postsynaptic potential, the functional form of $h(u)$ is proportional to both the $\text{swf}$ and the $\text{d wf}$:

$$h(u) = \text{d wf}(uT_x/L)\text{swf}(uT_y/L). \tag{5}$$

where $uT_x/L$ is the continuous distance that corresponds to discrete distance $u$ at the filter's sampling rate. If we set $\text{d wf}(r) = 1$, the functional form of the filter that corresponds to the $\text{swf}$ in Fig. 4(a) is given by $h(u) = 4 - |u|$ [Fig. 4(b)].

By means of several mechanisms retinal cells adapt to the signals that they carry. They change their overall gains, alter the shape of their receptive fields, and adjust their chromatic and temporal properties according to present and past input. The spatial components of these changes manifest themselves in the $\text{d wf}$'s and the overall gains of the cells. In contrast, $\text{swf}$'s are relatively constant. Our development below is for one level of adaptation, so the gains and the $\text{d wf}$'s remain constant. One can study different adaptive states by changing these components to match the different states or by using adaptive filters or nonlinearities.

The resampling operations in Fig. 3 are the basic mathematical tools necessary for defining and manipulating multirate systems. There are several means by which one can implement these operations; for example, the three steps can be combined into a single-step operation that does not explicitly insert or remove samples. We show in Subsection 2.A.2 that such a combined operation is analogous to the synaptic weighting in Fig. 2(a).

To derive $M$ and $L$ from anatomic parameters, we combine Eqs. (1)-(3) to obtain

$$\frac{M}{L} = \frac{T_i}{T_y} = \frac{\text{nnd}_a}{\text{nnd}_b} = \frac{d_a}{d_b}. \tag{4}$$

Equation (4) gives the ratio $M/L$ in terms of the anatomic densities for the one-dimensional case. The actual values for $M$ and $L$ are given by the smallest pair of integers $M$ and $L$ that satisfy Eq. (4). There is often flexibility in the exact choice of $M$ and $L$, since Eq. (4) can be exactly satisfied only when $d_a/d_b$ is a rational number. Provided that $M$ and $L$ are relatively prime, increasing $M$ and $L$ allows $M/L$ to approximate a given $d_a/d_b$ better and causes the sampling rates $T_i$ and $T_y$ to match nda and ndn more closely. We discuss other effects of an increase in $M$ and $L$ when we consider the two-dimensional case in Subsection 2.A.3.

In Fig. 3(c) $h(u)$ is a discrete spatial filter with $N_x$ samples that emulates the synaptic and dendritic weighting. In deriving $h(u)$ from anatomic parameters, one must consider three aspects of the filter: (a) functional form, (b) spatial extent, and (c) gain. We now discuss the relevant anatomy and its correlate in the model for each aspect.

a. Functional Form In Fig. 2(a) the number of offshoots on the terminal branches of an A cell indicates the number of synapses that it provides to the B cells beneath it. The number of these synapses is characterized by the synaptic weighting function (swf). Calculated from retinal data, $\text{swf}(x, y)$ gives the average number of synapses between a B cell and an A cell that synapses at location $(x, y)$ relative to the center of the B cell's dendritic field. Figure 4(a) shows the $\text{swf}$ for the one-dimensional array of cells in Fig. 2(a). In this example $\text{swf}(r) = 4 - 0.5|r|$, where $r$ is in micrometers. The change in voltage induced in a B cell by the A cells presynaptic to it is characterized by the dendritic weighting function (dwf). $\text{d wf}(x, y)$ gives the average change in voltage in the soma of a B cell induced by a unit injection of current from a presynaptic A cell that synapses at location $(x, y)$ relative to the center of the B cell's dendritic field. The $\text{d wf}$ reflects branching in the dendrites, dendritic diameters, and membrane and cytoplasmic characteristics. Though the $\text{swf}$ and the $\text{d wf}$ may be of any form, in practice they are often assumed to be on average circularly symmetric, such that $\text{swf}(x, y)$ and $\text{d wf}(x, y)$ are functions of the radius $r = (x^2 + y^2)^{1/2}$. With this assumption the weight between a presynaptic and a postsynaptic cell is a function of the distance between them.

Assuming that all synapses of the same type contribute equally to the postsynaptic potential, the functional form of $h(u)$ is proportional to both the $\text{swf}$ and the $\text{d wf}$:

$$h(u) = \text{d wf}(uT_x/L)\text{swf}(uT_y/L). \tag{5}$$

where $uT_x/L$ is the continuous distance that corresponds to discrete distance $u$ at the filter's sampling rate. If we set $\text{d wf}(r) = 1$, the functional form of the filter that corresponds to the $\text{swf}$ in Fig. 4(a) is given by $h(u) = 4 - |u|$ [Fig. 4(b)].

By means of several mechanisms retinal cells adapt to the signals that they carry. They change their overall gains, alter the shape of their receptive fields, and adjust their chromatic and temporal properties according to present and past input. The spatial components of these changes manifest themselves in the $\text{d wf}$'s and the overall gains of the cells. In contrast, $\text{swf}$'s are relatively constant. Our development below is for one level of adaptation, so the gains and the $\text{d wf}$'s remain constant. One can study different adaptive states by changing these components to match the different states or by using adaptive filters or nonlinearities.
b. Spatial Extent The number of class A cells that synapse on a class B cell varies from B cell to B cell. Similarly, the number of B cells to which an A cell is presynaptic varies from A cell to A cell. The average number of A cells that synapse on a single B cell is the convergence (C) (Fig. 1). In the early layers of vision, convergence reflects the degree to which a cell gathers information over the visual field. The average number of B cells on which a single A cell synapses is the divergence (D) (Fig. 1). Divergence is one of the mechanisms by which a point of visual information spreads laterally among the cells. In

Fig. 2(a) $C = 2.5$ and $D = 1.667$. Clearly, convergence and divergence are related. Freed et al. have shown that, for two arrays of cells as in Fig. 2(a),

$$\frac{C}{D} = \frac{d_A}{d_B}.$$

The filter's spatial extent $N_u$ is set so that the convergence and the divergence in the model match those of the anatomy. After the upsampler in Fig. 3(c) $L - 1$ of every $L$ samples in $v(u)$ are zero. In the model the convergence is the average number of samples from $x(n)$ that $h(u)$
convergence more closely and the model's matching

When Eqs. (4) and (6) are not satisfied exactly, as is typically the case given the precision of anatomic measurements, no value of \( N_s \) yields both \( C_n = C \) and \( D_n = D \). As is shown in Subsection 2.B below, varying \( N_s \) results in trade-offs between the model's matching the anatomic convergence more closely and the model's matching the anatomic divergence more closely.

c. Gain. The gain of \( h(u) \) is represented by the coefficient \( K \) in Eq. (5). The gain depends on the types of neurotransmitter and receptor in the synapse. When only one pair of cells is being modeled, \( K \) can generally be ignored except for its sign—excitatory synapses require positive \( K \), and inhibitory synapses require negative \( K \). When a model incorporates several sets of cells in parallel, the relative gains of the filters must be included. If it is intended that the values of \( x(n) \) and \( y(m) \) match the voltages in A and B cells, respectively, then we can set \( K \) by comparing the presynaptic and postsynaptic voltages for a spatially constant input to the A cells, using

\[
K = \frac{L^2 \times (\text{voltage in B cells})/(\text{voltage in A cells})}{\sum [\text{dwc}(uT/L)\text{wsc}(uT/L)]}
\]  

where the summation normalizes for the dc component of the convolved dwc and wsc. If we intend that the values of \( x(n) \) and \( y(m) \) relate only linearly to voltages in A and B cells, respectively, then Eq. (9) can be used or the \( K \) 's for the parallel filters can be made proportional to the total
Figure 5 demonstrates the space-varying filter implementation of Fig. 3(c) for "L = 2, M = 3, and filter h(u) with five weights. The model shown here corresponds to the synapses and the signals in Fig. 2. The samples in x(n) that compute a given sample in y(m) correspond to the class A cells that synapse on the corresponding class B cell. Not every B cell has the same convergence, just as not every filter g_m(n) has the same size. The anatomical convergence C is the average convergence for all the B cells, and the model convergence C_m is the average size of all the filters g_m(n). Different cells receive different number of synapses, and different space-varying filters have different weight values. We discuss this multirate variation more fully in Subsection 2.2.3.
convergence and divergence are given by both the swf and the dwf. Extending Eq. (5) to the two-integers that satisfy Eq. (12).

The graticules show which samples are selected along each axis, are square or rectangular in shape, and have an average of \( \frac{C}{2} \) samples. The number of samples along each axis is either the smallest integer greater than \( \frac{N_i}{L} \) or the largest integer less than \( \frac{N_i}{L} \). For example, the four space-varying filters that result from a 9 \times 9-weight \( h(u_1, u_2) \) for \( L = 2 \) and any \( M \) are \( 5 \times 5, 5 \times 4, 4 \times 5, \) and \( 4 \times 4 \) weights (Figs. 6(a) and 6(b)). Their average number of weights is 20.5, which matches the convergence \( C_m \), as one would expect.

To obtain space-varying filters that match the shapes of dendritic fields more closely (Fig. 10(b) below), we can make the direct-form filter circular in shape (Fig. 6(c)). This approach gives space-varying filters that are on average circular (Fig. 6(d)) instead of rectangular (Fig. 6(b)).

The circular filter boundary is defined by

\[
h(u_1, u_2) = 0 \quad \text{for} \quad (u_1^2 + u_2^2)^{1/2} > R_{cire},
\]

where we set the radius \( R_{cire} \) for the best match between \( C_m \) and \( C \) and between \( D_m \) and \( D \). For circular filters the model convergence and divergence are given by

\[
C_m = \text{(number of nonzero weights in } h)/L^2, \quad (18)
\]

\[
D_m = \text{(number of nonzero weights in } h)/M^2, \quad (19)
\]

respectively. The two direct-form filters in Fig. 6 have the same convergence, but the circular filter is slightly larger in compensation for the zero-valued weights.

When \( (d_x/d_y)^{1/2} \) is irrational, we can set \( M \) and \( L \) to satisfy Eq. (12) with an arbitrary degree of accuracy. The consequences of a larger \( L \) are a larger filter size \( N_x \times N_y \) and a greater variety of space-varying filters \( g_{n_1, n_2} \). If \( h(u_1, u_2) \) is square, the \( g_{n_1, n_2} \) will all differ in their weights but will never have more than four different sizes and shapes. If \( h(u_1, u_2) \) is circular, the \( g_{n_1, n_2} \) will be of a variety of shapes and sizes depending on locations of zero-valued weights in \( h(u_1, u_2) \). Thus changing \( M \) and \( L \) while keeping \( M/L \) approximately constant provides a means to improve the matches between the ranges of convergence and divergence in the model and those in anatomy.

Other sources of flexibility in the matching of variation in the model with that in anatomy arise when one ensures that the space-varying filters completely tile the presynaptic layer. Tiling refers to the degree to which filters touch every sample in their two-dimensional input array. The \( L^2 \) filters, constrained to match average anatomic properties, in some cases need additional variation to touch every input sample. Incomplete tiling corresponds to a situation in which presynaptic cells are not synapsed on by any postsynaptic cells. We demonstrate incomplete tiling in Subsection 2.B. There are at least three methods that introduce the variation needed for complete tiling:
Cone => CBb,

arbitrary test signal. Other cells such as CBb2-CBb4 and cone array, which in our analysis is either the signal shown in Fig. 7. The model's input is the output of the involving CBb, at 1° eccentricity. The two-stage model is 5

The cat cone => cone bipolar => on-/β ganglion cell pathway

Bipolar b, => on-/β Ganglion Cell Pathway

Application: Multirate Model of Cat Cone => Cone Bipolar

Deriving the Model from Anatomy

1. Deriving the Model from Anatomy

The cat cone => cone bipolar => on-β ganglion cell pathway is an ideal system for examination with multirate filtering, since much of its anatomy is known in great detail. The cone bipolar is of five morphological types, CBb1-CBb5, of which CBb1 contributes more than half of the cone bipolar input to the on-β cell.4 Here we model the pathway involving CBb1 at 1° eccentricity. The two-stage model is shown in Fig. 7. The model's input is the output of the cone array, which in our analysis is either the signal formed by cones, horizontal cells, and the eye's optics or an arbitrary test signal. Other cells such as CBb2-CBb4 and amacines also transmit signals from cones to on-β cells. These signals add to or may alter the CBb1 signal after it has reached the on-β cells, but there is no difficulty in analyzing the CBb1 path separately. Horizontal cells are partly responsible for generating the cone signal but do not have any influence on the signal once it has entered the bipolar cells. For clarity we will refer to the resampling factors in the model as M and L without subscripts, using the context to identify the particular synapse.

The densities for these cells are listed in Table 2. For the two synaptic steps involved, Table 3 lists the anatomic convergences, divergences, C/D ratios, and density ratios. Since the C/D ratios are not especially close to the density ratios, we use the more reliable density measurements to set M/L. Then find the circular filter sizes that best match the model to the anatomic convergences and divergences.

| Table 2. Densities and Nearest-Neighbor Distances of Cones, CBb, Cells, and on-β Cells at 1° Eccentricitya |
|---|---|---|
| Cell Class | Density (mm−2) | Nearest-Neighbor Distance (µm) |
| Cones | 24,200 | 5.2 |
| CBb, | 6,100 | 10.2 |
| On-β | 1,860 | 25 (est.) |

*aRefs. 4 and 30.

(1) Vary the M/L ratio within the anatomic range for measurements of \((d_x/d_y)^{1/2}\). If we change L and hence the number and the variety of space-varying filters, tiling may become more nearly complete. While it is easy to implement, because of the limited range of \((d_x/d_y)^{1/2}\) this method may not work in all cases.

(2) Permit the direct-form filter to cycle among several asymmetric weighting functions that are on average symmetric. The average symmetry ensures that the filters act isotropically. Methods (1) and (2) increase the range of nonzero weights (convergences) of the space-varying filters and can make the filters better mimic the variability of retinal dendritic fields. Method (2) requires performing a different analysis for each direct-form filter used.

(3) Use different patterns of sampling in different layers, such as rectangular and hexagonal.

We discuss examples of all these methods in Subsection 2.B.

B. Application: Multirate Model of Cat Cone => Cone Bipolar

1. Deriving the Model from Anatomy

The cat cone => cone bipolar => on-β ganglion cell pathway is an ideal system for examination with multirate filtering, since much of its anatomy is known in great detail. The cone bipolar is of five morphological types, CBb1-CBb5, of which CBb1 contributes more than half of the cone bipolar input to the on-β cell.4 Here we model the pathway involving CBb1 at 1° eccentricity. The two-stage model is shown in Fig. 7. The model's input is the output of the cone array, which in our analysis is either the signal formed by cones, horizontal cells, and the eye's optics or an arbitrary test signal. Other cells such as CBb2-CBb4 and amacines also transmit signals from cones to on-β cells. These signals add to or may alter the CBb1 signal after it has reached the on-β cells, but there is no difficulty in analyzing the CBb1 path separately. Horizontal cells are partly responsible for generating the cone signal but do not have any influence on the signal once it has entered the bipolar cells. For clarity we will refer to the resampling factors in the model as M and L without subscripts, using the context to identify the particular synapse.

The densities for these cells are listed in Table 2. For the two synaptic steps involved, Table 3 lists the anatomic convergences, divergences, C/D ratios, and density ratios. Since the C/D ratios are not especially close to the density ratios, we use the more reliable density measurements to set M/L. Then find the circular filter sizes that best match the model to the anatomic convergences and divergences.

a. Cone => CBb, Synapse For the cone => CBb, synapse (left-hand side of Fig. 7), Eq. (12) gives \(M/L = \sqrt{3.97} \) for which \(M = 2 \) and \(L = 1 \) is a close match. Since \(L = 1 \), there is only one space-varying filter, which is identical to the direct-filter form. Setting \(R_{nn} = 1 \) results in a filter with five nonzero samples in a plus-shaped \((-1)\) pattern. Equations (18) and (19) yield \(C_m = 5 \) and \(D_m = 1.25 \), which give a close match between model and anatomic convergences and divergences. For this synapse observation of the anatomic convergence indicates that it ranges between 4 and 7, with an average of 5.1.2 Because there is only one space-varying filter, convergence in the model does not vary from 5.0. The anatomy has more variation than does the model for this synapse, though the choice of other values for \(M \) and \(L \) could provide variation in the model's convergence while still matching the C/D ratio.

This synapse also provides an example of the tiling issue discussed in Subsection 2A.3. As Fig. 8(a) shows, a plus-shaped \((-1)\) symmetric filter with five weights and the ratio \(M/L = 2/1 \) does not completely tile the cone layer. One of four cone samples is missed. We do not demonstrate it here, but varying \(M \) and \(L \) does not complete tiling for this synapse. Values of \(M/L \) close to \(\sqrt{3.97} \) [Eq. (12)] yield tilings that are better than those provided by \(M/L = 2/1 \), but they are still incomplete. One solution that results in complete tiling is to cycle between two direct-form

| Table 3. Anatomic Properties for the Cone => CBb, and CBb, => on-β Synapses at 1° Eccentricitya |
|---|---|---|
| Synapse | Convergence | Divergence | Convergence/Divergence | Density Ratio | Convergence Range |
| Cone => CBb, | 5.1 | 1.2 | 4.25 | 3.97 | 6-7 |
| CBb, => on-β | 6-7 | 3 | 2-2.33 | 3.28 | 6-7 |

*aRefs. 3, 5, and 6.

The convergence range refers to the observed minimum and maximum number of presynaptic cells that converge on a postsynaptic cell for that synapse.
filters, one plus shaped and one cross shaped \((\times)\) \[Fig. 8(b)\]. These filters are rotated and radially scaled versions of each other. Since \(L = 1\), the space-varying filters are identical to the direct-form filters. Figure 8(c) shows a solution that introduces more variability in filter shape. Here we cycle among four different filters that are on average symmetric. Many other combinations of filters are possible. Another solution is to change the sampling in the on-\(\beta\) array to a slightly compressed hexagonal grid [Fig. 8(d)]. In this case a single plus-shaped filter touches every cone. These solutions show that varying from a single direct-form filter based on anatomic averages is sometimes needed for complete tiling of a presynaptic layer. In the frequency-domain analysis in Subsection 2.B.2 we use the tiling solution shown in Fig. 8(b).

Neither the swf nor the dwf is known for the cone \(\Rightarrow\) CBb, synapse. Since the CBb, dendritic field is so narrow, we assume that both functions are constant. From Eqs. (13) and (17), the plus-shaped and cross-shaped direct-form filters are

\[
h_{\text{cone}=\text{CBb},1}(u_1, u_2) = \begin{cases} K & u_1^2 - u_2^2 \leq 1, \\ 0 & \text{otherwise} \end{cases}
\]

\[
h_{\text{cone}=\text{CBb},2}(u_1, u_2) = \begin{cases} K & u_1 = 1 \text{ and } u_2 = 1, \\ 0 & \text{otherwise} \end{cases}
\]

respectively. These filters are shown in the insets of Fig. 9(a).

b. CBb, \(\Rightarrow\) on-\(\beta\) Synapse For the CBb, \(\Rightarrow\) on-\(\beta\) synapse (right-hand side of Fig. 7), Eq. (12) gives \(M/L = \sqrt{3.28}\), for which \(M = 16\) and \(L = 9\) match well. Because of the large difference in the \(C/D\) and density ratios, no value of \(R_{\text{con}}\) permits an exact numerical match of the model to both the anatomic convergence and the anatomic divergence. A value of \(R_{\text{con}}\) that gives a \(C_{\text{m}}\) within the range of \(C\) also gives a \(D_{\text{m}}\) that is significantly less than \(D\). While an increase in \(R_{\text{con}}\) will increase the model's divergence, it will also bring the model's convergence out of the anatomic range 6-7. We set \(R_{\text{con}} = 13.35\), which yields a direct-form filter with 561 nonzero weights, a \(C_{\text{m}}\) of 6.93, and a \(D_{\text{m}}\) of 2.19. These values also permit complete tiling of the CBb, array by the space-varying filters. While the observed convergence for this synapse ranges from 6 to 7, the space-varying filters have between four and nine weights. For this synapse the model has more variation than does the anatomy. However, since the observed range is based on only three on-\(\beta\) cells, it is likely that the actual anatomic convergence varies more widely than the range 6-7. Because the \(M/L\) ratios do not exactly match the density ratios for either synapse, the model sampling periods are not identical to the anatomic nnd's. We set \(T_{\text{cone}} = \text{nnd}_{\text{cone}} = 6.43\ \mu\text{m}\). By Eq. (3), \(T_{\text{CBb},1} = 12.8\ \mu\text{m}\) and \(T_{\text{cone}} = 23.2\ \mu\text{m}\). These values and the corresponding model densities are listed in Table 4.

Smith and Sterling argued that electrotonic decay along on-\(\beta\) cell dendrites at \(1^\circ\) eccentricity is insignificant, and therefore the dwf for the CBb, \(\Rightarrow\) on-\(\beta\) synapse is constant. The swf is not known in detail, but anatomic measurements show that CBb, cells near the middle of the on-\(\beta\) dendritic field tend to give many contacts (12-33) and that cells near the edge of the field tend to give few contacts (3-4).\(^3\) The radius of the on-\(\beta\) cell's dendritic field is approximately 20 \(\mu\text{m}\). Since, for this synapse, the variation in number of synaptic contacts is represented

\[
\begin{align*}
\text{Fig. 8. Demonstration of tiling issue and solutions for the cone }&\Rightarrow\text{ CBb, filter. Filled diamonds indicate the CBb, sample, and small open circles indicate the cone samples touched by the CBb, samples. The contacts are shown explicitly by the lines. The resampling operations give the CBb, array one quarter of the density of the cone array. (a) Cone and CBb, samples and filter as defined in the text. The five-weight, plus-shaped filters touch only three of every four cone samples. (b) Cycling among several direct-form filters that are on average symmetric completely tiles the cone array. Here two filters are used. (c) Same method as in (b) but here the cycling is among four different filters. (d) Changing from rectangular sampling to another sampling scheme (in this case, compressed hexagonal sampling for the CBb, array) permits tiling of the cone array by a single, symmetric filter.}
\end{align*}
\]
Fig. 9. Direct-form filters for the model. Spatial weighting functions are shown in the insets of (a) and on the left-hand sides of (b) and (c), and cross sections of their DTFT's are shown in (a) and on the right-hand sides of (b) and (c). The frequency axes are in units of normalized radian frequency and extend from zero to the Nyquist rate for each filter. The vertical lines show the cutoffs required for preventing aliasing. (a) $h_{cone} = c_{bb2}$ and $A_{cone} = c_{bb2}$ filters, DTFT's are shown along a 0° line ($\omega_1$ axis) and along a 45° line ($\omega_2 = \omega_1$), (b) $h_{cbb2} = \delta_{mod}(u_1, u_2)$, DTFT along a 0° line; (c) rescaled $A_{cbb2} = \delta_{mod}(u_1, u_2)$, DTFT along a 0° line.
Thus \( K = 0.004 \) in Eq. (22).

A constant input gives a constant output of the same intensity. For computational convenience we set the gains so that a computational model can yield an output whose samples cycle among several different intensity levels. As we will see, much of the type of variation in dendritic field size is in the number of synaptic contacts used by the model. As we will show, \( \text{on}-/3 \quad \Rightarrow \quad \text{CBb} \) is not sufficiently low-pass for the prevention of aliasing. One consequence is that the space-varying filters have slightly different zero-frequency gains, and an input consisting of a constant light level will yield an output whose samples cycle among several different but close values. One can easily solve this problem by rescaling the space-varying filters so that the sum of the coefficients in each filter is the same. The rescaled filter is plotted on the left-hand side of Fig. 9(b). This filter is plotted on the left-hand side of Fig. 10(b). A comparison of Figs. 9(b) and 9(c) shows that, for both the space and temporal domains (discussed in Subsection 2.B.2), this rescaling has little effect on the direct-form filters.

Since the two synaptic steps being modeled are in series, the overall gains of the filters can be set arbitrarily. For computational convenience we set the gains so that a constant input gives a constant output of the same intensity. Thus, \( K = 0.2 \) in Eqs. (20) and \( K = 0.004 \) in Eq. (22).

Table 5 summarizes the parameters of this model. As Fig. 10(a) shows, the \( \text{CBb} \Rightarrow \text{on}-/3 \) space-varying filters are of varying shape. In Fig. 10(a) four of the 81 space-varying filters are drawn, where filled diamonds indicate the \( \text{on}-/3 \) samples and small open circles indicate the \( \text{CBb} \) samples touched by the \( \text{on}-/3 \) samples. To compare these filters with the \( \text{on}-/3 \) cell dendritic fields shown in Fig. 10(b), we rescaled the weight values, to show the number of synaptic contacts used by the model. As we can see, much of the type of variation in dendritic field shape and in the number of synaptic contacts is embodied in the multirate model. Of course, retinal anatomic variation is much greater than the variation in the multirate model, but much of the flavor of the anatomy is represented. Note that, though the contacts for each \( \text{CBb} \) cell in Fig. 10(b) are spread over a small area, the model acts as if all the synapses from a \( \text{CBb} \) cell occur at one point. The model does not represent the locations of the individual synapses between a \( \text{CBb} \) cell and an \( \text{on}-/3 \) cell; rather it represents the total number of synapses between these cells.

### Table 4. Densities, Sampling Periods, and Nyquist Rates for Cell Arrays in the Multirate Model

<table>
<thead>
<tr>
<th>Cell Class</th>
<th>Density in Model ((\text{mm}^{-2}))</th>
<th>Sampling Period in Model ((\mu\text{m}))</th>
<th>Nyquist Rates in Model ((\text{cycles/degree}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cones</td>
<td>24,200</td>
<td>6.43</td>
<td>17.1</td>
</tr>
<tr>
<td>CBb</td>
<td>6,050</td>
<td>12.86</td>
<td>8.56</td>
</tr>
<tr>
<td>On-β</td>
<td>1.914</td>
<td>22.9</td>
<td>4.81</td>
</tr>
</tbody>
</table>

### Table 5. Filter and Resampling Parameters in the Multirate Model

<table>
<thead>
<tr>
<th>Synapse</th>
<th>( M )</th>
<th>( L )</th>
<th>( R_{\text{carr}} )</th>
<th>( K )</th>
<th>No. of Nonzero Weights</th>
<th>( C_\text{\textit{e}} )</th>
<th>( D_\text{\textit{e}} )</th>
<th>Convergence Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cone ( \Rightarrow \text{CBb} )</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0.2</td>
<td>5</td>
<td>5.00</td>
<td>1.25</td>
<td>5</td>
</tr>
<tr>
<td>\text{CBb} ( \Rightarrow \text{on}-/3 )</td>
<td>16</td>
<td>9</td>
<td>13.35</td>
<td>0.004</td>
<td>561</td>
<td>6.93</td>
<td>2.19</td>
<td>4-9</td>
</tr>
</tbody>
</table>

2. Analyzing the Model and the Effects of Aliasing

We examine the model primarily in the frequency domain. Because the sampling rate in the model decreases from one array to the next, the Nyquist rate decreases in successive arrays (Table 4). To prevent aliasing, the filters must remove the high frequencies from the signal in one array before they can alias in the next array. In particular, multirate filtering theory shows that direct-form filters in systems like that in Fig. 3(c) must attenuate components above their Nyquist rate/\( \text{max}(M, L) \) to prevent aliasing.\(^{11,12}\) The magnitudes of the discrete-time Fourier transforms\(^{27}\) (DTFT's) of the direct-form filters are shown in Fig. 9(a) and on the right-hand sides of Figs. 9(b) and 9(c). Figure 9(a) shows DTFT cross sections for the plus-shaped and cross-shaped filters along a 0° line \((\omega_1 = 0, \omega_2 = 0)\) and along a 45° line \((\omega_1 = \omega_2)\). Because these filters are simply rotated and scaled versions of each other, their DTFT's are also related by rotation and scaling. The DTFT's in Figs. 9(b) and 9(c) are essentially rotationally symmetric, and we show only their 0° cross sections. While all the filters in Fig. 9 are low-pass, they pass with significant magnitude frequency components above their Nyquist rate/\( \text{max}(M, L) \) cutoffs (shown by the vertical lines in the figure). High-frequency components in the input will alias in both sets of synapses and introduce spurious frequencies. Figure 11 demonstrates examples of this aliasing for the plus-shaped cone \( \Rightarrow \text{CBb} \) filter. In Fig. 11(a) the input temporal \( t_{\text{cone}} \) is a sinusoid of 12.75 cycles/degree. This sinusoid aliases in the \( \text{CBb} \) and \( \text{on}-/3 \) cell arrays to 4.37 cycles/degree. \( h_{\text{cone} \Rightarrow \text{CBb}} \) and \( h_{\text{CBb} \Rightarrow \text{on}-/3} \) attenuate this frequency component, but they are insufficient to prevent the input from being confused with 4.37-cycle/degree input. In Fig. 11(b) \( t_{\text{cone}} \) is a square wave of 6 cycles/degree. The \( \text{CBb} \) image (not plotted for clarity) shows the square wave, but the \( \text{on}-/3 \) image is quite distorted and demonstrates aliasing of several frequency components in the square wave. In the retina irregular sampling would cause this aliasing to manifest itself partly as low-frequency sinusoids and partly as broadband...
noise, both of which lead to significant degradation of the signal.

We characterize the degree of aliasing for all the frequencies with the frequency response functions in Fig. 12 (\( f \) denotes the spatial frequency in units of cycles/degree). In Fig. 12(a) the dashed curve shows the response of CBbi cells to frequencies below the Nyquist rate of the CBb cells (8.56 cycles/degree). Components of higher frequencies are aliased into the 0–8.56-cycles/degree baseband; the solid line shows the dashed-curve response plus the aliasing. As indicated by the large difference between the line and the curve, the aliased components constitute a large part of the response of the CBbi cells. This difference is also large in Fig. 12(b); in fact, the situation for the on-\( \beta \) cells is even worse because much of the CBbi signal that acts as input to the CBbi \( \Rightarrow \) on-\( \beta \) synapse consists of the aliased frequencies shown in Fig. 12(a). Figure 12 shows results for the plus-shaped filter. The cross-shaped filter displays a similar degree of aliasing. As we discuss in Section 3, the main anatomic reason that these filters are not sufficiently low pass is that their synapses have small convergences.

The aliasing in the CBbi and on-\( \beta \) images would be a problem in the cat retina if not for several factors, which include (1) the optic pointspread function, (2) the cone aperture, (3) cone–cone gap junctions, (4) temporal blurring from eye tremor, and (5) the low-pass nature of natural scenes. Several authors have shown that the optics and the cone aperture remove frequency components that would otherwise alias in the foveal cones of primates and humans. The cone–cone gap junctions and eye tremor also low-pass filter the image. The amplitudes of spectra of natural scenes generally drop in inverse proportion to spatial frequency. In humans the high frequencies have so little energy as to make negligible what aliasing does occur. It is unclear whether aliasing occurs. The analysis in Fig. 12 considers the worst-case situation of viewing a sharp, narrow line, which, in the limit of being infinitely thin, has a Fourier transform with constant magnitude along the frequency axis perpendicular to the line. The influence of these factors in the cat eye can be seen in the optics-to-cone frequency response \( t_{\text{ces}}(f) \) [Fig. 13(a)], which we derive from the optics-to-cone spatial impulse response as calculated by Smith and Sterling. The term optics-to-cone indicates that the response includes all spatial processing that occurs from the cornea to the cone pedicle outputs, that is, the eye’s optics, the cones and the cone aperture, the

---

Fig. 10. Comparison of typical space-varying filters for the CBbi \( \Rightarrow \) on-\( \beta \) synapse with on-\( \beta \) cell dendritic fields. (a) Space-varying filters \( g_{\text{in}}, g_{\text{out}}, g_{y}, \) and \( g_{z} \). Filled diamonds indicate the on-\( \beta \) samples, and small open circles indicate the CBbi samples touched by the on-\( \beta \) samples. The contacts are shown explicitly by the lines. The offset of the diamonds from the circles reflects the different alignments between samples in arrays of different densities. The numbers are the filter weight values rescaled so that they show the number of synaptic contacts from each CBbi cell in the model. (b) Tangential projections of on-\( \beta \) cell dendritic trees at 1° eccentricity. Different letters refer to different CBbi bipolars. Each occurrence of a letter indicates a synaptic contact between that bipolar and the on-\( \beta \) cell (on-\( \beta \) cells from Ref. 6).
Using the mathematics of multirate filtering, we can compute the CBb, and on-β cell frequency responses in terms of the optics-to-cone frequency response. The frequency response for the output of the one-dimensional system shown in Fig. 3(c) is given by

\[
Y(\exp(j\omega)) = \frac{1}{M} \sum_{r=-\infty}^{\infty} H[\exp\left(\frac{jr\pi}{M}\right)]
\times X[\exp\left(\frac{j\omega}{M}\right)],
\]

where \(H[\exp(j\omega)]\) is the DTFT of \(h(u)\) and \(j = \sqrt{-1}\).

Applying Eq. (23) to Fig. 7, substituting for \(L\) and \(M\), and

cone-cone gap junctions, and the A and B horizontal cells. Smith and Sterling calculated the cat optics-to-cone spatial impulse response by deconvolving the on-β cell receptive field with a function based on the synaptic weighting between cones and on-β cells. Frequency components above 5.48 cycles/degree are attenuated to 1/100 of the maximum of \(i_{\text{cone}}(f)\). Thus the 12.75- and 6-cycle/degree signals in Fig. 11 are lost to neural noise before they can alias.
Converting radian frequency to spatial frequency gives

\[ I_{\text{CBb}}(f) = \frac{1}{2} \left| \sum_{n=0}^{N} H_{\text{cone} \to \text{CBb}}[\exp(j0.39f_n - j\pi n)] \right| \times I_{\text{cone}}[\exp(j0.39f_n - j\pi n)], \]  

(24)

\[ I_{\text{on-\beta}}(f) = \frac{1}{16} \left| \sum_{n=0}^{N} H_{\text{CBb} \to \text{on-\beta}}[\exp(j8.976f_n - j\pi n/8)] \right| \times I_{\text{CBb}}[\exp(j8.976f_n - j\pi n/8)]. \]  

(25)

where \( H_{\text{cone} \to \text{CBb}}[\exp(j\omega)] \) is the DTFT of \( h_{\text{cone} \to \text{CBb}} \), \( H_{\text{CBb} \to \text{on-\beta}}[\exp(j\omega)] \) is the DTFT of \( h_{\text{CBb} \to \text{on-\beta}} \), and \( I_{\text{CBb}}(f) \) and \( I_{\text{on-\beta}}(f) \) are the frequency responses of the CBb and on-\beta images, respectively. Here \( h_{\text{cone} \to \text{CBb}} \) refers to either the plus-shaped or the cross-shaped filter. For clarity we have presented the one-dimensional equations and defined each output in terms of the previous signal in the system. The two-dimensional results are similar. These frequency responses are plotted in Figs. 13(b) and 13(c) for the plus-shaped filter.

As we can see in Fig. 13(b), the dashed and solid curves are identical for the CBb image. There is no aliasing, since frequencies above the CBb Nyquist rate are highly attenuated in the cone image. As Fig. 13(c) shows, there still remains a small amount of aliasing in the on-\beta image. Nevertheless, we regard this aliasing as insignificant for several reasons:

1. As indicated by the small difference between the curves, the aliasing forms a nearly insignificant part of the on-\beta output.
2. The gain for the aliased frequencies never exceeds 1.6% of the maximum on-\beta frequency response and is likely at or barely above the level of neural noise.
3. Because the sampling of all three classes of cell is highly irregular, the small amount of aliasing that may occur scatters into broadband noise.

Thus, for broadband images, the attenuation of high frequencies performed before or at the cone level prevents aliasing that would otherwise occur in CBb and on-\beta cells. These results are essentially identical for both the plus-shaped and cross-shaped filters. The root-mean-square difference for the images produced by the plus-shaped and cross-shaped filters is 0.0079 for the CBb images and 0.0093 for the on-\beta images. These two filters produce such similar outputs because they are similar in those frequencies passed by the cones and differ significantly only in the frequencies attenuated by the cones [compare the attenuated region in Fig. 13(a) with the DTFT's in Fig. 9(a)].

We compare our calculations with results from cat psychophysics. Hall and Mitchell show that cats can both detect gratings and discriminate between vertical and horizontal gratings with equal ability up to between 8.5 and 9 cycles/degree. Their results suggest that aliasing does not play a part in the detection of the gratings. On-\beta and off-\beta cells have the same densities, and the Nyquist rate of each class of cell at the area centralis is approximately 6.5 cycles/degree.

Since the Nyquist rate of both classes considered as one array is \( 6.5 \sqrt{2} = 9.2 \) cycles/degree, Hall and Mitchell suggest that the on-\beta and off-\beta
cells are interpreted by the cat brain as a single sampling mosaic.

We include both on-β and off-β cells by doubling the CBb₁ and on-β cell densities. Mathematically, this step reduces the CBb₁ and on-β cell array sampling periods by a factor of \( \sqrt{2} \), scales their frequency axes in Fig. 13 by the same amount, and doubles the cone ⇒ CBb₁ divergence to 2.4 (Eq. (12)). The maximum grating acuity in the cat is determined by the center of the retina, the region with the highest cell densities. To compare our analysis with that in Ref. 44, we assume that the convergences and the divergences of the on-β and off-β cell pathways are the same at 0° eccentricity as they are at 1° eccentricity (where the anatomic data are available). Cone density at 0° is typically 30,000/mm², while in our model \( d_{\text{cone}} = 24,200 \) at 1°. The assumption permits us to reinterpret the frequency axes in Fig. 13 for 0° eccentricity by rescaling the axes by 30,000/24,200. Thus the abscissa of Fig. 13(c) is rescaled from its present value of 4.81 to 
\[
\left( \frac{4.81 \times 30,000}{24,200} \right) \sqrt{2} = 8.43 \text{ cycles/degree.}
\]
Figure 13(c) then implies that frequency components up to approximately 8.43 cycles/degree are passed by the model with significant gain and without aliasing. Beyond this frequency, components are severely attenuated, and what aliasing remains is scattered into broadband noise. This cutoff is in agreement with Hall and Mitchell's measurement of 8.5-9 cycles/degree.

It is of interest to determine to what degree the optics of the cat eye alone are responsible for protection of the cells from aliasing. The frequency response of the cat's optics is given by Fig. 14(a).47,48 If we substitute this response for \( I_{\text{opt}}(f) \), we effectively ignore the effects of cone aperture, cone–cone gap junctions, A and B horizontal cells, and blurring by eye tremor. Figures 14(b) and 14(c) show that the optics are adequate for the removal of practically all the aliasing in the CBb₁ cells but permit significant aliasing in the on-β cells. The plotted results are for the plus-shaped filter, but they are similar to those for the cross-shaped filter. Clearly the majority of the aliasing is removed by the optics, but additional attenuation of frequencies above 4.81 cycles/degree is necessary for its complete prevention. This attenuation is largely performed by the cone aperture and the cone–cone gap junctions. Owing to the irregular sampling, much of the aliasing in the on-β cell array would scatter into broadband noise, but the significant energy of this noise would still greatly degrade the on-β signal.

3. DISCUSSION

Many anatomic properties are a consequence of the different densities of cell classes. Models based on multirate filtering can incorporate these properties quite naturally. The efficient implementations that are possible with multirate models permit rapid computation of the outputs for arbitrary inputs. The closed-form solutions in space and frequency domains provide a means of analyzing in detail how anatomic properties dictate the responses of cells. These features permit the derivation of generalized input/output relationships for arbitrary anatomic densities, convergences, divergences, and other cell properties. In this paper we developed a means for the modeling of synapses between several cell layers that uses multirate filtering,
thereby introducing the opportunity for a model of the effects of postreceptoral filtering. We discussed methods that ensure that a set of filters modeling synapses touches every sample in the presynaptic array. In applying the modeling technique to the cone → CBb1 → on-β cell pathway, we calculated the frequency responses of the CBb1 and on-β cells based on the cone frequency response, examined how aliasing in these cells is prevented, and compared our results with cat psychophysics. We showed that the optics of the cat eye are insufficient to prevent aliasing in these cells independently. Multirate analysis demonstrates that the highest spatial frequency that can be passed by the retina without aliasing is determined not always only by the densities of cones, bipolar cells, and ganglion cells but also by the synaptic and the dendritic weighting between these cells.

Because of the detail with which spatial anatomic information is incorporated into multirate models, analysis of these models can potentially permit the study of several retinal properties. The propagation of noise along converging synaptic paths in multiple-cell layers can be modeled by the association of noise processes with each synapse. The influence of convergence, divergence, synaptic and dendritic weighting, and synaptic gain on receptive fields can be examined by calculation of the appropriate transfer functions. Trade-offs between the number of synaptic layers and the volume of dendritic trees can be calculated as an architecture varies among parallel, hierarchic, and various hybrids. The approach to these trade-offs is reported in an early form in Refs. 13–15.

A distinct difference between the cat retina and our model is that retinal cells are usually laid out in a disordered manner, whereas the samples in the model are in a regular array. The consequences of this irregular sampling depend on whether subsequent neural processing senses the locations of the cells and uses this information. Ahumada described mechanisms whereby the visual system could determine the location of its photoreceptors and alter its processing accordingly. Hirsch and Miller showed that human-acuity measurements are matched up to 1.5° eccentricity by primate cone and data scaled to the human retina without any correction for sampling disorder. They conclude that cone positions are known to the visual system, at least within this range. We adopt this view for the cat retina.

When sample locations are known for a finite array of irregularly spaced samples, signal processing theory shows that perfect reconstruction of a bandlimited input is possible, regardless of the degree of irregularity, provided that there is no noise in the signal. Noise interferes with the reconstruction of signals sampled on either regular or irregular lattices; but the more the sampling deviates from a regular lattice, the more susceptible to noise is the reconstruction. Thus, in Fig. 2, irregularly spaced samples of x(n) can be regarded as an enhancement of the noise in y(m) caused by noise in x(n).

Irregular sampling also has consequences in the implementation of the photoreceptors and irregular spacing of postreceptoral cells. Photoreceptors sample a continuous light image, and their exact locations are given importance by virtue of the spatial content of the image. Postreceptoral cells can be regarded as establishing patterns of connectivity from one layer to the next. Since the cell bodies can be shifted without a change in the patterns of connections, their exact spatial positions are not overly significant. The situation is analogous in the implementation of the model. Once the filters are designed, the samples can be mathematically irregularly displaced without any effect on their values or on the filters that connect them. The situation is somewhat different for designing the model. The relative locations of samples in all the model's arrays are used in the design of the space-varying filters. While the sample locations are given importance in this manner, they can be regarded as average relative distances between the terminal branches of a presynaptic cell and the center of a postsynaptic cell's dendritic field. This approach matches the definitions of swf(x,y) and dwf x,y as average weights for class A cells whose terminal branches are located at (x,y) relative to the center of a class B cell dendritic field.

Even with regular sampling the multirate approach displays some of the variation and the irregularity seen in retinal anatomy. The irregularity is unexpected, because multirate models are based only on average measures. Yet, given different presynaptic and postsynaptic cell densities, there is no choice but that of variety in dendritic fields, even if the cells are in a regular array. Since this variation [Fig. 10(a)] occurs with period L along each axis and is not so great as anatomic variation [Fig. 10(b)], it suggests that anatomic cell-to-cell irregularity in the retina has two components: (1) variation that is due to the different relative locations of cells in presynaptic and postsynaptic cell arrays as required by different cell densities and (2) variation that is due to partly random branching, direction, and length of cell growth.

Multirate filtering provides a means of examining the amplification or the aliasing of frequency components that propagate through cell layers of different densities. For example, for the case of resampling as in Fig. 3(c1), if the filter has the appropriate cutoff, then the maximum frequency that can be passed by the system is the lower of the two Nyquist rates of the input x(n) and the output y(m). If the filter has a higher cutoff, as is the case in the cone → CBb1 → on-β example, components in x(n) will alias in y(m) unless they are attenuated before x(n). If the filter has a lower cutoff, a smaller range of frequencies is passed to y(m). Thus the highest frequency passed without aliasing is determined not only by the densities of the cell layers but also by the swf and the dwf between the layers. The present analysis suggests that the attenuation of high frequencies in the cones prevents aliasing that would otherwise occur in CBb1 and on-β cells. The analysis also suggests that, while assuming responsibility for most of this attenuation, the cat's optics cannot independently prevent all the aliasing in these cells.

The two filters in the cone → CBb1 → on-β model permit aliasing primarily because their convergences are so small. As we showed in Subsection 2.A.2, convergence corresponds to the average number of weights in the space-varying filters. The fewer weights there are in a low-pass filter, the less well the filter approximates an ideal low-pass filter and the more it passes frequency components above the ideal cutoff. Two-dimensional space-varying filters with only five to seven weights cannot possibly act
sufficiently low pass for the prevention of most of the aliasing in the model, and the prefiltering in the cone image is necessary. In this regard, it is of interest that the convergences of the cone => A and the cone => B horizontal cell synapses near the area centralis in the cat retina are of the order of 140 and 90, respectively. The ratios of presynaptic to postsynaptic cell densities between cones and either type of horizontal cell are much greater than those for either of the synapses in the cone => CB1 => on-β pathway. These greater density ratios provide the potential for much more aliasing than that in our example, even with the cone prefiltering. The large convergences may be necessary for preventing the aliasing associated with low horizontal cell densities.

ACKNOWLEDGMENTS

We thank Peter Sterling for many helpful discussions and for his help with the anatomical pictures. This study was supported by U.S. Air Force Office of Scientific Research grant 91-0082. B. Levitan was also supported by a National Institutes of Health fellowship under grant 5-T32-GM07170.

REFERENCES

Conversions Between Parallel and Hierarchic Architecture Analysis Multirate Filter Banks
Bennett Levitan and Gershon Buchsbaum

Abstract—We derive general conversions between equivalent parallel and hierarchic analysis multirate filter banks (MRB's) as well as sufficient conditions for existence and uniqueness of the conversions. We use MRB's with arbitrary, rational number changes in sampling rate between successive outputs and arbitrary LTI filtering for each output. Conversion consists of commuting sampling rate expanders, sampling rate compressors, and filters to turn one MRB into the form of the other. For a class of MRB's we call "well-formed," the conversions between architectures are one to one.

I. INTRODUCTION

An analysis multirate filter bank (MRB) consists of a set of filters that produces several different sampling rate, or spatial scale, versions of a signal. MRB's can operate very efficiently by allowing differential allocation of processing and storage resources to the spatial scales [2], [6], [11], [14]-[17], [19], [20]; however, these advantages depend heavily on the architecture. In a hierarchic architecture, the nth output is computed by operations on the n-1st output; the first output is computed directly from the input. In a parallel architecture, all outputs are computed directly from the input. If the corresponding outputs of two MRB's (with the same or different architectures) are equal for all inputs, the MRB's are equivalent. Compared to a parallel architecture, an equivalent hierarchy has the advantages of: i) smaller filter sizes, ii) lower sampling rates for most computation, and iii) a smaller total number of connections between processing elements, a feature important for "hardwired" implementations [1], [5], [9], [12], [15]. The hierarchy's disadvantages include: i) slower operation in hardwired implementations, ii) less flexibility in choosing filters, iii) less straightforward design, and iv) susceptibility to noise and errors propagating through successive outputs.

To benefit from these and other tradeoffs between equivalent hierarchic and parallel analysis MRB's requires a method to convert between the parallel and hierarchic architectures and means to determine when such conversions are allowed. These issues have been examined for some MRB's [2]-[4], [12], [18] but not for the general case. In this correspondence, we develop equations and conditions for the conversions.

II. THEORY

A. Definitions of Multirate Filter Bank Architectures

In the parallel multirate filter bank, each output \( y_n(x_n) \) is computed from the original input \( y_0(x_0) \) in three steps: i) sampling rate expansion by factor \( L_n \), ii) filtering by arbitrary, linear time-invariant (LTI) parallel filter \( h_n(u_n) \), and iii) sampling rate compression by factor \( M_n \). These basic operations of multirate filtering are described extensively in [5], [17]. Fig. 1(a) shows the parallel MRB

---


corresponding hierarchic MRB variables. The spatial variables for the filters and outputs are indexed to differentiate between coordinates on different levels.

in the \( z \) domain, where \( H_n(z) \) is the \( z \) transform of \( h_n(u_n) \). To distinguish between parallel and hierarchic MRB variables, all parallel MRB variables are in boldface type.

The hierarchic multirate filter bank successively produces outputs \( y_i(x_i) \) through \( y_N(x_N) \) in \( N \) stages. The nth stage computes \( y_n \) from \( y_{n-1} \) in three steps: i) sampling rate expansion by factor \( L_n \), ii) filtering by arbitrary, LTI hierarchic filter \( h_n(u_n) \), and iii) sampling rate compression by factor \( M_n \). Fig. 1(b) shows the hierarchic MRB in the \( z \) domain, where \( H_n(z) \) is the \( z \) transform of \( h_n(u_n) \).

The sets of parallel and hierarchic MRB's can be partitioned into equivalence classes [10]. Each parallel equivalence class contains an infinite number of equivalent MRB's whose resampling factors satisfy \( M_n/L_n = K_n \), where \( K_n \) is a rational number. From elementary properties of prime numbers, it follows that each class contains one and only one MRB whose resampling factors satisfy

\[ M_n \text{ and } L_n \text{ are relatively prime, } n = 1 \cdots N. \]

Similarly, each hierarchic MRB class contains an infinite number of equivalent MRB's whose resampling factors satisfy \( M_n/L_n = K_n \), where \( K_n \) is a rational number. In each class there is one and only one MRB whose resampling factors satisfy

\[ M_n \text{ and } L_n \text{ are relatively prime, } n = 1 \cdots N. \]

We call MRB's satisfying (1) or (2) "well formed." As shown in the next section, compared to equivalent non-well-formed MRB's, well-formed MRB's are more likely to satisfy the conversion conditions derived below. Because each equivalence class is uniquely identified by one well-formed MRB, for well-formed hier-

Manuscript received April 5, 1990; revised September 6, 1991. This work was supported in part by AFOSR Grant 91-0082. The work of B. Levitan was also supported by an NIH fellowship under Grant 5-T32-GM07170.

The authors are with the Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA 19104-6392.

IEEE Log Number 9202795.

1053-587X/92$03.00 © 1992 IEEE
archie and parallel MRB's that satisfy the conversion conditions, the conversions are one to one.

B. Relations between Multirate Filter Bank Architectures

1) Conversion from Hierarchic MRB to Parallel MRB:

a) General conversion: The basic step for MRB conversion is the "adjacent pair" commutation (Fig. 2). Adjacent pair commutations convert two adjacent stages into a single equivalent stage. Using rules for commuting multirate operations with filters [4] (these rules are also referred to as the "noble identities" [17]), the $L$-fold expander is commuted to the left and combined with the first expander. Similarly, the $M_{n-1}$-fold compressor is commuted to the right and combined with the second compressor (Fig. 2(b)). The filter system functions are expanded by the factors of the resamplers commuted with their filters. The filters are then joined (Fig. 2c). If the resulting resampling factors are not relatively prime, their greatest common divisor can be removed if the combined filter is also downsampled by the common divisor (Fig. 2d). This method is similar to that used in [4], [7].

Adjacent pair commutations are not always possible, since an $L$-fold expander and an $M$-fold compressor can commute if and only if $L$ and $M$ are relatively prime [7], [17]. By removing the greatest common divisor of the resampling factors in adjacent pair commutations, the reduced factors are more likely to be relatively prime to the factors of other stages during subsequent commutations. For this reason, removing common divisors in each stage by converting a hierarchic MRB to its well-formed equivalent [10] increases the likelihood of its conversion.

$n - 1$ adjacent pair commutations turn the first $n$ stages of a hierarchic MRB into a single stage computing output $y$, directly from $y_0$. The stages can be commuted in any sequence. The set of relatively prime conditions for a particular commutation sequence is the "hierarchic to parallel commutation condition" for that sequence. If a particular sequence requires commuting factors that are not relatively prime, conversion in that sequence is not possible. A hierarchic MRB can be converted into an equivalent parallel MRB if at least one sequence of commutations is possible for each output. The final adjacent pair commutation in a sequence insures that $M_i$ and $L_i$ are relatively prime. Thus, the parallel MRB derived from a hierarchic MRB satisfies (1) and is well formed.

Conversion with any sequence of commutation gives equivalent parallel factors

$$L_n = \prod_{i=1}^{n} L_i / \gcd \left( \prod_{i=1}^{n} L_i, \prod_{i=1}^{n} M_i \right)$$

and

$$M_n = \prod_{i=1}^{n} M_i / \gcd \left( \prod_{i=1}^{n} L_i, \prod_{i=1}^{n} M_i \right)$$

where $\gcd(A, B)$ denotes the greatest common divisor of $A$ and $B$ [10]. The equivalent $H_n(z)$ is the product of multiply resampled hierarchic filters $H$, through $H_n$ and depends on which of the potential $(n - 1)!$ possible sequences of conversion is used. Each sequence has a different commutation condition and removes common divisors in a different manner.

b) Relatively prime factors case: We can derive closed-form equivalent filters for a less general case. Let all resampling factors commuted in the conversion be relatively prime; that is,

$$M_i$$ and $$L_i$$ are relatively prime, $$i = 1 \cdots N, j = i + 1 \cdots N.$$
Fig. 3. Conversion of hierarchic MRB into parallel MRB for relatively prime factors case. (a) Result of using $z$-domain identities on Fig. 1(b) for $y_2$. (b) Result for $y_3$. 

$n$, the parallel MRB can be written in terms of the hierarchic MRB as

$$ M_n = \prod_{i=1}^{n} M_i, \quad L_n = \prod_{i=1}^{n} L_i \quad n = 1 \cdots N \quad (6,7) $$

$$ H_n(z) = \prod_{j=1}^{n} H_j(z^{M/j}) \quad n = 1 \cdots N. \quad (8) $$

where

$$ W_{i,j} = 1, \quad W_{j,j} = \prod_{i=1}^{j-1} M_i \prod_{i=j+1}^{n} L_j \quad n = 2 \cdots N, j = 1 \cdots n. \quad (9) $$

Equations (6) and (7) result from dropping the gcd terms in (3) and (4). Equation (8) states that the $z$-domain filter for level $n$ of the equivalent parallel MRB is the product of expanded versions of hierarchic filters $H_j$ through $H_n$. It is valid for all sequences of conversion since, in any sequence, filter $H_j$ commutes with all expanders to its right and all compressors to its left. $W_{i,j}$ is the product of factors of all expanders and compressors that commute with $H_j$.

The relatively prime factors case is often satisfied in practice. The sampling scheme most often used in hierarchic MRB's is reducing the sampling rate by an integer factor from $y_n$ to $y_n$. This scheme corresponds to setting $M_n = constant$ and $L_n = 1$ for all $n$ and trivially satisfies (5).

2) Conversion from Parallel MRB to Hierarchic MRB: The parallel MRB definitions of $y_n-1$ and $y_n$ precisely define the $n$th hierarchic stage. To convert a parallel MRB into a hierarchic MRB, consider a system in which $y_n$ is calculated by the parallel MRB and then filtered by a hierarchic stage to yield $y_n$. (Fig. 4(a)). We perform an adjacent pair commutation on Fig. 4(b), but do not remove gcd $\{L_n-1, L_n, M_n-1, M_n\}$ (Fig. 4(b)). This commutation is possible only if $L_n$ and $M_n-1$ are relatively prime. To increase the likelihood of conversion, many potential common divisors can be removed by making the parallel MRB well formed before the commutations. The resulting stage (Fig. 4(b)) is similar to the parallel MRB for $y_n$ (Fig. 4(c)); however, because it may have nonrelatively prime resampling factors, they cannot be directly compared. To compare the two, we insert $C_n$-fold expanders and compressors in the parallel MRB (Fig. 4(d)) and commute them to get a stage whose coefficients may not be relatively prime (Fig. 4(e)). Comparing Figs. 4(b) and (e) gives

$$ L_1 = L_1, \quad L_n = C_n \cdot L_n/L_{n-1} \quad n = 2 \cdots N \quad (10) $$

$$ M_1 = M_1, \quad M_n = C_n \cdot M_n/M_{n-1} \quad n = 2 \cdots N \quad (11) $$

$$ H_n(z) = H_n(z), \quad H_n = H_n(z^{C_n/M_n})/H_{n-1}(z^{L_n/M_{n-1}}) \quad n = 2 \cdots N. \quad (12) $$

We can write the “parallel to hierarchic commutation condition” as: A parallel MRB can be converted into an equivalent hierarchic MRB if all $M_{n-1}$ and $L_n$ are relatively prime for $n = 2 \cdots N$. $L_n$ is calculated from the given parallel resampling factors using (10). By setting

$$ C_n = L_{n-1} \cdot M_{n-1}/[gcd(L_n, L_{n-1}) \cdot gcd(M_n, M_{n-1})] \quad (13) $$

$M_n$ and $L_n$ can be made relatively prime. Thus, the hierarchic MRB derived from a parallel MRB satisfies (2) and is well formed.

A second condition arises from the practical requirement that the filters $H_n(z)$ in (12) be stable. $H_n(z)$ is stable provided its ROC includes the unit circle. The given parallel filters $H_n(z)$ are stable. Hence, $H_n(z^{C_n/M_n})$ and $H_{n-1}(z^{L_n/M_{n-1}})$ in (12) are also stable, because raising $z$ to a constant does not cause poles or zeros to move on or off the unit circle. However, $H_n(z) = H_n(z^{C_n/M_n})/H_{n-1}(z^{L_n/M_{n-1}})$ may still not be stable, since $H_{n-1}(z^{L_n/M_{n-1}})$ may have zeros on the unit circle which would exclude the unit circle from $H_n(z)$'s ROC. Hence, a parallel MRB can be converted into a hierarchic MRB with stable filters if and only if the ratios of parallel filters $H_n(z^{C_n/M_n})/H_{n-1}(z^{L_n/M_{n-1}})$ are stable for $2 \leq n \leq N$.

On a more intuitive level, this condition requires that any frequency component required in output $y_n$ of the parallel MRB be retained in output $y_{n-1}$. Clearly, this component need also be retained in outputs $y_i$ through $y_{n-2}$. This point highlights an inherent difference between the two architectures: The $n$th filter in a hierarchy of stable filters has access only to frequency components represented in signal $y_{n-1}$, while parallel MRB filters have access to all components in the input $y_0$.

In practically every MRB we have encountered in the hierarchic processing literature, the outputs are indexed in order of decreasing sampling rate. In such MRB's, the filters $H_n(\omega)$ are nonzero for $0 \leq |\omega| \leq \pi/\max(L_n, M_n)$ and negligible at higher frequencies. These filters satisfy the stability condition, because the cut-offs decrease sufficiently from level $n-1$ to $n$. In the subband coding literature, however, filters are generally not of this form and will often not satisfy the stability condition. Subband coding analysis MRB's generally cannot be converted to a hierarchic equivalent; however, they could be implemented with difference-pyramid type hierarchies [1], [2], [12], [14], [15]. For MRB's with at least one output whose sampling rate is above that of $y_0$, the stability condition does not apply to any $\omega$ such that $\pi \cdot M_n/L_n < |\omega| \leq \pi$. 
Since these frequency components cannot be represented in \( y_0 \), they always contribute nothing to \( y_0 \).

### III. Scaled Parallel Filters

Many MRB applications use scaled, or self-similar, parallel filters [1], [2], [11], [12], [19]. These filters are of the same shape but scaled differently in height and width on each level. To incorporate scaled filters into an MRB, we define the parallel filters as scaled, sampled versions of function \( f(t) \) where \( t \) is a continuous variable:

\[
h_t(u) = A_n \cdot f(u/B_n), \quad n = 1 \cdots N. \tag{14}
\]

We assume the MRB’s satisfy the commutation conditions. \( A_n \) and \( B_n \) are positive, real-valued coefficients that allow arbitrary vertical and horizontal scaling of \( f \). \( f \) must have a continuous argument since the \( B_n \) are arbitrary. If we define discrete function \( f_n(x) = f_n(x/B_n) \) and \( F_n(z) \) as its \( z \)-transform, the \( z \)-transform of (14) can be written

\[
H_n(z) = A_n \cdot F_n(z), \quad n = 1 \cdots N. \tag{15}
\]

To solve the hierarchic filters, we substitute (15) into (12) yielding

\[
H_n(z) = \frac{A_n \cdot F_n(z^{c(M_n-1)})}{A_{n-1} \cdot F_{n-1}(z^{c(M_n-1)})}, \quad n = 2 \cdots N. \tag{16}
\]

Equation (16) gives the hierarchic filters for scaled parallel filters based on \( f(t) \).

### VI. Discussion

This correspondence presents relations between parallel and hierarchic implementations of analysis MRB’s. Fig. 5 summarizes the theory. The theory gives closed-form equations and conditions for converting a parallel MRB into an equivalent well-formed hierarchic MRB with stable filters. It gives a method and conditions to convert a hierarchic MRB into an equivalent well-formed parallel MRB and closed-form equations for the relatively prime factors case. Since MRB’s of either architecture are members of equivalence classes containing an infinite number of MRB’s, the conversions are obviously not unique. However, for well-formed MRB’s satisfying the conversion conditions, conversion is one to one. For conversion in either direction the commutation conditions are sufficient but may not be necessary, since other methods not based on commutation might still be able to give an equivalent MRB.

MRB conversion allows taking advantage of the property tradeoffs between hierarchic and parallel MRB architectures. If speed is the primary consideration, a hardwired parallel architecture can be used. If the hardware is limited, a hardwired hierarchic architecture is a better choice. On serial computers, the hierarchic is both faster and requires less storage for filter coefficients than the parallel architecture. The conversions are easily extended to image signals if rectangular sampling is used. The conversions are also useful in the modeling of naturally occurring systems, such as parts of the nervous system [8].

In practice, many practical MRB’s easily satisfy the commutation and stability conditions. MRB’s in the hierarchic literature commonly use low-pass or Gaussian filters and have relatively prime factors. These MRB’s always fulfill the commutation and stability conditions. Other MRB’s normally use either low-pass or bandpass filters. MRB’s with low-pass filters will satisfy the stability condition, since the filter cutoffs typically decrease with decreasing sampling rate. Parallel MRB’s with bandpass filters do not satisfy the condition, but could be implemented with difference-pyramid type hierarchies [1], [2], [12], [14], [15].

### References


Appendix F

Conversion and Trade-offs between Scaled Gaussian Parallel and Hierarchic Analysis Multirate Filter Banks

BENNETT LEVITAN AND GERSHON BUCHSBAUM

Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, 220 South 33rd St., Philadelphia, Pennsylvania 19104-6392

Received September 20, 1991; accepted July 22, 1992

Scaled Gaussian analysis multirate filter banks (MRBs) are analysis MRBs whose filters are Gaussians scaled differently in width and height. They are frequently employed in image processing and visual systems modeling. We define generalized scaled Gaussian analysis MRBs for both parallel and hierarchic architectures and derive closed-form equations for conversions between them. The MRBs have arbitrary rational number changes in sampling rate between successive outputs, and arbitrary vertical and horizontal scaling of the Gaussian filters for each output. We calculate the number of multiplications, number of additions, and total number of filter coefficients to compare the parallel and hierarchic architectures as implemented with direct form and time-varying filters. In all cases, the parallel MRB requires considerably more multiplications, additions, and filter coefficients than the equivalent hierarchic MRB. However, the relative differences are far less severe in the time-varying case. We also derive a useful approximation for the discrete time Fourier transform of a Gaussian.

INTRODUCTION

To detect and manipulate image features, many image processing systems operate on several spatial scales. For example, image recognition systems apply the same detection algorithm to small and large versions of the same objects [1-4]. It is often beneficial to distinguish between spatial scales. The technique of subband coding, for instance, operates efficiently by differentially allocating processing and storage resources to the scales [1, 5-9]. In many cases, multispatial scale processing is performed with analysis multirate filter banks (MRBs) [1, 5, 8, 10-15]. Analysis MRBs consist of a set of filters and resamplers that produce several different sampling rate versions of an image. They attain great computational savings by manipulating large spatial scale features in low sampling rate outputs. Analysis MRBs can be implemented with parallel, hierarchic or other architectures. In the parallel architecture, each output is produced by filters and resamplers acting on the input image. In the hierarchic architecture, the nth output is produced by filters and resamplers acting on the n - 1st output.

An analysis MRB often used in image processing is the scaled Gaussian analysis MRB [5, 10-14]. These are MRBs all of whose filter impulse responses are Gaussians scaled differently in width and height. Reasons that scaled Gaussian MRBs are popular include: (i) Two-dimensional Gaussians are separable and hence easily implemented. (ii) As shown below, filters of the MRB equivalent to a given scaled Gaussian MRB of the other architecture are always Gaussian. (iii) As also shown below, provided the Gaussians of a parallel MRB increase sufficiently in width between successive outputs, the equivalent hierarchic filters are real. With arbitrary filters, the hierarchic equivalent filters may be IIR, complex, or unstable. (iv) The $V^G$ operator commonly used in image processing can be well matched by taking the difference between appropriately Gaussian-filtered images [16]. (v) Gaussians are very useful in modeling the receptive fields of cells in mammalian visual systems [17-21]. Parts of the visual systems can be modeled as an MRB with Gaussian or related filters [13, 14, 21, 22].

In this paper, we derive conversions between scaled Gaussian hierarchic and parallel MRBs and calculate some of their properties. Conversions and properties have been derived for some MRBs, but not for the general case [5, 10-12, 15, 23, 24]. The MRBs that have been examined downsample by the same factor between outputs and change the sampling rate only by integer factors. Conversions and properties for more general cases were derived in [25, 26]. In this paper, Gaussians have arbitrary vertical and horizontal scaling for each output, and the sampling rate may change by a different rational num-

---

1 This work supported by Grant AFOSR-91-00-82 from the Air Force Office of Scientific Research. B. Levitan was also supported by a National Institutes of Health fellowship under Grant 5-T32-GM07170.
Responses are defined as scaled, sampled versions of prototypical filters into a parallel MRB, the filter impulse response from $y_i(-1)$ in three steps: (i) sampling rate expansion by factor $L_i$, (ii) filtering by LTI parallel filter $h^N_i(u')$, and (iii) sampling rate compression by factor $M^N_i$ (Fig. 1a). In the hierarchical MRB, outputs $y_i(x_i)$ through $y_N(x_N)$ are produced successively in $N$ stages. The $n$th stage computes $y_n$ from $y_{n-1}$ in three steps: (i) sampling rate expansion by factor $L_n$, (ii) filtering by LTI parallel filter $h^N_n(u'_n)$, and (iii) sampling rate compression by factor $M^N_n$ (Fig. 1b).

**Parallel to Hierarchic Conversion.** To incorporate scaled filters into a parallel MRB, the filter impulse responses are defined as scaled, sampled versions of prototype function $f_c(t)$ where $t$ is a continuous variable [26],

$$h_n(u_n) = A_n f_c(u_n/B_n) \quad n = 1, \ldots, N,$$  \hspace{1cm} (1)

where $A_n$ and $B_n$ are positive, real-valued coefficients that allow arbitrary vertical and horizontal scaling of $f_c$; $f_c$ must have a continuous argument since the $B_n$ are arbitrary. From [26], provided basic conditions for conversion are met, if we define discrete function $f_c(x) = f_c(x/B_n)$ and $F_c(\omega)$ as its discrete time Fourier transform (DTFT), the hierarchical filters are the inverse DTFT of

$$H_n(\omega) = \frac{A_n F_c(\omega C_n/M_{n-1})}{A_{n-1} F_{n-1}(\omega L_n/M_{n-1})}$$  \hspace{1cm} (2)

where $C_n = L_{n-1} M_{n-1}/[\gcd(L_{n-1}, M_{n-1}) \gcd(M_n, M_{n-1})]$ and $\gcd(a, b)$ is the greatest common divisor of $a$ and $b$. $L_n$ in (2) is the equivalent hierarchic upsampling factor for stage $n$. It and the equivalent downsampling factors are given by

$$L_1 = L_1, \quad L_n = C_n L_{n-1}/M_{n-1}, \quad n = 2, \ldots, N, \quad M_1 = M_1, \quad M_n = C_n M_{n-1}/M_{n-1}, \quad n = 2, \ldots, N. \quad (3, 4)$$

For scaled Gaussian filters, $f_c(t)$ is defined as $f_c(t) = \exp(-t^2/\sigma^2)$. Substituting $f_c$ into (1) gives the parallel filters $h^N_n(u'_n) = A_n \exp(u'_n^2/\sigma^2 B_n^2)$. For scaled Gaussian filters, we note that the DTFT of $f_c(x) = \exp(-x^2/\sigma^2 B_n^2)$ is

$$F_c(\omega) = \sqrt{\pi} \sigma B_n \sum_{m=-\infty}^{\infty} \exp \left( -\sigma^2 B_n^2 (\omega - 2\pi m)^2 \right)$$  \hspace{1cm} (5)

In the Appendix, we show that (5) can be excellently approximated in the interval $-\pi \leq \omega \leq \pi$ by $F_c(\omega) \approx \sqrt{\pi} \sigma B_n \exp(-\sigma^2 B_n^2 \omega^2/4)$ for $\sigma B_n \geq 1.8981$ to within 0.1%. Substitution into (2) gives hierarchical filters

$$H_n(\omega) = \frac{A_n B_n}{A_{n-1} B_{n-1}} \exp \left( -\frac{\sigma^2 (B_n^2 C_n^2 - B_{n-1}^2 L_n^2) \omega^2}{4M_{n-1}^2} \right) \quad n = 2, \ldots, N. \quad (6)$$

With the DTFT approximation, we can consider (6) as a constant multiplied by the approximated DTFT of a Gaussian with variance $\sigma^2 (B_n^2 C_n^2 - B_{n-1}^2 L_n^2)/M_{n-1}^2$. The inverse transform is

$$h_n(u_n) = \frac{A_n B_n M_{n-1}}{A_{n-1} B_{n-1}} \sqrt{\pi} \sigma (B_n^2 C_n^2 - B_{n-1}^2 L_n^2)^{1/2} \exp \left( -\frac{M_{n-1}^2 \omega^2}{\sigma^2 (B_n^2 C_n^2 - B_{n-1}^2 L_n^2)} \right) \quad n = 2, \ldots, N. \quad (7)$$

FIG. 1. Multirate filter bank architectures: (a) parallel MRB and (b) hierarchical MRB.
Scaled Gaussian Parallel and Hierarchical MRBs

Equation (7) gives the \( n = 2 \) filters for the hierarchical MRB equivalent to a given parallel MRB with scaled Gaussian filters. For \( n = 1 \), \( h_1(u_1) = h_1(u_1) \).

Because \( H_1(\omega) \) has no poles, the \( h_n(u_n) \) are FIR and stable. The hierarchical filters are Gaussian, since they result from the convolution of two Gaussians. Examples of filters from equivalent parallel and hierarchical MRBs are shown later in Fig. 3. If \( B_n \leq B_n/[n/L_{n-1}] \), then \( B_n^2 C_n^2 - B_n^2 L_n \) in the denominator of (7) is negative or zero, and the equivalent hierarchical filters are undefined or complex. In practice, this situation can be avoided by proper choice of \( B_n \) or reordering the outputs.

Hierarchic to Parallel Conversion. As in the parallel case, to incorporate scaled filters into a hierarchical MRB, the filter impulse responses are defined as scaled versions of a continuous prototype function \( f_c(t) \):

\[
h_n(u_n) = A_n f_c(u_n/B_n), \quad n = 1, \ldots, N.
\]

where again \( A_n \) and \( B_n \) are positive, real-valued coefficients that allow arbitrary vertical and horizontal scaling of \( f_c \). From [26], defining discrete function \( f_c(x) = f_c(x/B_n) \) and \( F_c(\omega) \) as its DTFT, \( H_1(\omega) = A_n F_n(\omega) \), and the equivalent parallel filter is

\[
H_n(\omega) = \prod_{i=1}^{n} A_i \prod_{j=1}^{n} F_n(\omega W_{j,n}).
\]

where

\[
W_{1,1} = 1, \quad W_{j,n} = \prod_{i=1}^{j-1} M_i \prod_{i=n}^{N} L_i,
\]

and \( M_i \) and \( L_i \) are relatively prime, \( i = 1, \ldots, N, j = i + 1, \ldots, N \). The constraints on \( M_i \) and \( L_i \) are sufficient to satisfy the basic conditions for hierarchical to parallel conversion and allows us to easily write the closed-form solutions given in (9) [26]. For scaled Gaussian filters, we set \( f_c(t) = e^{-t^2/\sigma^2} \). Substituting this into (8) gives

\[
h_n(u_n) = A_n e^{-u_n^2/\sigma_n^2}, \quad n = 1, \ldots, N.
\]

From the Appendix, the DTFT of the Gaussian \( f_c(x) \) for \( B_0 \sigma \geq 1.8981 \) is

\[
F_n(\omega) = \left( \frac{\pi}{\sigma} \right)^{n/2} B_n e^{-\omega^2 B_n^2/4}.
\]

Substitution into (9) gives

\[
H_n(\omega) = \sigma^2 B_n^{n^2+1} \prod_{i=1}^{n} A_i \exp \left( -\frac{\sigma^2 B_n^2}{4} \sum_{j=1}^{n} W_{j,n}^2 \right).
\]

Using the approximation, (11) can be regarded as a constant multiplied by the DTFT of Gaussian with variance \( \sigma^2 B_n^2 \sum_{j=1}^{n} W_{j,n}^2 \). Its inverse transform yields the parallel filters and completes the conversion.

Example Conversion. We demonstrate the conversions by implementing a four-output scaled Gaussian parallel MRB and its equivalent hierarchical MRB as derived above. Table 1 lists the factors and filter lengths used. \( A_n \) were chosen so each Gaussian has volume equal to \( L_n \). These values maintain the DC value of the input signal in each output. \( \sigma \) and \( B_n \) were chosen to make \( H_n(\omega) \) negligible beyond \( \pi/\max(M_n, L_n) \), the largest frequency that can be passed without aliasing or imaging [27]. \( M_n = 3^n \) and \( L_n = 2^n \) for all four outputs. Parallel MRB filter lengths were selected so the equivalent hierarchical filters included greater than 99.9% of the area they would occupy if infinitely extended.

Figure 2 shows the input image and outputs of the MRBs. Visually, the corresponding outputs are identical. All root-mean-square error differences between corresponding outputs are very small, and are caused mostly by edge effects. However, successive outputs show an increasing rms difference. We attribute this trend to a problem fundamental to hierarchical processors: Since the images and filters are quantized and the filters are of finite length, each output has a small amount of quantization error and aliasing error. In a hierarchical processor, these errors propagate from level to level and accumulate. While parallel processors also have quantization errors and aliasing, the errors do not propagate. For this reason,

<table>
<thead>
<tr>
<th>( n )</th>
<th>( A_n )</th>
<th>( B_n )</th>
<th>( L_n )</th>
<th>( M_n )</th>
<th>( L_n )</th>
<th>( M_n )</th>
<th>( t_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.06577</td>
<td>2.2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>0.01989</td>
<td>8.0</td>
<td>4</td>
<td>9</td>
<td>2</td>
<td>3</td>
<td>101</td>
</tr>
<tr>
<td>3</td>
<td>0.01052</td>
<td>22.0</td>
<td>8</td>
<td>27</td>
<td>2</td>
<td>3</td>
<td>345</td>
</tr>
<tr>
<td>4</td>
<td>0.00513</td>
<td>63.0</td>
<td>16</td>
<td>81</td>
<td>2</td>
<td>3</td>
<td>540</td>
</tr>
</tbody>
</table>
successive images in the hierarchic and parallel MRBs may increasingly differ.

Figure 3 shows the filters used to make these outputs. Each filter \( h_n \) or \( h_n^* \) is plotted against its own spatial axis \( u_n \) or \( u_n^* \). While \( h_1 \) and \( h_2 \) are similar as are \( h_3 \) and \( h_4 \), there is in general no reason for the hierarchic filters to be similar. \( h_1 \) is identical to \( h_1^* \). The other \( h_n \) differ from the \( h_n^* \) in several respects: (i) Their peak amplitudes are larger; (ii) Their energy is much more tightly compressed about the origin; (iii) While the parallel filters are successively more low pass, the hierarchic filters do not become more or less low pass in any order. These differences are due to the incremental nature of the hierarchy.

Properties of Architectures

The computational complexity of a system is the number of multiplications and additions it requires. In this section, we calculate the computational complexity and total filter storage required for the parallel and hierarchic MRBs as computed by a serial (single instruction stream, single data stream) machine. We give results for both one-dimensional MRBs and two-dimensional MRBs using separable filters. As in many applications, the "parallel" architecture, in which each output is computed directly from the input, is actually implemented on a serial computer; only one output is calculated at a time. The
complexity for the parallel architecture is thus the sum of
the complexities for each output. In a true parallel im-
plementation, in which the \(N\) outputs are computed simulta-
nuously, the complexity is the maximum of the complexi-
ties for each output.

The properties are based on MRB filter lengths. Since
parallel filter \(h_n\) is the convolution of sampling rate-
expaned versions of hierarchic filters \(h_1\) through \(h_n\) \([26]\),

\[
t_a = 1 + \sum_{j=1}^{n} W_{j,a}(t_j - 1).
\]

(14)

where \(t_a\) is the number of coefficients in \(h_a\). \(t_a\) is the
number of coefficients in \(h_n\), and the \(W_{j,a}\) are as defined in
\([10]\). Since the Gaussian filters are FIR, we can use \([10]\)
to solve \(14\) for finite \(t_a\), giving

\[
t_a = \frac{t_a - L_a(t_a - 1) - 1}{W_{a,n}} + 1.
\]

(15)

Parallel MRBs that can be converted to equivalent hi-
erarchic MRBs yield integer \(t_a\) in \(15\). For simplicity in
the derivations, we consider filter impulse responses of
odd length and later show the substitution necessary for
even-length filters.

We define \(S_n\) as the number of samples in signal \(y_n\).
From \([27]\), in the parallel MRB

\[
S_n = \left\lfloor \frac{S_0 L_n}{M_n} \right\rfloor
\]

(16)

and in the hierarchic MRB,

\[
S_n = \left\lfloor \frac{S_{n-1} L_{n-1}}{M_{n-1}} \right\rfloor = \left\lfloor \frac{S_{n-2} L_{n-2}}{M_{n-2}} \right\rfloor \left\lfloor \frac{L_{n-1}}{M_{n-1}} \right\rfloor \left\lfloor \frac{L_{n-1}}{M_{n-1}} \right\rfloor
\]

(17)

where \(\lfloor x \rfloor\) indicates the smallest integer greater than or
equal to \(x\). Though \(L_n/M_n = L_1 L_2 \cdots L_{n-1}/M_1 M_2 \cdots M_n\)
\([Eq. (13)]\), Eqs. \((16)\) and \((17)\) do not always give the same
results. However, the differences are always small or
zero for typical resampling factors. To compare proper-
ties and results in different architectures, we truncate the
outputs at the smaller of the two sizes those given by
\((16)\). The two-dimensional \(n\)th outputs contains \(S_n^2\) sam-
ple.

**Direct Form Filters.** To calculate the computational
complexity of the MRB, we total the number of multipli-
cations and additions in all the stages. In the \(n\)th stage of
a parallel MRB, \(y_0\) is first expanded by \(L_n\) and filtered by
\(h_n\). With direct form filters, the stage performs a total of
\(S_0 L_n t_n\) multiplications, or \(L_n t_n\) multiplications per input
sample (with appropriate padding of the input signal at
the edges). Taking advantage of symmetry in the odd-
length Gaussian filters, there are only \(L_n(t_n - 1)/2 + 1\)
multiplications per input sample. Thus, the total number of
multiplications per input sample in the parallel MRB is

\[
m_{df,p} = \sum_{k=1}^{N} L_k (t_k - 1)/2 + 1.
\]

(18)

In this notation, "df" or "tv" stand for direct form or
time-varying filters, and "p" or "h" stand for the parallel
or hierarchic architectures. The number of additions in a
filter is one less than the number of filter coefficients,
thus the total number of additions is

\[
a_{df,p} = \sum_{k=1}^{N} L_k (t_k - 1).
\]

(19)

As discussed below, the number of filter coefficients de-
deps only on the architecture (parallel vs. hierarchic)
and not on whether the filters are direct form or time-
varying. The total number of coefficients is the sum of the
number of independent coefficients in all filters:

\[
c_p = \sum_{k=1}^{N} \left\lfloor \frac{(t_k - 1)}{2} + 1 \right\rfloor.
\]

(20)

where again we have taken advantage of filter symmetry.

For two-dimensional signals, the results are similar.
Since the filters are separable, the \(S_0 L_n(t_n - 1)/2 + 1\)
multiplications for the one-dimensional signal are re-
peated \(S_0\) times along one axis and \(S_0\) times along the
other. This gives \(2S_0 L_n(t_n - 1)/2 + 1\) multiplications
for the \(n\)th stage, or \(2L_n(t_n - 1)/2 + 1\) multiplications
per input sample. Thus, for two-dimensional signals, the
number of multiplications and additions (Eqs. \((18)\) and
\((19)\) are doubled. However, since the same filter is used
along each axis, the number of filter coefficients \((20)\) re-
mains the same.

For the one-dimensional, hierarchic MRB, the \(n\)th
stage performs \(S_{n-1} L_n(t_n - 1)/2 + 1\) multiplications. To
normalize by the number of samples in \(y_0\), we approxi-
mate \((16)\) with \(S_n/S_0 = L_n/M_n\). This approximation dif-
fers from \((16)\) by at most \(1/S_0\) and is excellent, since \(S_n \gg
1\). For the first stage, dividing out \(S_0\) gives \(L_1(t_1 - 1)/
2 + 1\) multiplications per input sample. For all other
stages, there are \((L_{n-1}/M_{n-1}) L_n(t_n - 1)/2 + 1\) multipli-
cations per input sample. Thus, the total number of multipli-
cations and additions per input sample in the one-di-
Mensional hierarchic MRB is

\[
m_{df,h} = L_1 \left\lfloor \frac{t_1 - 1}{2} + 1 \right\rfloor + \sum_{k=1}^{N} \frac{L_k - 1}{M_k - 1} L_k \left\lfloor \frac{t_k - 1}{2} + 1 \right\rfloor.
\]

(21)
TABLE 2
Properties for Direct form and Time-Varying Implementations of Equivalent Parallel and Hierarchic MRBs ($N = 4$):
Two Examples Are Shown; the Signal Is One-Dimensional

<table>
<thead>
<tr>
<th>$L$</th>
<th>$M$</th>
<th>$t$</th>
<th>Architecture</th>
<th>No. of Coefficients</th>
<th>Direct form</th>
<th>Time-varying</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No. of</td>
<td>No. of</td>
<td>No. of</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>coeff/s</td>
<td>coeff/s</td>
<td>coeff/s</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>9</td>
<td>Parallel</td>
<td>364</td>
<td>4,886</td>
<td>19.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hierarchic</td>
<td>20</td>
<td>24</td>
<td>7.22</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>9</td>
<td>Parallel</td>
<td>884</td>
<td>61,080</td>
<td>15.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hierarchic</td>
<td>20</td>
<td>41</td>
<td>6.15</td>
</tr>
</tbody>
</table>

where the “2D” indicates the result is for two-dimensional signals. As in the parallel case, the number of filter coefficients is the same for one and two-dimensional signals.

Table 2 shows these properties for two example MRBs. In all cases, the MRBs are based on a hierarchic MRB having $t_n = t$, $L_n = L$, and $M_n = M$ for $n = 1, \ldots , N$, the most common type of hierarchic MRB used. As shown in the table, $m_{df,p}$ and $a_{df,p}$ exceed $m_{df,h}$ and $a_{df,h}$ by several orders of magnitude, because the parallel filters are so much larger than the equivalent hierarchic filters. The ratios of $c_{df,p}$ to $c_{df,h}$ also reflect this difference. The parallel values are larger because the parallel MRB does in single steps the same filtering that the hierarchic MRB does incrementally. Table 3 shows that the relative values for the two-dimensional MRB properties are similar to those for the one-dimensional case.

**Time-Varying Filters.** Direct form filters can be useful since they are easy to design and require little overhead computation. However, they are highly inefficient for multirate filtering. Because the filtering in each stage is performed after the sampling rate expander, where the sampling rate is highest: (i) the $L_n - 1$ zeros inserted between input samples by the expander (in a parallel-type stage) are multiplied by the filter weights, even though the zeros do not influence the outputs, and (ii) $M_n - 1$ of

TABLE 3
Properties for Direct form and Time-Varying Implementations of Equivalent Parallel and Hierarchic MRBs ($N = 4$):
Two Examples Are Shown; the Signal Is Two-Dimensional

<table>
<thead>
<tr>
<th>$L$</th>
<th>$M$</th>
<th>$t$</th>
<th>Architecture</th>
<th>No. of Coefficients</th>
<th>Direct form</th>
<th>Time-varying</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No. of</td>
<td>No. of</td>
<td>No. of</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>coeff/s</td>
<td>coeff/s</td>
<td>coeff/s</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>9</td>
<td>Parallel</td>
<td>364</td>
<td>9,772</td>
<td>39.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hierarchic</td>
<td>20</td>
<td>35</td>
<td>10.38</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>9</td>
<td>Parallel</td>
<td>884</td>
<td>122,160</td>
<td>31.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hierarchic</td>
<td>20</td>
<td>62</td>
<td>9.26</td>
</tr>
</tbody>
</table>
every $M_n$ outputs $h_n$ calculates are ignored by the compressor.

Time-varying filters are much more efficient [8, 27]. At the cost of some overhead computation, they ignore the inserted zeros and only calculate the samples passed by the compressor. They achieve these savings by cycling among $L_n$ different impulse functions, each composed of different samples from the direct form filter $h_n$. These impulse responses are either $[t_n/L_n]$ or $[t_n/L_n]$ samples in length, where $[x]$ indicates the largest integer less than or equal to $x$. Their average size is $t_n/L_n$, which we use below as filter length. Thus, the number of multiplications in the $n$th stage is $(S_nL_n/M_n)(t_n/L_n)$, or $t_n/L_n$ multiplications per input sample, where $S_nL_n/M_n$ is the number of output samples computed. Because the $h_n$ is asymmetrically broken into $L_n$ functions, it is not possible to take advantage of its symmetry. The number of additions for the $n$th stage is $(S_nL_n/M_n)(t_n/L_n - 1)$, or $(L_n/M_n)(t_n/L_n - 1)$ additions per input sample. For the time-varying, one-dimensional, parallel MRB, the total number of multiplications and additions per input sample is

$$m_{tv,p} = \sum_{k=1}^{N} t_k/M_k$$

$$a_{tv,p} = \sum_{k=1}^{N} L_k/M_k (t_k/L_k - 1).$$

For two-dimensional parallel MRBs, the number of multiplications and number of additions are double the one-dimensional values. Since the $L_n$ impulse functions are composed of samples from the direct form filter, for both the one- and two-dimensional time-varying filter, parallel cases, the total number of independent filter coefficients stored is the same as for the direct form, parallel case. For the time-varying, one-dimensional hierarchic MRB, the $n$th stage performs $(S_{n-1}L_n/M_n)(t_n/L_n)$ multiplications and $(S_{n-1}L_n/M_n)(t_n/L_n - 1)$ additions per input sample. Thus, the number of multiplications and additions per input sample are

$$m_{tv,h} = t_1/M_1 + \sum_{k=2}^{N} (L_{k-1}/M_{k-1})(t_k/M_k)$$

$$a_{tv,h} = (L_1/M_1)(t_1/L_1 - 1)$$

$$+ \sum_{k=2}^{N} (L_{k-1}/M_{k-1})(L_k/M_k)(t_k/L_k - 1).$$

For two-dimensional signals, the results are

$$m_{tv,h,2D} = 2t_1/M_1 + 2 \sum_{k=2}^{N} (L_{k-1}/M_{k-1})^2 t_k/M_k$$

$$a_{tv,h,2D} = 2(L_1/M_1)(t_1/L_1 - 1)$$

$$+ 2 \sum_{k=2}^{N} \left( \frac{L_{k-1}}{M_{k-1}} \right)^2 (L_k/M_k)(t_k/L_k - 1).$$

The total number of filter coefficients stored for both the one- and two-dimensional time-varying, hierarchic cases is the same as for the direct form, hierarchic case.

Table 2 and 3 show these properties for the same MRBs as used in the direct form case. As shown in the tables, $m_{tv,p}$ and $a_{tv,p}$ exceed $m_{tv,h}$ and $a_{tv,h}$, but by a much smaller factor than in the direct form case. While the parallel filters are still much larger than the equivalent hierarchic filters, the efficient time-varying implementation lessens the significance of this difference.

CONCLUSION

This paper gives closed-form conversions between parallel and hierarchic scaled Gaussian MRBs and calculates properties of these architectures. We have shown that the hierarchic architecture always requires fewer multiplications, additions, and coefficients than the equivalent parallel architecture. The hierarchy's advantages are much smaller for time-varying filters than for direct form filters. However, hierarchic MRBs can suffer from the propagation and accumulation of quantization and aliasing error in successive stages. Additionally, in hardwired implementations, while the hierarchic MRB requires far fewer hardwired multiplications, the parallel MRB operates more rapidly [25]. Thus, despite its greater complexity, there are applications for which the parallel architecture may be more useful. Conversion allows taking advantage of these and other differences between the architectures.

APPENDIX

Approximation of Discrete Time Fourier Transform of a Gaussian

The Discrete Time Fourier Transform of the Gaussian $\exp(-x^2/\sigma^2)$ is

$$\sqrt{\pi} \sigma \sum_{m=-\infty}^{\infty} \exp \left( -\frac{\sigma^2(\omega - 2\pi m)^2}{4} \right).$$

Equation (32) is the sum of an infinite number of Gaussians displaced by $\pm 2\pi m$. We are interested in the minimum $\sigma$ such that the predominant contribution to (32) in the region $R = (-\pi \leq \omega \leq \pi)$ is the $n = 0$ Gaussian. To simplify the derivation, we drop the coefficient $\sqrt{\pi} \sigma$. 

Let \( V_1 \) be the area under the center Gaussian in \( R \)
\[
V_1 = \int_{-\sigma}^{\sigma} \exp(-\sigma^2 \omega^2/4) \, d\omega = 2 \sqrt{\pi} \text{erf}(\pi \sigma/2)/\sigma \tag{33}
\]
where \( \text{erf}(\cdot) \) is the error function and \( V_2 \) be the area in \( R \) due to the other Gaussians
\[
V_2 = \sum_{n=n_{0}}^{\infty} \int_{-\sigma}^{\sigma} \exp \left( -\frac{\sigma^2 (\omega - 2\pi n)^2}{4} \right) \, d\omega \tag{34}
\]
We define \( V_2 = V_2^+ + V_2^- \) where \( V_2^+ \) is the sum in (34) over \( n = 1 \) to \( \infty \), and \( V_2^- \) is the sum over \( n = -1 \) to \(-\infty \). Due to symmetry, \( V_2 = 2 \cdot V_2^+ \). The problem can be restated as finding the minimum \( \sigma \) such that
\[
\left( V_1 + V_2 \right)/V_1 = \left( V_1 + 2 \cdot V_2^+ \right)/V_1 = 1 + 2^\delta \quad 0 < \delta \ll 1
\]
For \( n > 0 \), the exponents in (34) are maximized in \( R \) at \( \omega = \pi \) and \(-\pi \). Thus, an upper bound for \( V_2^+ \) is
\[
V_2^+ < 2\pi \sum_{n=1}^{\infty} \exp \left( -\frac{\sigma^2 (\pi - 2\pi n)^2}{4} \right)
= 2\pi \exp(-\sigma^2\pi^2/4) \sum_{n=1}^{\infty} (\exp(\sigma^2\pi^2))^{-n+1}. \tag{36}
\]
Since \( \sigma \) and \( \pi \) are both positive, \( \exp(\sigma^2\pi^2) > 1 \). Noting that \( \sum_{n=1}^{\infty} x^{n-1} = 1 + 1/(1-x) \) for \( x > 1 \) and \( n > 2 \), substitution into (36) gives
\[
V_2^+ < 2\pi \exp(-\sigma^2\pi^2/4) \left[ 1 + \sum_{n=1}^{\infty} (\exp(\sigma^2\pi^2))^{-n} \right]. \tag{37}
\]
Applying the identity \( \sum_{n=0}^{\infty} x^n = 1/(1-x) \) for \( |x| < 1 \) to (37) gives
\[
V_2^+ < 2\pi \exp(-\sigma^2\pi^2/4) \left[ \frac{1}{1 - \exp(-\sigma^2\pi^2)} - \exp(-\sigma^2\pi^2) \right]. \tag{38}
\]
Finally, substituting for \( V_1 \) and \( V_2^+ \) in (35) and simplifying gives
\[
\frac{2\sigma \sqrt{\pi}}{\text{erf}(\pi \sigma/2)} \left[ \frac{1}{1 - \exp(-\sigma^2\pi^2)} - \exp(-\sigma^2\pi^2) \right] \leq \delta. \tag{39}
\]
For \( \delta = 0.001 \), \( \sigma \) must be greater than approximately 1.8981. Thus, the approximation to (5) is excellent when \( \sigma B_\pi \geq 1.8981 \). By way of example, when \( \sigma = 1.8981 \), \( V_1 = 3.735 \). The contribution of the \( n = 1 \) Gaussians to \( V_2 \) is \( 4.634 \times 10^{-5} \). For \( n > 1 \), the contributions are even more negligible. This approximation is quite useful, since a Gaussian with standard deviation 1.8981 drops to below 0.001 its peak value only five samples from the origin.

REFERENCES


---

**BENNETT LEVITAN** received the B.S. (academic honors and distinction) in electrical engineering from Columbia University in 1985 and the M.S. degree in bioengineering from the University of Pennsylvania in 1987. He is presently pursuing a joint degree combining medicine and a Ph.D. in bioengineering at the University of Pennsylvania. He was awarded the Measey Foundation Combined Degree Scholarship in 1986 and the National Institutes of Health Medical Scientist Training Program in 1987. His research interests include multirate filtering theory and its application to retinal modeling, artificial life, and genetic algorithms. Mr. Levitan is a member of Tau Beta Pi, the Optical Society of America, ARVO, and IEEE.

**GERSHON BUCHSBAUM** received the B.Sc. and M.Sc. degrees in electrical engineering with distinction in 1974 and 1975, and the Ph.D. degree in 1978, all from the Tel Aviv University, Tel Aviv, Israel. Since 1979 he has been with the Department of Bioengineering at the School of Engineering and Applied Science, University of Pennsylvania, Philadelphia. In 1984 he was awarded the Presidential Young Investigator Award from the National Science Foundation. His research interests are concentrated in human vision, particularly image coding and analysis by the visual system. Dr. Buchsbaum is a member of IEEE, the Optical Society of America, and the Association for Research in Vision and Ophthalmology.
Group 3: Color constancy and interactions of space and color


Appendix H: A Multi-Stage neural network for Color Constancy and Color Induction, IEEE Transactions on Neural Networks, in press, 1995
Network Simulations of Retinal and Cortical Contributions to Color Constancy

SUSAN M. COURTNEY,*† LEIF H. FINKEL,∗‡ GERSHON BUCHSBAUM*

Received 10 May 1993; in revised form 28 December 1993; in final form 2 June 1994

A biologically-based neural network simulation is used to analyze the contributions to color perception of each of several processing steps in the visual system from the retina to cortical area V4. We consider the effects on color constancy and color induction of adaptation, spectral opponency, non-linearities including saturation and rectification, and spectrally-specific long-range inhibition. This last stage is a novel mechanism based on cells which have been described in V4. The model has been tested with simulations of several well known psychophysical color constancy and color induction experiments. We conclude from these simulations the following: (1) a simple push-pull spectrally specific contrast mechanism, using large surrounds analogous to those found in V4, is very effective in producing general color constancy and color induction behavior; (2) given some spatio-temporal averaging, receptor adaptation can also produce a degree of color constancy; (3) spectrally opponent processes have spatial frequency dependent responses to color and brightness contrast which affect the contribution of the V4 mechanism to color constancy in images with nonuniform backgrounds; and (4) the effect of the V4 mechanism depends on the difference between center and surround while the effect of adaptation depends on the total sum of inputs from both center and surround and therefore the two stages cooperate to increase the range of stimulus conditions under which color constancy can be achieved.

Color constancy Color induction V4 Adaptation

INTRODUCTION

Human color perception is not a simple function of the wavelengths of light reflected from a small area on a single surface. Instead, color depends on the spatial distribution of the wavelengths of light present in the entire image. The two most common phenomena which demonstrate this dependence are color constancy and color induction. Color constancy is the tendency of the colors of surfaces to remain more constant than would be suggested by the physical composition of the reflected light under changing illuminance conditions. It is thought that color constancy contributes to object recognition by allowing more reliable judgments about the object's surface properties regardless of the ambient light. A related phenomenon, color induction, is the change in the color of a surface due to its juxtaposition with other colored surfaces. Color induction enhances the color contrast in a scene and probably aids in object detection and surface segmentation.

Color constancy has been the subject of investigation for many years and a large variety of approaches have been attempted. Some were based on learning and judgment (e.g. Helmholtz, 1866; review by Jameson & Hurvich, 1989). Others have attempted to explicitly separate the reflectance from the illuminant by either computational theory (e.g. Buchsbaum, 1980; Maloney & Wandell, 1986; Rubin & Richards, 1982; D'Zmura & Lennie, 1986; Gershon & Jepson, 1989; Brainard & Wandell, 1991; D'Zmura & Iverson, 1993a, b) or linear filter theory (Faugeras, 1979). Additional well known theories include Land's Retinex (Land & McCann, 1971), various adaptation mechanisms (e.g. Hering, 1878; Helson, 1938; Judd, 1940; Brill & West, 1986; Brainard & Wandell, 1992), and spectrally-specific contrast based algorithms (Lucassen & Walraven, 1993). Most of these approaches attempted to identify one particular mechanism for achieving color constancy, or emphasized the importance of the contribution of one mechanism over another.

This emphasis has resulted in a retina vs cortex debate. Many researchers point to the need for two types of processing, one slow and one fast, one multiplicative and one subtractive (e.g. Hayhoe, Benimoff & Hood, 1987) to explain color constancy and color induction data. However, the different computational properties of these biological processes with regard to their effects on color constancy and color induction were not extensively studied, nor has much been said about the interactions...
of these retinal and cortical processes. This paper attempts to examine these issues and to determine what advantage this multistage system has for producing color constancy.

Receptor adaptation and retinal spectrally opponent processes have been studied in depth, psychophysically, physiologically, and computationally, for their contributions to color processing and color coding. Interest in the cortical color mechanisms, particularly V4, has developed relatively recently and the results are more controversial. The first physiological evidence for the importance of the cortex in color constancy was reported by Zeki (1983) who recorded from individual cells in V4 whose responses, unlike those in V1, appeared to follow human color perception rather than wavelength. Several V4 lesion studies have had mixed results concerning the apparent effect of such lesions on color perception (Walsh, Kulikowski, Butler, & Carden, 1992; Heywood, Gadotti, & Cowey, 1992). Schein and Desimone (1990) reported that there are regions quite distant (up to 16 deg) from the classically-defined receptive fields of V4 cells which can influence a cell's response if the center of its classical receptive field is also stimulated. They called these regions silent surrounds. The existence of long-range lateral connections in V4 (Yoshioka, Levitt & Lund, 1992) and the dramatic reduction in ipsilateral surround suppression after section of the corpus callosum (Desimone, Moran, Schein & Mishkin, 1993) suggest that these large surrounds may be mediated by a mechanism within V4. The silent surrounds in V4 are sensitive to nearly the same wavelengths as the center of the receptive field, creating a spectrally-specific response which is functionally akin to "cone-specific contrast" (see Lucassen & Walraven, 1993). "Cone-specific contrast" appears, from psychophysical experiments, to be a necessary component of human color constancy (Tiplitz Blackwell & Buchsbaum, 1988b; Lucassen & Walraven, 1993; McCann, McKee, & Taylor, 1976). However, because of modifications to the cone inputs preceding cortical stages it is difficult to quantify the response of the V4 cells directly in terms of "cone-specific contrast".

Two psychophysical experiments, one using a split corpus callosum patient (Land, Hubel, Livingstone, Perry & Burns, 1983) and the other using binocularly fused stimuli (Shevell, Holliday & Whittle, 1992), demonstrate a significant influence from cortical processing in constancy and induction phenomena. In addition, regions significantly separated from the test area have been demonstrated by psychophysical experiments to be

---

**FIGURE 1.** Caption on facing page.
very influential in determining perceived color (Tiplitz Blackwell & Buchsbaum, 1988a; Valberg & Lange-Malecki, 1990; Wesner & Shevell, 1992). The spatial dimensions of these phenomena are too large to be easily explained by known retinal structures. In addition, the speed with which a significant portion of this effect occurs, rules out the combination of receptor adaptation and eye movements as the sole mechanism for long-range color induction.
In order to explore the effects of both retinal and cortical processing on color constancy and color induction, we simulated a multi-stage neural network which includes three processes: receptor adaptation, spectral opponency, and spectrally-specific long-range inhibition. Each stage includes a saturating and rectifying nonlinear response function. Neural networks have been used before for implementing a variety of color constancy algorithms: lightness algorithms similar to Retinex (Hurlbert & Poggio, 1988; Moore, Allman & Goodman, 1991), a color categorization method using double opponent cells (Dufort & Lumsden, 1991), and an algorithm which uses contrast across boundaries to fill-in enclosed regions (Grossberg, 1987). In these simulations, as in other color constancy studies, the emphasis has been on describing a specific mechanism for achieving color constancy. In the current network simulation, which includes a new mechanism for cortical level processing, the specific effects of each processing stage and the interactions between processes were controlled and observed. We will show that a system which includes both retinal and cortical processes can produce the general behavior of both color constancy and color induction. In addition, we will demonstrate that while the differences between the spatial and chromatic properties of these processes sometimes leads to complex interactions between stages, all of these processes cooperate so that together they can produce greater contrast sensitivity and color constancy in a larger range of stimulus conditions than can any of the stages alone.

**NETWORK ARCHITECTURE**

An overview of the model is shown in Fig. 1(a). The cortical mechanism is shown in greater detail in Fig. 1(b). The network was simulated using NEXUS, an interactive neural simulator designed for large scale models (Sajda & Finkel, 1992). The complete network consists of over 11,000 cells and approx. 1.65 million connections. Below we will describe how each stage was implemented in the simulation. Table I summarizes the most significant parameters in the model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Factors in choice of parameter value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_c$</td>
<td>Connection strength between cells</td>
<td>Chosen to create receptive field shapes found physiologically, different for each cell type</td>
</tr>
<tr>
<td>$\sigma_i$</td>
<td>Threshold of cell $i$</td>
<td>Chosen so that most inputs fall in middle of response range, different for each cell type, cone threshold changes with adaptation state</td>
</tr>
<tr>
<td>$\beta_i$</td>
<td>Slope of linear portion of cell's response function</td>
<td>Chosen in combination with $\sigma_i$ to give the appropriate dynamic range for each processing stage, different for each cell type</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Width of adaptation weighting function</td>
<td>Small value for fixation or very short presentation time experiments, large value for experiments with free eye movements ($\theta \approx 3.0$, relatively small compared to cortical silent surrounds, large compared to center of spectrally opponent receptive fields $0 &lt; \theta &lt; \text{diameter of image}$)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Fraction of total long term adaptation achieved</td>
<td>Dependent upon length of viewing time ($\alpha = 0.2; 0 &lt; \alpha &lt; 1$)</td>
</tr>
<tr>
<td>$c_1, c_2$</td>
<td>Coefficients for push-pull mechanism</td>
<td>Chosen together with $x$ to give a total average constancy shift of 20% in accordance with psychophysical data ($c_1 = c_2 = 0.25; 0 \leq c_1, c_2 \leq 1$)</td>
</tr>
</tbody>
</table>

(i) Input

The first stage corresponds to the cone responses. The input image is a $27 \times 27$ array, in which each entry defines the color at that location. The array is converted to three $27 \times 27$ arrays of cone activation levels: $R, G, B$. Therefore, an input image unit has a corresponding set of three units (analogous to one cone of each type) in the first layer of the network. Each entry in the input image is specified either by a Munsell reflectance spectrum and an illuminant spectrum, or in CIE notation $(x, y, Y)$.

When the reflectance and illuminant spectra were specified, the image was converted, at each point, to the three normalized cone activation levels by using the Vos–Walraven (Vos & Walraven, 1971; Vos, 1978) cone action spectra $[r(\lambda), g(\lambda), b(\lambda)]$, in steps of 10 nm:

$$ R = \sum_{\lambda = 400}^{700} k_i r(\lambda) \mathcal{R}(\lambda) \mathcal{I}(\lambda) \Delta \lambda $$

$$ G = \sum_{\lambda = 400}^{700} k_i g(\lambda) \mathcal{R}(\lambda) \mathcal{I}(\lambda) \Delta \lambda $$

$$ B = \sum_{\lambda = 400}^{700} k_i b(\lambda) \mathcal{R}(\lambda) \mathcal{I}(\lambda) \Delta \lambda $$

where $\mathcal{R}(\lambda)$ is the reflectance spectrum, a fixed property of the surface, and $I(\lambda)$ is the illuminant, which may change with the particular viewing condition and, therefore, may change the (perceived) color of the surface. (Because inputs are computed from the reflectance and no other surface properties are considered, we will refer only to the reflectance spectra, not to a real or simulated surface.) The coefficients $k_{1,2,3}$ are constants which normalize the sensitivity spectra so that all cone types in the simulated array have the same peak sensitivity. Therefore, the three types of first layer units ("cones") have responses of the same order of magnitude and we designed the matching procedure to depend upon the relative responses of the three simulated, color pathways [Section (vi)]. For those cases in which the image was specified in CIE notation, the image was converted to cone activation levels by applying the transformations for Vos–Walraven action spectra (Vos, 1978; Wyszecki & Stiles, 1982, p. 615) and then normalizing using the same coefficients $k_{1,2,3}$. 

![Figure 1(a)](image-url) An overview of the model is shown in Fig. 1(a).

![Figure 1(b)](image-url) The cortical mechanism is shown in greater detail in Fig. 1(b).
(ii) Cell responses and nonlinearities

In the simulation of the network model, the total input to cell, \( Q_i \), is determined by a weighted sum of the activities of all cells connected to cell \( i \):

\[
Q_i = \sum_{j=1}^{n} \omega_{ij} A_j
\]

where \( A_j \) is the activity of cell \( j \), \( \omega_{ij} \) is the connection strength from cell \( j \) to cell \( i \). The cells of the network corresponding to the cone layer have a Naka–Rushton response function (Naka & Rushton, 1966):

\[
A_i = \frac{Q_i}{Q_i^* + \sigma_i} \quad (3)
\]

where \( x \) is a constant from 0.7 to 1.0. In the simulation results shown here \( x = 0.9 \). The general behavior of the system was not very sensitive to the value of this parameter. \( \sigma_i \) is the threshold of cell \( i \). The input, \( Q_i \), for a cone is the cone activation level \( R, G, \) or \( B \) calculated from the input image as described above in equation (1).

In all other stages, cell activity is determined by a sigmoidal response function of the input:

\[
A_i = (\text{max} - \text{min}) \left( \frac{1}{1 + \exp[-(Q_i - \sigma_i)\beta_i]} \right) + \text{min} \quad (4)
\]

where \( A_i \) is the activity of cell \( i \), max and min are the maximum and minimum possible activity levels for cell \( i \), \( \sigma_i \) is the threshold of the cell, and \( \beta_i \) is proportional to the slope of the linear portion of the curve (see Fig. 2).

(iii) Adaptation

We assume an initial long-term adaptation to a uniform neutral background [see Walraven, Enroth-Cugell, Hood, MacLeod, and Schnapf (1990) for a review of psychophysical and physiological studies on adaptation]. The amount of threshold shift \( \Delta \sigma \), is determined by the difference between the cone activation level for the neutral background stimulus and the cone activation level for the new stimulus. Because adaptation is dependent on the temporally weighted average of its input, the adaptation shift for a cone is dependent not only on the point in the image directly corresponding to that cone position, but also on the surrounding area to which the cone may be exposed during eye movements, or from optical blur. We approximated this temporal effect by a two-dimensional Gaussian spatial weighting function, because for the psychophysical experiments we were interested in studying, there was generally either a fixation point, or a central test patch around which one could assume eye movements were centered. In the simulation, the amount of the shift follows a sigmoidal function of the difference between the neutral and the current stimuli and is proportional to the length of viewing time. These constraints are incorporated into the simulation by calculating the threshold shift for the receptor adaptation by using the equation:

\[
\Delta \sigma = \alpha \left\{ 2M \left( \frac{1}{1 + \exp[-(Q_i - Q_{\text{neutral}})\beta_i]} \right) - M \right\}
\]

\[
M = \left[ \sum_{j=0}^{n} \frac{(Q_j - Q_{\text{neutral}})^2}{2 \pi \theta^2} \right] \exp \left[ -\frac{(x^2 + y^2)}{2 \theta^2} \right]
\]

FIGURE 2. Nonlinear response function of each cell. The parameters are set for each stage so that most stimuli produce responses in the linear range of this function. The slope of the linear portion of the curve is proportional to \( \beta \). Each cell’s input is the weighted sum of the activities of all the cells connected to it. \( \sigma \) is the “threshold” which is defined for mathematical clarity to be at the center of the linear portion of the response. Saturation and rectification occur when the cell’s output nears its maximum and minimum outputs respectively.

FIGURE 3. Response curves for cones in the simulation under a range of values for the adaptation threshold. (a) Shows the sigmoidal limits of the adaptation range. The luminance level of the adapting stimulus was increased linearly, but the threshold values reach an asymptote at both ends of the range. (b) Same as (a) in log-linear coordinates.
where \( \sigma \) is the threshold; \( \beta \) is the proportional to the slope of the linear portion of the function; \( Q \) is the cone activation level (i.e., R, G, or B) due to current image pixel \( i \); \( Q_{\text{new}} \) is the cone activation level due to standard neutral at image pixel \( i \); \( n \) is the number of pixels in the image; \( x, y \) is the horizontal and vertical distances from pixel \( i \) to the center of the cone's receptive field when fixated on the center of the image; \( \theta \) is the width of a Gaussian weighting function which varies with the degree of fixation required for the experiment; \( \alpha \) is the fraction achieved within the stimulus presentation time of the total difference in long-term adaptation states between the neutral state and the state for the new stimulus.

\( \alpha \) is proportional to the time of exposure. As \( \alpha \) increases, the size of the threshold shift increases, following a sigmoidal curve ranging from \(-M\) to \(+M\) where \( M \) is the difference between the weighted average activation level for the current image and the activation level for a uniform neutral background (see Fig. 3). In the current study, \( \alpha \) was held fixed at 0.3 and \( \theta \) was held fixed at 3.0. However, we wished to include this flexibility in the model because eye movements do affect the adaptation state. With longer exposure time, the cell will be able to better adapt (larger \( \alpha \)) to its new stimulus. Under certain experimental conditions, longer exposure time may also allow for more eye movements. The spatial extent of the weighting function broadens with more eye movements. In the extreme case of very long exposure time and completely random eye movements over the entire field of view, the weighting function would be flat and the cone would adapt to the field average. This dependence of the parameters \( \theta \) and \( \alpha \) on eye movements and viewing time, allows the effects of the adaptation stage of the simulation to vary with the experimental conditions being considered. This is important because the extent of eye movements in psychophysical experiments has been shown to affect color perception (Cornellissen & Brenner, 1991).

(iv) Spectral opponency

For the purpose of studying the effect of spectral opponency, we include only a single stage for this process, instead of the hierarchy of opponent cell types observed physiologically between the retina and V4. We wished to study the effects of spectral opponency as a mathematical operation rather than attempt to simulate the specific anatomical implementation. Opponency can occur at many levels of the visual system from cone gap junctions to the cortex (see review by Lenne & D'Zmura, 1988). We avoid the term "color opponency", because it has often been used in reference to psycho-

Opponent processing is achieved in the simulation by subtracting responses of spectrally opponent cone types and is generally based on the properties of LGN parvo-
cellular type I receptive fields. In the simulation, each "cell" receives excitatory input from a single cone in the center of its receptive field and inhibitory input from several cone types surrounding the center using a difference of Gaussians synaptic weighting function (Lenne & D'Zmura, 1988). The surrounds receive input from all cones in their receptive fields, however the synaptic weights are different for each cone type. The surround input is most heavily weighted toward the cone type(s) opponent to the center cone type. For example, opponent cells whose centers receive excitatory input from R cones receive inhibitory surround input from both R and G cones, but the amplitude of the synaptic weighting function for the G cones is twice that for the R cones. The opposite ratio was used for the G center cells. The R and G centered cells, thus, do not differ from each other just by a negative sign, but have linearly independent cone input combinations. B center cells receive inhibitory input which is equally weighted between the R and G cones. Altogether there are three linearly independent combinations. Off-center cells were created by using the same weighting functions, but with opposite sign, and their thresholds were lower than those of the on-center cells, giving them a higher spontaneous activity level. Therefore, the off-center cells responses were greatest when the magnitude of the stimulus in the center of the receptive field was less than that in the surround. The off-center cells of course do not add additional independent combinations to the three resulting from the on-center cells. In addition, primate retina and LGN cells do not have perfectly balanced centers and surrounds (Derrington & Lenne, 1984). Rather, the center strength (volume of two-dimensional Gaussian sensitivity profile) is roughly twice that of the surround, allowing these cells to have a significant response to homogeneous fields as well as to edges. Likewise, the spectrally opponent stage in the simulation has a 2:1 center/surround sensitivity ratio.

(v) Higher cortical processing

The next stage in the network is designed to respond according to the primary chromatic properties of the analogous cells in V4 (Schein & Desimone, 1990). These cells have large, suppressive surrounds each of which has a wavelength sensitivity similar to that of the center of the receptive field [see Fig. 1(b, c)]. These large surrounds had little or no effect on the cell's activity unless the center was also stimulated, and were therefore termed "silent surrounds". In the simulation, the "classical receptive field" (Schein & Desimone, 1990) receives excitatory input from a single class of spectrally opponent cells. These same type cells provide inhibitory input to the "silent surround" outside the classical receptive field. The "silent" behavior of the surrounds could be explained either by shunting inhibition (a multiplicative suppression of the excitatory input to a cell) or by rectified inhibition (the absence of effective inhibition in the resting state because of a very low spontaneous activity level). We chose to use rectified inhibition in the simulation because it is often found in cortical neurophysiological measurements while shunting inhibition appears to be rare in the cortex (Berman,
Douglas, Martin & Whitteridge, 1991). This is achieved by setting the thresholds of the silent surround cells so that the resting levels are very low. The effect of this rectification, together with the 2:1 center:surround weighting of the spectrally opponent cells, is to make the V4 cells in the simulation primarily dependent on the spectral sensitivity of the centers of the opponent cells which provide input to the cell. In this sense, the responses of the V4 cells in the simulation are measuring the difference in activity between the contributions of cones of the same type in the center and the surround. Therefore, we refer to the response of the V4 cells in the simulation as measuring spectrally-specific contrast.

Desimone, Schein and their colleagues (Moran, Desimone, Schein, & Mishkin, 1983; Desimone & Schein, 1987) reported that the effect of stimulation in the silent surround decreases with increasing distance from the classical receptive field. Psychophysical results also show a decrease in the effect of inducing regions with increasing distance (e.g. Tiplitz Blackwell & Buchsbaum, 1988a; Valberg & Lange-Malecki, 1990; Wesner & Shevell, 1992; Zaidi, Yoshimi, Flanagan & Canova, 1992). To incorporate these observations into the simulation, the inputs to the surround are weighted according to distance from the center by a negative exponential function [see Fig. 1(b)].

The strengths of the centers and silent surrounds of V4 cells appear to be well balanced; stimulation of the surround can completely inhibit the response to stimulation of the center (Schein & Desimone, 1990). Because the silent surround cells in V4 respond only when there is a difference, either in wavelength or luminance, between the center and the distant surround, these cells are particularly well suited for carrying information about contrast. However, for those images that have little spectrally-specific contrast, or an unknown or non-gray average chromaticity (e.g. blue sky, green forest), the d.c. (or local average chromaticity) information is also important. It is significant, therefore, that approx. 10% of the cells found in V4 did not have silent surrounds. The cells without silent surrounds have the same classical receptive field response as those cells with silent surrounds. These cells have the capacity to carry the (spatial) d.c. portion of the signal, i.e. to respond to homogeneous fields as well as edges and small spots. These center-only cells have been included in the network and we refer to them as "local reference cells" because they provide the normalizing reference information for the contrast cell responses.

The responses of analogous V4 stage "cells" in the simulation were created directly using the outputs of the spectrally opponent stage. A positive contrast cell receives its input, excitatory from the center and inhibitory from the surround, from on-center spectrally opponent cells. Therefore, the positive contrast cells respond to images for which the input to its classical receptive field is greater than the input to its silent surround. We have also included negative contrast cells which receive input from off-center cells, and therefore respond when the center input is less than the surround input. While, to our knowledge, there has been no systematic study of off-center cells in V4, given the symmetry of on- and off-populations of cells in earlier stages and the common observation that color constancy and color induction are seen in negative as well as in positive contrast stimuli, it seems reasonable to propose a negative contrast cell analogous to the positive contrast silent surround cells. Alternatively, the functions of both the negative and positive contrast cells in the simulation could be achieved by the V4 cells, also described by Schein and Desimone (1990), which had silent surrounds with both spectrally-specific inhibition and spectrally opponent excitation.

In order to combine the physiological information from the local reference and contrast cells into a simple set of outputs which could be compared to human color perception, we combined the outputs of these V4-like cells into a simple push–pull mechanism. [This stage is shown in Fig. 1(b).] We used one reference cell for every pair of positive and negative contrast cells. The output of this final network stage is determined by the response of the local reference cells, enhanced by the positive contrast cells, or inhibited by the negative contrast cells. This is given by the equation:

\[ O = B + c_1 P - c_2 N \]  

where \( O \) is the output, \( B \) is the local reference response, \( P \) is the positive contrast response, \( N \) is the negative contrast response, and \( c_1 \) and \( c_2 \) are constants. The constants \( c_1 \) and \( c_2 \) were chosen, together with \( x \) to give an average constancy shift of 20% of the distance between the color of the reflectance under the standard illuminant and the color of the reflectance under the test illuminant. This is consistent with psychophysical data (Tiplitz Blackwell & Buchsbaum, 1988b).

(ii) Matching procedure

After the image was processed by these three model stages, we needed to assess the input–output relationship in a manner similar to the psychophysical experiments. We, therefore, used a process analogous to the psychophysical matching paradigm (see Fig. 4). The final
output is a single set of three cells whose receptive fields are centered on the middle of the input image. (Because the sizes of the receptive fields increase with each subsequent stage in the network model, the dimensions of the network layers decrease progressively in order to reduce edge effects.) If the outputs for two different images are equal, then the centers of the two images are said to “match”.

In order to do this matching efficiently, for each set of simulation parameters, outputs were determined for 2625 colors [from Table I(6.6.1) of Wyszecki and Stiles (1982) which lists CIE coordinates for Munsell colors] using a standard background and illuminant. Unless otherwise indicated, reported matches were made using calculated input images corresponding to a single small square (3 x 3 input units) of a Munsell reflectance against a uniform gray (Munsell N6.0) background under CIE standard illuminant C. These standard outputs are then stored with their corresponding input values in a look-up table. Then, when the test image is shown to the network, its output is compared to the stored outputs for the standard images. The standard input color which corresponds to the stored output closest to the test image output is reported as the “match”.

SIMULATION RESULTS
(i) General constancy and induction abilities

We tested the network with various stimuli to determine how well it would follow human perception in the primary aspects of color constancy and color induction. The first simulated experiment tested brightness constancy and brightness induction. The center of the image was a single small patch (3 x 3 units) of the gray Munsell reflectance N6.75. The background of the first test image was Munsell reflectance N6.0. Constancy was tested using several different luminances of a spectrally flat illuminant. Matches were made using a N6.0 background and CIE standard illuminant C which gives a luminance of approx. 43 cd/m² for the N6.75 reflectance. Therefore, for the N6.75 center reflectance under other illuminants, perfect luminance constancy would be achieved if the matches also had a luminance of 43 cd/m².

The results are shown in Fig. 5. Because the chromatic changes under these conditions were small, only the luminance results are shown. The input luminances are shown by the black columns. The gray columns represent matches to the center of the first test image under the different levels of illuminant. For the N6.0 surround condition (the same surround used for the match condition), the match luminances were equal to the physical luminance (43 cd/m²) of N6.75 under the standard illuminant for all the different test illuminants, demonstrating brightness constancy. For the second test image a lighter background, N7.5, was used. The matches for this image are shown by white columns. Again, all illuminant conditions produce matches of the same luminance, demonstrating brightness constancy. However, the presence of the lighter surround shifts the luminance matches to a smaller value, the correct shift direction for brightness induction.

For color constancy, 10 different colored reflectance patches were used with three different illuminants. The reflectances were chosen, one of each Munsell hue, as a representative sample of Munsell chips of moderate luminances. One illuminant peaked at 440 nm, one at 560 nm and one at 660 nm. Again CIE standard illuminant C was used for the match condition. For both the match and test images, the background was Munsell reflectance N6.0. For most of the reflectance-illuminant pairs (25 out of 30), some degree of color constancy was obtained by the network. Figure 6 shows results for two of the 10 reflectance patches under the three colored illuminants.
illuminants and illuminant C. A match is considered as “achieving some degree of color constancy” if the difference (in color space) between the color of the match and the “true color” is less than the difference between the “true color” and the “physical color”. Both “true color” and “physical color” are defined by their computed coordinates in the RGB space described earlier. “True color” is the computed coordinates of the reflectance under standard illuminant conditions, and “physical color” is computed coordinates of the reflectance under the test illuminant. A “shift toward constancy” is a shift of the match toward the true color and away from the physical color. In the color constancy tests, the matches made by the network are somewhat color constant, but do not completely compensate for the illuminant change. The size of the constancy shift is different for each reflectance–illuminant pair and the amount of compensation can be varied by changing \( z, c_1, \) and \( c_2 \). However, we were not able to achieve perfect color constancy for all stimuli with any of the parameter combinations that we tried. This is not unexpected since human color “constancy” is also imperfect (see review in Beck, 1972; Tiplitz Blackwell & Buchsbaum, 1988b).

In a second test of color constancy, we simulated the McCann Mondrian experiment (McCann, McKee & Taylor, 1976). The experimental set-up is shown in Fig. 7(a). Two identical Mondrians were simulated, one under a standard neutral illuminant (CIE illuminant C) and the other under a combination of illuminants chosen so that the center colored patch [purple-blue in the example shown in Fig. 7(b)] under that illuminant would have the same \( R, G, B \) as a gray (N7.5) patch under the standard illuminant. Matches were made using the Mondrian as the background rather than the neutral
uniform field used in the other simulations [see Fig. 7(a)]. The color chosen for the center patch of the Mondrian under the standard illuminant to match the center of the Mondrian under the second illuminant, again, showed a tendency toward constancy, but not perfect compensation. For perfect constancy, the match would have to be identical to the color of the test patch under neutral illumination. For no color constancy, the match would have been equal to the color of the gray patch under neutral illumination.

Next, to test the spatial properties of color induction, we used small (3 × 3 units) reflectance patches surrounded by an annulus the width of which varied from 0 to 4 input units. The center patches and the surrounding annuli were separated by a neutral gap of 0 to 4 units in width. The diameter of the V4 surrounds in the simulation was 11 × 11 input units. Beyond the annulus, the background was the same neutral as the gap. The stimulus is shown in Fig. 8(a). As the width of the gap was increased, the amount of induction decreased [see Fig. 8(b)]. When the gap was 4 units wide, the annulus was outside the receptive field of the V4 cells and there was almost no induction. The induction effect did not disappear in the presence of a small gap as it would with a contrast mechanism which was highly localized. In addition, if the gap width is fixed and the width of the annulus is increased, the amount of induction increases [see Fig. 8(c)]. These results agree with those presented for the analogous psychophysical experiment by Tiplitz Blackwell and Buchsbaum (1988a).

The observation that induction is still noticeable when a neutral gap separates center and annulus, suggests that this same, large, spatially distributed spectrally-specific contrast mechanism could also account for the color context effects in psychophysical experiments by Wesner and Shevell (1992) in which they demonstrated that local contrast alone could not entirely account for color appearance. Wesner and Shevell used monochromatic lights to test color context effects, using the color cancellation method for a unique yellow center. The results for a simulation of these experimental conditions are shown in Fig. 9. The stimulus is shown in the figure inset. The stimulus used for the simulation consists of a central test spot (3 × 3 input units), an adjacent surrounding annulus (1 unit wide), and a distant surrounding annulus (3 units wide) immediately outside the adjacent annulus. The simulation was done using matches instead of cancellation, but the general results are the same. The results show that both areas adjacent to the test spot and distant areas affect the predicted color match. Green in either the adjacent or distant surround shifts the appearance of the center toward red. Red in the distant surround shifts the appearance of the center toward green. Increased luminance of the test spot relative to the surround luminance decreases the induction effect.

(ii) The roles of V4 and adaptation

To understand what each stage contributes to color constancy and color induction, we repeated several of these simulated experiments with various stages in the network eliminated or modified. By eliminating the adaptation stage, we found that many of the general properties of color constancy could be achieved by the cortical spectrally-specific push-pull mechanism alone.
Likewise, for most reflectance–illuminant pairs, adaptation alone also results in some degree of constancy. An example is shown in Fig. 10. However, each of these stages works best in different situations.

The size and direction of the color shift depends on stimulus conditions. An illuminant which causes a larger shift in the color signal (reflectance times illuminant) for the center test patch than for the surrounding background, such as a red test spot on a neutral background under red illuminant, causes an adaptation shift in the direction of constancy. However, the V4 stage for this stimulus increases the illuminant’s influence rather than decreasing it because the R cone contrast is positive. On the other hand, an illuminant which causes a larger shift in the background than in the test spot, such as a blue illuminant on a yellow test spot with a neutral
FIGURE 9. Results for the simulation of a color context stimulus, shown at the top right. The test spot is yellow, the adjacent surround is green (540 nm), and the distant surround is either green, red, or white. The figure shows the change in the (R/G) ratio from the neutral surround condition to the match for the yellow spot with the various colored surrounds. The results show that both areas adjacent to the test spot and distant areas affect the predicted color match. Green in either the adjacent or the distant surround increases the R/G ratio of the match, while red decreases the ratio. Increased luminance of the test spot relative to the surround luminance decreases the induction effect.

FIGURE 10. Color constancy simulation results plotted as in Fig. 5. The left graph shows the matches obtained when the V4, spectrally-specific contrast stage is eliminated and, therefore, only adaptation contributes to color constancy. The right graph shows the matches when adaptation is eliminated and, therefore, only the cortical spectrally-specific contrast mechanism contributes.
background, results in a large shift toward constancy at the V4 stage. In fact, the V4 stage overcompensates, causing color induction. The adaptation stage contribution depends on the degree of localization. Highly localized adaptation (perfect fixation) in this case results in almost no constancy shift, while less localized adaptation does cause a shift toward constancy.

The main reason for the difference in the color constancy contributions of these two stages can be seen in the spatial sensitivity profiles of each mechanism (see Fig. 11). The adaptation stage sums its input across both the test spot and the background; it is not spatially opponent. Whether most of the contribution is from the test spot or the background depends on how localized the adaptation is. The cortical contrast cells, on the other hand, receive antagonistic inputs from center and surround. Therefore, the effect of the V4 mechanism will depend on the difference between center and surround while the effect of adaptation will depend on the sum of inputs from both center and surround.

In situations where both stages, separately, would be effective in producing color constancy, their effects are sometimes antagonistic. Localized adaptation can decrease the contrast of the inputs to the center and surround of the V4 cells, making the V4 stage less effective. In some such cases, the size of the constancy shift with both stages is actually less than for either stage alone. However, the multi-stage system is more

FIGURE 11. Two stimulus conditions, each of which favors a different mechanism in the network for achieving color constancy. (a) A red spot will reflect a red illuminant more strongly than will a gray background of equal lightness. Therefore, the adaptation mechanism, which is most sensitive to the test spot will respond well to the red illuminant and provide good color constancy. The spectrally specific contrast mechanism, on the other hand, has a positive contrast response, and therefore, enhances rather than diminishes the effect of the illuminant. (b) The opposite stimulus condition. The blue illuminant is reflected best by the background. This leads to little response from the adaptation mechanism but a good response from the negative contrast cells in the final layer of the network.
consistent than either stage alone, because for the cases in which one of the stages alone would fail to produce constancy, the other stage can generally compensate. We tested the network with 10 colored test spots on a neutral background under 3 illuminants, as described earlier. Without the V4 stage, the system shifts the match toward constancy for 20 of 30 stimuli. Without the adaptation stage (but with V4) the system succeeds for 22 out of 30. With all stages included, the system achieves some degree of constancy for 25 out of 30 stimuli. For four out of the five stimuli for which the complete system does not achieve constancy, neither adaptation alone nor the V4 mechanism alone could produce constancy. The complete system is capable of producing color constancy in a broader range of stimulus conditions than can be handled by either stage alone. The complete system, therefore, also has a slightly better average color constancy performance.

We wanted to have some quantitative measure with which to compare the amount of constancy achieved by each of the stages in the model. Although \((R^2 + G^2 + B^2)^{1/2}\) (where \(R\), \(G\), and \(B\) are the normalized cone activation levels) cannot be considered a true measure of "color distance" because \(R\), \(G\), and \(B\) are not orthogonal and also because the "distances" do not correspond to perceptual distances, it is a good intuitive measurement and incorporates both color and brightness. CIELUV color differences, \(\Delta E^*\), is a less intuitive measure, but one which does correspond to perceptual distances (Wyszecki & Stiles, 1982, p. 166). \(\Delta E^*\) was also computed for each input-match pair and these numbers gave similar results. Figure 12 shows histograms of the \((R^2 + G^2 + B^2)^{1/2}\) "distances" from the actual matches under various colored illuminants to the ideal constancy match. The combination of both adaptation and V4 results in both a slightly smaller mean distance and a smaller range of distances than either stage alone.

(iii) Spectrally opponent vs spectrally-specific stages

The spectral sensitivities and center-surround organization of receptive fields in the opponent stage modify the inputs to the spectrally specific cortical stage. The effect that this intermediate stage has on the final output depends on the spatial structure and spectral composition of the input image (i.e. the segment sizes, spatial frequency content, number of edges, amount of chromatic and luminance contrast at the edges). In the following section we examine the effects that the opponent stage has on the input signal that it provides to the final stage of the network, and the effect that these modifications have on the output of the network.

Responses to high and low spatial frequency stimuli. If a low spatial frequency input (such that center and surround of the receptive field receive approximately the same input) to a spectrally opponent R–G cell changes in color, from yellow to red, without changing in luminance, the cell will receive both an increase in excitation and a decrease in inhibition. Spectral opponency, therefore, results in a high gain for low spatial frequency purely chromatic signals. On the other hand, the response to a low spatial frequency luminance stimulus will be attenuated because the increase (or decrease) in excitation will be offset by the increase (or decrease) in inhibition. At high spatial frequencies, this response relationship is reversed for cells which are spatially as well as spectrally opponent. A cell whose inhibitory surround falls partially on the darker side of a luminance edge will receive less inhibition than a cell which has both center and surround receiving input entirely from the higher luminance region (see Fig. 13). This center–surround receptive field structure, therefore, leads to enhancement of the cells responses to luminance edges. This is shown by the response of the spatially opponent cells in the network simulation.

On the other hand, at an equiluminant chromatic edge, a spectrally and spatially opponent cell may receive more inhibition from a surround which receives input partly from the other side of the color edge, if the surround is more sensitive to that color. As the response of the spectrally and spatially opponent layer of the network shows, this increase in inhibition results in a blurring of the chromatic edge response, an attenuation of high chromatic spatial frequencies (see Fig. 13). Because the V4-type spectrally-specific contrast cell has a very large receptive field, both the high and low spatial frequency responses of the spectrally opponent cells, which comprise the input to the V4 cell, are linearly summed.

The effect on color induction. There has recently been some discussion in the field regarding the effect of the image spatial structure on color induction. Valberg and Lange-Malecki (1990) presented evidence that the color induction shift caused by a Mondrian background was the same as the induction caused by a homogeneous background whose chromaticity and luminance were equal to the spatially weighted average of the Mondrian background. This homogeneous background was termed...
the "equivalent surround". Two additional psychophysical studies have since shown that perhaps the equivalent surround calculation must include some edge enhancement before the spatial average is computed (Brown, 1993; Wesner & Shevell, 1993). Because the opponent stage of the network causes luminance edge enhancement and the silent-surround stage calculates a spatially weighted average, we expected the simulation to show similar behavior.

We tested this hypothesis with the current simulation by using several input images whose surrounds had identical average color and luminance properties, but had an increasing number of high frequency edges. The stimuli and the results are shown in Fig. 14. The equivalent surround hypothesis predicts that such surrounds would have identical induction effects on the center test patch. The simulation outputs showed only very small changes with increased number of edges in the surround if the edges were purely chromatic. As explained above, a chromatic edge is not enhanced by the spectrally opponent cells. There was also no significant change when the luminances and saturations of all regions in the surround were such that all of the opponent cells were operating in the linear range of their response functions. However, there was a change in the output when some of the regions in the image produced responses outside the linear range of the opponent cells. For these images, the edge enhancement caused by the opponent cells was not symmetric across the edges. Therefore, the spatially weighted surround calculated by the silent-surround cells was different for each of the different images.

The effect on color induction of high spatial frequencies in both color and luminance has also been shown psychophysically in a different paradigm. Zaidi et al. (1992) showed an attenuation in the magnitude of color induction when an equiluminant surround included high spatial frequencies. Shevell and Wesner (1990) found a larger decrease in the magnitude of color induction when a thin white ring, equiluminant with the surround color, was placed in the surround, than when a black ring was placed in the surround. Zaidi et al. (1992) argue that this could also be explained by an attenuation of color induction by high spatial frequency chromatic signals in the inducing surround. However, neither Zaidi et al. (1992) nor Wesner and Shevell (1990) found this
attenuation when the high spatial frequencies in the inducing surround were due to luminance changes.

This behavior is also shown by the current network. Although the V4 stage linearly sums, its inputs from spectrally opponent cells, those inputs depend nonlinearly on the spatial frequency properties of color and luminance variations in the image. High spatial frequencies in color cause an attenuation of the color signal at the spectrally opponent layer, while high spatial frequencies in luminance are enhanced. Therefore, if the color regions within the inducing surround are equiluminant, the presence of high spatial frequencies will attenuate the input to the spectrally specific contrast stage and subsequently will reduce the amount of induction relative to that induced by a homogeneous surround.

To test this, we used an input image similar to that used by Wesner and Shevell (1990), a yellow test spot with either a red surround or a green surround. The red surround was either spatially homogeneous, or contained a thin ring around the test spot which was either black or a white which was equiluminant with the surround.

The results are shown in Fig. 15. The presence of the thin white ring in the surround significantly diminishes the color induction effect on the yellow center. However, the thin black ring, causes much less decrease in the color in output for different surround spatial structures.
FIGURE 15. (a) Final output of the network for a yellow center stimulus with either a red or green surround. The surrounds were either homogeneous, contained a thin black ring, or contained a thin white ring which was equiluminant with the surround. (b) Spatial configuration of the stimulus. (c) Response of the R on-center spectrally and spatially opponent layer of the network to the yellow center, red background stimulus which includes a thin white ring around the center which is equiluminant with the red surround. (d) Response of the R on-center spectrally and spatially opponent layer of the network to the yellow center, red background stimulus which includes a black ring in the surround. The network shows an enhanced response to the red surround when the black ring is present, while the white ring induces no such response. The enhanced response around the black ring provides more input to the induction mechanisms of the V4 layer which compensates for the loss of input from the introduction of the black ring. Therefore, the white ring decreases the induction effectiveness of the red surround, but the black ring does not.
induction. The equivalent surround hypothesis (Valberg & Lange-Malecki, 1990) implies that the black ring would cause a greater decrease in induction than the white ring. The chromaticities of spatially averaged surrounds in the two cases are identical, but the luminance of the surround with the white ring is higher. Psychophysical experiments have shown that, in general, higher luminance surrounds have greater induction effects than lower luminance surrounds (Jameson, Hurvich & Varner 1979). However, the presence of the spectrally opponent stage before the spatial integration and induction of the spectrally specific contrast stage, results in the opposite effect. The luminance contrast of the black ring causes an enhancement of the response to the red surround in the area surrounding the black ring (see Fig. 15). The chromatic contrast of the white ring, on the other hand, causes a decrease in the response of the spectrally opponent cells to the red surround near the white ring.

The frequency dependent responses of the spectrally opponent cells provide the inputs to the final stage of the network. Therefore, the surround is spatially integrated only after the stimuli have been altered by the spectrally opponent stage. The net effect of the increased contrast response at the opponent stage in both color and luminance is to enhance the contribution of the spectrally-specific contrast stage which uses that contrast information in the push-pull mechanism. These enhancements are most significant when the positive and negative changes in inputs occur in cells which are operating in different parts of their response curves, and which, therefore, have different gains. In these cases, the changes caused by the opponent cells on either side of an edge do not cancel each other when the V4 cells sum these inputs. Because the opponent stage increases contrast under certain conditions, it can sometimes contribute to color and brightness induction via the V4 stage of the network. However, color induction decreases with increasing distance between center and inducing surround (e.g. Tiplitz-Blackwell & Buchsbaum, 1988a). The color opponent stage works in the opposite direction of these observations for high spatial frequency stimuli by decreasing color contrast. This property, combined with the lower sensitivity, compared to other stages, to global luminance changes and to spectrally-specific contrast, preclude consideration of the color opponent stage as a direct contributor to color constancy and color induction.

(iv) Nonlinearities

With limited adaptation, high or low luminance inputs can cause the cells to respond in the nonlinear portion of the response curve, saturating or rectifying respectively the cells’ responses (see Figs 2 and 3). Allowing time for adaptation to take place prevents this to some degree, but if the luminance change is too large even long term adaptation will not compensate enough. The predicted color matches then move toward neutral (Fig. 16). This is analogous to surfaces appearing either “washed out” in very bright light or “muddied” in very dim light. However, this situation makes matching difficult because very different inputs (stimuli) can give very similar outputs (percepts).

The luminance levels of the images used for testing this system and the values for the slope and threshold of each cell type were chosen to prevent the cell responses from being significantly saturated or rectified under most circumstances. The nonlinearities were nonetheless important to the behavior of the network, however, because even at mid-range luminance levels saturated colors can cause individual color channels to respond outside of their linear range. The effect is that the gain of the induction shifts in each color channel will be different for different input images. For example, the gain of the induction in the B cells of the network is much less for a saturated yellow input than for a blue input. The effects of the nonlinearities on the spectrally-opponent stage responses were described in the previous section. However, the qualitative effects of the nonlinearities on the general behavior of the network were usually small.

DISCUSSION

We have described a model of color constancy and color induction which is based on aspects of the anatomy and physiology of the primate visual system. In a number of simulations, the model demonstrates a degree of constancy and induction which is similar to that shown by human psychophysics in previously reported experiments. The color constancy literature has generally been a retinal vs cortical mechanism debate. Although many have argued for a two-stage process (e.g. Walraven, 1976; Werner & Walraven, 1982; Arend & Reeves, 1986; Shevell et al., 1992), the emphasis has often been on determining which stage is most important. The results of this simulation emphasize the importance of both retinal and cortical mechanisms in a cooperative multi-stage system. By systematically adding or eliminating each processing stage during the simulations, we could draw several conclusions about the contributions and interdependence of these stages.

We have shown that a simple push-pull contrast mechanism which uses V4 silent surround cells can, alone, accomplish a significant amount of both color constancy and color induction effects. Given some amount of spatio-temporal integration, adaptation alone can also achieve a degree of color constancy. We found that the effects of receptor adaptation and the V4 mechanism both depend on the particular stimulus conditions and that when one stage cannot achieve color constancy another stage can often compensate. Therefore, a system which combines both adaptation and cortical processing can achieve color constancy in a greater range of stimulus conditions than can either stage alone.

The approach taken here is quite different from many of the previous approaches to color constancy. Recently, much attention has been given to computational theories
for color constancy which estimate the surface reflectance by mathematically separating the illuminant from the reflectance. This is generally done by describing the reflectance and illuminant each as a sum of three basis functions (see review by Lennie & D'Zmura, 1988). The resulting set of equations is underdetermined. In order to solve this set of equations, these models require either restrictions on the reflectances, such as a gray average chromaticity, some a priori knowledge of the illuminant, or assumptions about the mathematical structure of reflectances and illuminants (Buchbaum, 1978, 1980; Brill, 1978; Maloney & Wandell, 1986; D'Zmura & Lennie, 1986; Gershon & Jepson, 1989; Rubin & Richards, 1982; Dannemiller, 1989; Troost & de Weert, 1991a; Brainard & Wandell, 1991). A comprehensive mathematical analysis of the problem, generalizing the earlier approaches and using multiple surfaces and/or illuminants, is given by D'Zmura and Iverson (1993a, b).

There are additional restrictions which allow solutions to the reflectance-illuminant separation problem. One solution is to require the number of photoreceptors to be greater than the dimension of the reflectance space: (Maloney & Wandell, 1986). This solution enables simple reflectance-illuminant separation algorithms. However, this implies that one must severely limit the reflectances that the algorithm can use or that more than three photoreceptor types be involved. Another attempt was made by Faugeras (1979) who developed a filter to separate the reflectance and illuminant by taking the logarithm of the reflectance-illuminant product, thus turning the product into a sum which may be separated by a linear filter. However, the algorithm encounters difficulties when both the illuminant and the reflectance vary so that their spatial Fourier spectra overlap. The current approach does not try to explicitly calculate the reflectance or illuminant spectrum, and so does not require any of these assumptions or restrictions.

Several "lightness" algorithms which have been previously proposed for color constancy, including the Retinex (Land & McCann, 1971), have been shown to be mathematically equivalent to a local spatial derivative plus a normalization term (Hurlbert, 1986). Similarly.
some have argued that the Retinex is essentially the same as Von Kries adaptation in that each is a renormalization of color channel activities relative to some white reference (see review by Jameson & Hurvich, 1989). In this sense, the adaptation stage and the spectrally-specific contrast stage in the simulation are also similar, as are the multiplicative adaptation and the subtractive adaptation mechanisms for color constancy described by previous researchers (Hayhoe et al., 1987; Hayhoe & Wenderoth, 1991). However, there are several important differences in the operations described here which allow them to cover different stimulus conditions and, therefore, to cooperate in their contributions to color constancy.

First, the cone adaptation stage has a permanent white reference which is set by the midpoint of the range of possible threshold values. The adaptation stage also has a long-term adaptation reference which is usually close to neutral because it is established through exposure to many different stimuli over a long period of time. Faster, more localized adaptation effects are deviations from this long term reference. The reference for the spectrally-specific contrast stage is the activity of the "local reference" cells. The spectrally-specific contrast reference is not fixed, but instead changes with each new image. This reference is not usually neutral. The spectrally-specific contrast mechanism described here is also different from most "lightness" algorithms in that the normalizing reference is measured locally, rather than globally. In addition, the spatial profiles of the two constancy mechanisms are different. The effect of localized adaptation depends more heavily on the central test spot, while the effect of the large surrounds in the spectrally-specific contrast operation are more affected by the background stimulus.

The effect of the spectrally-specific contrast operation can be increased by the spectral and spatial opponency of the preceding stage which enhances color and brightness contrast for low spatial frequencies. The simulation results regarding the effects of image spatial structure on color induction are in agreement with the psychophysical results of Wesner and Shevell (1990). The simulation results confirm the assertion of Zaidi et al. (1992) that the psychophysical results could be explained by a mechanism which selectively attenuates high frequency chromatic stimuli before color induction takes place. The spectrally opponent cells reduce the effectiveness of high frequency chromatic inputs in the surround prior to spatial integration and induction by the spectrally-specific mechanism. This agreement of the model with the psychophysical data lends support to the idea that at least part of the color induction mechanism must lie beyond the stage which gives a low-pass response for color stimuli and a band-pass response for luminance stimuli. In other words, there are color induction mechanisms beyond the retina. If there are additional post-retinal contrast enhancing processes, these will also alter the equivalent surround of a complex image if these processes take place before the spatial integration in V4.

Another difference between the retinal and cortical stages which has not yet been incorporated into the current model is the existence of more than three distinct color channels in the cortex. This paper addresses the processing of color information in terms of color constancy and color induction, but does not address the more complex problems of image representation. Spectral sensitivities of cortical cells have been shown to have peaks at many different wavelengths (Zeki, 1980; Lennie, Krauskopf & Sclar, 1990; Schein & Desimone, 1990) indicating a more distributed representation of color information in the cortex. There are also questions remaining about how the processed color information is then integrated with information about image segmentation and object perception. The output of the simulation, which represents color information at a single point in the image is, most likely, highly simplistic.

Our goal was to examine the effects and interactions of color processing mechanisms rather than specific cellular mechanisms. We cannot rule out other possible implementations of the processing stages used in this network because many anatomical substrates can accomplish very similar processing tasks. We intentionally abstracted some of the anatomical details so that the emphasis would be on the information processing mechanisms themselves. The model is robust enough that the primary results do not depend on any particular parameter value or anatomical implementation. Eventually, we would like to make the simulation and our predictions for psychophysical and physiological experiments more quantitative. This will require more anatomical detail and more indepth parameter optimization. There are many parameters in the simulation which are not directly determinable from current physiological data. As a first step, however, we wished to address more general questions regarding color information processing in the visual system, independent of the specific anatomical implementation.

The implementation of the adaptation stage in the simulation was particularly difficult because it is a dynamic process in an otherwise static model. In order to calculate what the adaptation state should be, assumptions had to be made regarding the previous stimuli presented to each cone during the adaptation period. This depends on the conditions of the experiment being simulated. The results will be different for different types of viewing conditions (e.g. haploscopic, simultaneous match and test stimuli, or memory matches). We assumed in these simulations that prior to each stimulus presentation, there was long-term adaptation to a moderate luminance neutral uniform field. We calculated the adaptation shift (away from the neutral adaptation state) each time a new stimulus was presented, whether that stimulus was the test stimulus or the match. A gaussian weighting function was used because each image had a central region of interest and was either symmetrical about that central region, or had a Mondrian background which had a random distribution of color patches. As has been discussed in numerous psychophysical studies, the experimental conditions can
greatly affect the adaptation state of the visual system. The same is true with our simulation.

One possible extension of this model involves solving the problem of image scale invariance which, in the case of human color perception, means that the color of an object does not change significantly with size, provided that the regions surrounding that object are scaled in the same proportion. In other words, this is the common observation that objects don’t change color as we walk toward them. One possible solution is dynamic receptive fields which adapt to match the spatial scale of the stimulus. Pettet and Gilbert (1992) have recently found physiological evidence for stimulus-dependent dynamic receptive fields in cortical area V1. In addition, Moran and Desimone (1985) reported cells in V4 and inferior temporal cortex whose responses depended on the state of attention of the animal. While evidence for very rapid stimulus dependent receptive field changes is still preliminary, if such mechanisms do exist then these dynamic properties could be incorporated into the V4 mechanism described here to allow for image scale invariance.

There are also additions that could be made in order to include other aspects of color perception. For example, it is known from psychophysics that there are contributions to brightness induction from binocular depth information (Schirillo & Shevell, 1993) and surface segmentation (White, 1979). There are also task dependent surface/illuminant segregation influences (Arend & Reeves, 1986; Troost & de Weert, 1991b; Craven & Foster, 1992). There are many interacting processes involved in color perception and no single mechanism can be credited with achieving “color constancy”. In addition to color constancy and color induction, the stages of the network described here are rather basic processes and each stage is likely to serve many other roles in the visual system as well. The present study provides a different perspective in the debate as to whether retinal or cortical mechanisms have a greater contribution to color constancy and color induction. Although others have suggested the need for both retinal and cortical visual color processing, this paper emphasized the distinct roles of each stage and the interactions between the stages. The two levels of processing have important but different effects on color constancy and color induction, not necessarily greater or smaller effects.

REFERENCES


Acknowledgements—The authors would like to thank Paul Sajda for extensive help in implementing the network using the NEXUS neural simulator. This work was supported by grants from the Office of Naval Research, N00014-90-J-1864 and N00014-93-J-0681; Air Force Office of Scientific Research, 91-0082 and 92-J-0316; The Whitaker Foundation, and the McDonnell-Pew Program in Cognitive Neuroscience.
A biologically-based multi-stage neural network is presented which produces color constant responses to a variety of color stimuli. The network takes advantage of several mechanisms in the human visual system, including retinal adaptation, spectral opponency, and spectrally-specific long-range inhibition. This last stage is a novel mechanism based on cells which have been described in cortical area V4. All stages include non-linear response functions. The model emulates human performance in several psychophysical paradigms designed to test color constancy and color induction. We measured the amount of constancy achieved with both natural and artificial simulated illuminants, using homogeneous grey backgrounds and more complex backgrounds, such as Mondrians. On average, the model performs as well or better than the average human color constancy performance under similar conditions. The network simulation also displays color induction and assimilation behavior consistent with human perceptual data.

INTRODUCTION

Color perception can contribute to object recognition if the perceived color is a fixed attribute of the object surface. However, the color signal which reaches the eye, or any other detector, is the product of the surface reflectance and the incident illuminant. Therefore, this signal changes as the illuminant changes, from noon sunlight to hazy sunset, or from incandescent to fluorescent artificial lighting. The term "color constancy" has been used to describe the ability of humans and other animals [25, 30] to discount a portion of the illuminant in order to make more reliable judgments of surface color. Color induction, a related phenomenon, is the change in the perceived color of a surface due to its juxtaposition with other colored surfaces. Color induction enhances the color contrast in a scene and probably aids in object detection and surface segmentation. Color constancy and color induction demonstrate that human color perception depends on the spatial distribution of the wavelengths of light present in the entire image. Cameras and most artificial vision systems, on the other hand, are not color constant. Photographs taken in fluorescent light often look green.

We would like both to understand how the human visual system accomplishes the task of discounting the illuminant, and also to build a practical artificial vision system which has color constancy. These are not necessarily distinct tasks. A logical way to design an artificial color constant system is by reverse engineering of the existing biological design, the human visual system. An ideal color constant system should show the following characteristics:
1) Accurate object reflectance color determination
2) Flexibility, i.e. good color output for a large variety of images.
3) Few assumptions about the image and little a priori knowledge.

The human visual system, as will be explained in what follows, seems to favor flexibility and fewer assumptions over complete accuracy. The color constancy achieved by the biological system is thus only approximate, but it is able to handle a large variety of scenes under very different lighting conditions. Depending upon the application, one may want to design a system which has perfect color constancy, one which more closely mimics the human visual system, or one which is a compromise between the advantages and disadvantages of each. An example of an application which would require more of an emphasis on accurate color discrimination would be quality control of textiles in multiple factories which have different lighting conditions. An application which would emphasize the duplication of human color perception would be catalogue printing in which the color of a picture of an item as perceived by a human looking at the catalogue should be the same as the color perceived by a human looking at the actual object being sold.

Most previous computational algorithms for color constancy [3, 8, 13, 14, 15, 17, 37, 42] have favored accuracy over flexibility. Most of these try to explicitly recover the reflectance spectrum from the reflectance illuminant product. Because the resulting set of equations is underdetermined, assumptions regarding either the reflectances or the illuminant must be made in order to find a solution. When these assumptions are not met, the algorithms can make color predictions far from those reported by humans.

One algorithm for achieving color constancy which has received much attention is the retinex theory developed by Land [31-34]. The retinex calculates the relative "lightness" of each area of a scene within three separate channels, each sensitive to a different region of the visual spectrum. A variety of algorithms have been suggested for computing the lightness values. (For example [31-34], [38], [23]). The most recent method of calculation [33] bears a resemblance to the receptive fields of cells recorded by Schein and Desimone [44] in cortical area V4 in that it effectively subtracts a large "surround" region from a very small "center" with similar wavelength sensitivity.

The retinex theory successfully predicts many of the basic properties of human color perception, but there are limitations in the algorithm in certain situations as has been demonstrated by Brainard and Wandell [2] and in a neural network by Moore, Allman, and Goodman [39]. Because the retinex records only contrast, the interior of large uniformly colored areas become grey. In addition, in order for the contrast information to be converted to color and luminance information, one of two assumptions must be made. Either the average chromaticity of the reflectances in a scene must be constant for all
images (known as the grey world assumption) or the brightest region in an image must be a white reflectance. In addition, the retinex must assume that the illumination varies slowly in space while reflectances have sharp chromatic borders. Some of these assumptions were also used by many of the separation of reflectance and illuminance theories. There are many examples of images that do not fit these conditions (such as scenes with shadows or simple images which contain a strongly colored object). In such cases the retinex will predict colors far from those perceived by humans.

Neural networks have been used for implementing the Retinex and a variety of other color constancy algorithms. As mentioned above, Moore et al. [39] implemented Land's retinex algorithm within a neural network structure. They also incorporated a modification in order to eliminate the color washout problem described above by multiplying the surround factor by the "edginess" of the area. Hurlbert and Poggio [24] demonstrated that, using a number of different "learning" methods in a neural network, including least squares gradient descent and back propagation, a linear operator could be found which is similar to that proposed in the Retinex algorithm.

Grossberg [18] also developed a neural network simulation which obtains color information from the contrast at boundaries and then fills in the color into each segment of the image. It successfully demonstrates many aspects of brightness and color perception. However, because only the contrast information is preserved, the system must rely on the "grey world" assumption to determine the mean level to which the contrast refers. Therefore, as with all theories which use this assumption, the predictions of the network can be inconsistent with psychophysical results in scenes that have a non-neutral average chromaticity. Also, because it fills in each segment with the average color determined from the surrounding boundaries, it loses information about subtle variations in the reflectance within a segment.

The network model developed by Dufort and Lumsden [12] uses double opponent cells and features output cells which behave qualitatively like the "color constant" cells reported by Zeki [58] in V4. The parameters of the network were optimized to create output response curves which correspond to color naming categories developed by psychophysical studies. The network design directly incorporates several aspects of color psychophysics. At present, however, the network addresses only hue constancy. The saturation and lightness dimensions of color signals are also affected by changes in the illuminant, and require a constancy mechanism as well.

This paper describes a system which is based on the primate visual system, from the retinal to cortical area V4. Each stage contributes both to producing better color constancy and to the flexibility of the system. No assumptions are made about the
mathematical structure of the reflectances and illuminants, the average chromaticity of the reflectances in the image, or about the existence of chromatic or luminance gradients in either the reflectances or illuminants.

NETWORK ARCHITECTURE

An overview of the processing mechanisms in the network is shown in figure 1. The network was simulated using NEXUS, an interactive neural simulator designed for large scale models [43]. The complete network consists of over 11,000 cells and approximately 1.65 million connections. Below we will describe the properties of each stage of the network. Table 1 summarizes the most significant parameters in the model.

i. Input

The first stage corresponds to the cone responses. The input image is a 27x27 array, in which each entry defines the color at that location, specified in either Munsell color notation or in CIE (Commission Internationale de l'Eclairage) notation (x, y, Y). The array is converted to three 27x27 arrays of cone activation levels: R, G, B. Therefore, an input image unit has a one-to-one correspondence with a set of three units (analogous to one cone of each type) in the first layer of the network. For example, an image defined using Munsell spectra would be converted, at each point, to the three normalized cone activation levels by using the Vos-Walraven [51, 52] cone action spectra (\(r(\lambda), g(\lambda), b(\lambda)\)) which are shown in figure 2.

\[
\begin{align*}
R &= \sum_{\lambda=400}^{700} k_1 r(\lambda) R(\lambda) I(\lambda) \Delta\lambda \\
G &= \sum_{\lambda=400}^{700} k_2 g(\lambda) R(\lambda) I(\lambda) \Delta\lambda \\
B &= \sum_{\lambda=400}^{700} k_3 b(\lambda) R(\lambda) I(\lambda) \Delta\lambda
\end{align*}
\] (1)

where \(R(\lambda)\) is the reflectance spectrum and \(I(\lambda)\) is the illuminant. The coefficients \(k_{1,2,3}\) are constants which normalize the sensitivity spectra so that all cone types in the simulated array have the same peak sensitivity. Therefore, the three types of first layer units ("cones") have responses of the same order of magnitude, and we may use the same dynamic range for all chromatic cell types in subsequent stages. For those cases in which the image was specified in CIE notation, the image was converted to cone activation levels by applying the transformations for Vos-Walraven action spectra [51, 52, 56] and then normalized using the same coefficients, \(k_{1,2,3}\):
\[ R = k_1 Y \left[ 0.155 \left( \frac{x'}{y'} \right) + 0.543 - 0.037 \left( \frac{1 - x' - y'}{y'} \right) \right] \]

\[ G = k_2 Y \left[ -0.155 \left( \frac{x'}{y'} \right) + 0.457 - 0.030 \left( \frac{1 - x' - y'}{y'} \right) \right] \]

\[ B = k_3 Y \left[ 0.007 \left( \frac{1 - x' - y'}{y'} \right) \right] \]  

(2)

where \( x' \) and \( y' \) are the Judd modified [24] 1931 CIE chromaticity coordinates. \( Y \) is the luminance in cd/m\(^2\) of the stimulus and is used here to scale the cone responses for luminance.

**ii. Cell Responses and Nonlinearities**

In the simulation of the network model, the input layers of the network correspond to the three cone types of the human retina. These cells have a Naka-Rushton type response function [41]:

\[ A_i = \frac{Q_i^x}{Q_i^x + \sigma_i^x} \]  

(3)

where \( Q_i \) is the total input to cell \( i \), \( \sigma_i \) is the threshold of cell \( i \), and \( x \) is a constant from 0.7 to 1.0. In the simulation results shown here \( x = 0.9 \), but the general behavior of the system was not very sensitive to the value of this parameter. In all other stages, cell activity is determined by a standard sigmoidal response function of the input:

\[ A_i = (\max - \min) \left( \frac{1}{1 + \exp(-Q_i - \sigma_i \beta_i)} \right) + \min. \]

(4)

**iii. Adaptation**

(a) **Motivation**

The nonlinear response function of both the cones and the cells in higher layers of the network, gives the system a limited dynamic range. However, the light level varies in our daily environment from \( 10^{-4} \) to \( 10^{+5} \) cd/m\(^2\). In order to keep the cell responses in all stages within the linear range of their response functions, the cones must be able to shift their thresholds to accommodate the overall level of incoming light. Although there may also be mechanisms for adaptation in later stages of the visual system, individual primate cones are known to change their sensitivities according to the amount of light available [46].

Adaptation was recognized early on as a probable contributor to color constancy in humans. (e.g., [21], [50]) Although adaptation alone, defined as a multiplicative gain
change of individual photoreceptors, has been mathematically proven to be incapable of achieving perfect color constancy \([6, 11]\), it can produce some degree of color constancy if the mechanism includes integration across space and time. In human psychophysical experiments, color constancy has been shown to depend on length of presentation time \([4, 19, 20]\) and eye movements \([9]\). This implicates receptor adaptation because it depends on the temporal integration of activity in spatially localized mechanisms. Therefore, the inclusion of adaptation in our simulation allows us to better mimic human perception by incorporating the differences in color perception under different viewing conditions.

\((b)\) Implementation

In the simulation, we assume an initial long-term adaptation to a uniform neutral background. (see \([53]\) for a review of psychophysical and physiological studies on adaptation) The amount of threshold shift is determined by the difference between the cone activation level for the neutral background stimulus and the cone activation level for the new stimulus. Because adaptation is dependent on the temporally weighted average of its input, the adaptation shift for a cone is dependent, not only on the point in the image directly corresponding to that cone position, but also on the surrounding area to which the cone may be exposed during eye movements, or from optical blur.

We approximated this temporal averaging effect by a two-dimensional Gaussian spatial weighting function, because for most viewing conditions which have been tested psychophysically, there is either a fixation point, or a central test patch around which one can assume eye movements were centered. In the simulation, the amount of the shift follows a sigmoidal function of the difference between the neutral and the current stimuli and is proportional to the length of viewing time. These constraints are incorporated into the simulation by including one adaptation cell for each receptor. The adaptation cell shifts the threshold of its corresponding receptor according to the following equation:

\[
\begin{align*}
\sigma_{\text{new}_i} - \sigma_{\text{neut}} &= \alpha \left\{ 2G \left( \frac{1}{1 + e^{-(Q_i - Q_{\text{neut}})\beta_i}} \right) - G \right\} \\
G &= \sum_{i=0}^{n} (Q_i - Q_{\text{neut}}) \left( \frac{1}{2\pi\theta^2} \right) e^{-\left( \frac{-(x^2+y^2)}{2\theta^2} \right)}
\end{align*}
\]  

\[ (5) \]
where

\[ \sigma = \text{threshold} \]
\[ \beta = \text{proportional to the slope of the linear portion of the function.} \]
\[ Q_i = \text{cone activation level (i.e. R, G, or B) due to current image pixel } i \]
\[ Q_{\text{neut}} = \text{cone activation level due to standard neutral input at image pixel } i \]
\[ n = \text{total number of matrix entries comprising the image}, \]
\[ x, y = \text{the horizontal and vertical distances from entry } i \text{ to the center of} \]
\[ \text{the cone's receptive field when fixated on the center of the image} \]
\[ \theta = \text{the width of a Gaussian weighting function which varies with the} \]
\[ \text{degree of fixation required for the experiment} \]
\[ \alpha = \text{the fraction achieved within the stimulus presentation time of the} \]
\[ \text{total difference in long-term adaptation states between the neutral} \]
\[ \text{state and the state for the new stimulus.} \]

\( \alpha \) is proportional to the time of exposure. As \( \alpha \) increases, the size of the threshold shift increases, following a sigmoidal curve ranging from \(-G\) to \(+G\), where \( G \) is the difference between the weighted average activation level for the current image and the activation level for a uniform neutral background. (See figure 3) In the current study, \( \alpha \) was held fixed at 0.3 and \( \theta \) was held fixed at 3.0. However, we wished to include this flexibility in the model so that in future studies of the current network, we could make closer comparisons with human psychophysical data. With longer exposure time, the adapting cell will be able to better adapt (larger \( \alpha \)) to its new stimulus. Under certain experimental conditions, longer exposure time may also allow for more eye movements. The spatial extent of the weighting function broadens (larger \( \theta \)) with more eye movements. In the extreme case of very long exposure time and completely random eye movements over the entire field of view, the weighting function would be flat and the cone would adapt to the field average. This dependence of the parameters \( \theta \) and \( \alpha \) on eye movements and viewing time allows the effects of the adaptation stage of the simulation to vary with the type of viewing conditions being considered. (See Table 1 for additional information on parameter values.)

iv. Spectral opponency

Spectrally opponent cells are excited by one region of the spectrum and inhibited by a different region. In the network spectrally opponent cells are obtained by subtracting responses of spectrally opponent cone types and are generally based on the properties of LGN parvocellular type I receptive fields, which are spatially opponent as well. [22] This means that the excitatory and inhibitory regions are spatially segregated into center and surround regions. The spectral opponency gives the cell a high gain for low spatial frequency color changes, i.e. a color change which is spatially homogeneous across both
center and surround regions. The reason for this high gain is that such a color change will result in the combination of either an increase in excitation and a decrease in inhibition or a decrease in excitation and a increase in inhibition. For example, for a cell which is excited by red light in the center of its receptive field and inhibited by green light in its surround, a color change of the stimulus from green to red will result in increased excitatory input to the center and decreased inhibitory input to the surround.

High spatial frequency stimuli, such as edges, for which the center and surround regions receive different inputs, result in a different type of response. In the case of high spatial frequency stimuli which differ only in luminance, spatial opponency results in enhanced cell responses, because the excitatory regions may receive input from the higher luminance portion of the stimulus while the inhibitory regions receive input from the lower luminance portion of the stimulus, or vice versa. High spatial frequency stimuli which differ only in color, however, may result in diminished responses. For example, a red-green edge stimulating a red excitatory center, green inhibitory surround cell, will result in a smaller response than would a homogeneous red stimulus, because the green portion of the stimulus will inhibit the surround region of the receptive field, even though the center may be receiving excitation from the red side of the stimulus edge. This will result in the blurring of purely chromatic edges (i.e. edges with no luminance difference).

A layer of spectrally and spatially opponent cells contributes to the color constancy abilities of the current system in two ways. First, the increased chromatic gain enhances the overall sensitivity of the final stage of the network. Second, the additional enhancement of luminance edge responses acts in a way analogous to the "edginess" factor used by Moore et al. [39] to eliminate the washout of the color response to large homogeneously colored regions seen in the Retinex algorithm [33]. In their improved color constancy network, the main calculation was

\[ \text{output} = \text{center} - (\text{surround} \times \text{edginess}) \]

where the edginess was determined by the average of the absolute values of the local spatial derivatives. The spatially opponent second stage of the current network accomplishes something similar in that the responses of this layer are enhanced in regions with luminance edges. One difference between the current operation and the Moore et al edginess factor is that the Moore et al operation enhances both color and luminance edges. The spectrally and spatially opponent cells enhance luminance edge responses, but diminish responses at equiluminant color edges. In real images, however, equiluminant edges are rare. These are mostly laboratory color stimuli designed to control effects of luminance. The enhanced response provides the input to the final stage of the network, described in the next section, which is also a center-surround operation. Therefore, the
final stage will produce larger color shifts in "edgy" regions and smaller shifts in large homogeneously colored regions.

To produce these cells in the simulation, each "cell" receives excitatory input from a single cone in the center of its receptive field and inhibitory input from both R and G cone types surrounding the center [35]. No inhibitory surround in the network receives input from B cones because physiological recordings show no significant B cone input in LGN cell inhibitory surrounds [35, 60]. The surround input is most heavily weighted toward the cone type(s) opponent to the center cone type. The number of R and G cells providing input to the surrounds are equal for all cell types. However, for R center cells, the amplitude of the synaptic weighting function for the G cone input to the surround is twice that of the R cone input to the same cell. The center strength (volume of 2D Gaussian sensitivity profile) is twice that of the surround, allowing these cells to have a significant response to homogeneous fields as well as to edges.

v. Higher Cortical Processing
(a) Motivation

The next stage in the network is similar to the Retinex and related color constancy algorithms in that it measures and uses spectrally-specific contrast [36, 48], what Land called "lightness" [31]. However, the method of normalizing contrast responses relative to a reference level is handled differently in the current system. In this simulation, the final stage is designed to respond according to the primary chromatic properties of cells in V4 [44]. Most cells in V4 have large, suppressive surrounds which have approximately the same wavelength sensitivity as the center of the receptive field. (See fig. 1) These large surrounds are called "silent surrounds" because they had little or no effect on the cell's activity unless the center was also stimulated. Desimone and Schein and their colleagues [10, 40] reported that the effect of stimulation in the silent surround decreases with increasing distance from the classical receptive field. Psychophysical results also show a decrease in the effect of inducing regions with distance. (e.g., [47], [49], [54], [57])

The strengths of the centers and silent surrounds of V4 cells appear to be well balanced; stimulation of the surround can completely inhibit the response to stimulation of the center [44]. Because the cells in V4 with silent surrounds respond only when there is a difference, either in wavelength or luminance, between the center and the distant surround, these cells are particularly well suited for carrying information about spectrally specific contrast. The significance of spectrally specific contrast in the visual system has been demonstrated psychophysically [36, 38, 48]. However, for those images that have little spectrally specific contrast, or an unknown or non-grey average chromaticity (e.g. blue
sky, green forest), the DC (or spatial average) information is also important. It is significant, therefore, that approximately 10% of the cells found in V4 did not have silent surrounds. The cells without silent surrounds have the same classical receptive field response as those cells with silent surrounds. These cells have the capacity to carry the (spatial) DC portion of the signal, i.e. to respond to homogeneous fields as well as edges and small spots. These center-only cells have been included in the network and we refer to them as "local reference cells" because in the network they provide the normalizing reference information for the contrast cell responses.

(b) Implementation

To incorporate these observations into the simulation, the responses of analogous V4 stage "cells" in the simulation were created directly using the outputs of the spectrally opponent stage. A positive contrast cell receives its input, excitatory from the center and inhibitory from the surround, from a single type of on-center spectrally opponent cell. Therefore, the positive contrast cells respond to images for which the input to its classical receptive field is greater than the input to its silent surround. We have also included negative contrast cells which receive input from off-center cells and, therefore, respond when the center input is less than the surround input. The inputs to the surround are weighted according to distance from the center by a negative exponential function. (See figure 1.) The "silent" nature of the V4 cell surrounds was implemented in the simulation by rectified inhibition, which was achieved by giving the V4 cells in the simulation, like those recorded physiologically, very low levels of spontaneous activity. Therefore, inhibitory input from the surround was only effective when there was also excitatory input from the center, "classical" receptive field.

In order to combine the physiological information from the local reference and contrast cells into a simple set of outputs which could be compared to human color perception, we combined the outputs of these V4-like cells into a simple push-pull mechanism. (This stage is shown in figure 1b.) This is the output of the final network stage and it is determined by the response of the local reference cells, enhanced by the positive contrast cells, or inhibited by the negative contrast cells, as given by the equation

\[ O_x = B_x + c_1 P_x - c_2 N_x \]

where \( O \) is the output, \( B \) is the local reference response, \( P \) is the positive contrast response, \( N \) is the negative contrast response, and \( c_1 \) and \( c_2 \) are constants. The constants \( c_1 \) and \( c_2 \) are chosen, together with \( \alpha \) (in equation 5), to increase or decrease the size of the constancy shift. In the simulation, we used \( \alpha = 0.3 \) and \( c_1 = c_2 = 0.2 \). This choice
enabled a minimum of 20% constancy (measured as distance in RGB space, see below) for all stimuli. In addition, these parameter values resulted in network behavior which was similar to the corresponding psychophysical data [48].

RESULTS

i. Measuring Constancy

An output to a particular reflectance in an image is considered as "achieving some degree of color constancy" if the difference (in color space) between the network's predicted color of the reflectance (the output) and the "true color" (the color under neutral illumination with a neutral background) is less than the difference between the "true color" and the "physical color" (the color as calculated from the power spectrum of the reflectance times the illuminant). "Physical color" is what would be expected in a completely non-color constant system, such as a camera or photometer. A "shift toward constancy" is a shift of the output toward the true color and away from the physical color.

When evaluating the behavior of the network the "true color" response is defined as the output of the network for a reflectance under standard neutral illuminant and background conditions. We call the "true color" of a reflectance an "eigencolor" of the system because it is an internal reference which is not altered or "corrected" by the system in the way that those inputs which have non-neutral backgrounds or illuminants are. The "eigencolor" response represents "perfect constancy". A reference was also needed for a non-constant response, which would correspond to the "physical color" reference, reflectance times illuminant.

In human perception, stimuli which are viewed "in isolation", either with a completely black background or viewed at a distance through a hole or aperture in a grey barrier, result in a close correlation between the physical color and the perceived color. In other words, in "aperture viewing mode" color constancy disappears. ([29], also called "void viewing mode" in [59], similar to "film colors" reviewed in [1]) Without a surround which also reflects the illuminant change, the visual system cannot determine whether a change in the color signal from the test spot is due to an illuminant change or a reflectance change.

Stimuli analogous to the aperture viewing condition were created for the network by simulating a change in the illuminant on only the center test spot, leaving the background as a neutral reflectance under a neutral illuminant. (See figure 4) As with a human viewing a color in aperture mode, the outputs to these stimuli could be considered essentially non-color-constant, because there is no change in the surround color signal to indicate that the change is due to a different illuminant and not to a change in reflectance.
The "aperture colors" will not be perfectly non-constant because of a small amount of compression from the adaptation of those receptors directly on the test spot. Therefore the method underestimates slightly the contribution of adaptation in the simulation, but the effect is minor compared to the size of the color constancy shifts under natural viewing conditions. Simulation stimuli for "natural viewing conditions", in which the illuminant was applied to the entire image, are the test stimuli.

The Euclidean distance in output space \( \Delta(O_R^2 + O_G^2 + O_B^2)^{1/2} \) was calculated between each aperture color response and the corresponding eigencolor response. This distance was then compared to the distance between the test stimulus output and the eigencolor output. The percent constancy achieved by the network is calculated by:

\[
\frac{D_{ea} - D_{et}}{D_{ea}} \times 100
\]

where \( D_{ea} \) is the distance from eigencolor to aperture color response, and \( D_{et} \) is the distance from eigencolor to test stimulus response. In this way, a quantitative measure was obtained for the network's ability to discern the "true" color of the center reflectance by using information about the illuminant from the surround. We will use this measure to assess and compare the effectiveness of the network in different simulated stimulus and background conditions. Obviously, the end points with 0% and 100% correspond exactly to their psychophysical counterparts. Correspondence of computed values between 0% and 100% cannot be taken as numerically equivalent to psychophysics. Psychophysical color space is not Euclidean nor linear in RGB along the curve from 0% to 100% and the computed distance measures may not exactly match.

### ii. Color Constancy with a Homogeneous Neutral Background

#### (a) Natural Illuminants

Because of the importance for practical applications, and for evolutionary significance, we begin by looking at the color constancy abilities of the network under natural illuminants, such as the various phases of daylight. The power spectra used are from tabulations of the CIE standard illuminants, A, B, C, D55, D65, and D75 [56]. We will refer to them here by qualitative descriptions of their sources: incandescent light (or full radiator), direct sunlight, average daylight (averaged over all times of day from dawn to dusk), overcast daylight, natural daylight (a single phase of daylight), blue skylight. These spectra are shown in figure 5. The natural illuminants vary primarily in the short (blue) and long (red) wavelength regions with very little change in the middle (green) wavelengths.
The network outputs for three Munsell reflectances (2.5yr7/10, 10b6/10, and 2.5bg6/10,) under these illuminants are shown in figure 6. The stimuli consisted of a single small square reflectance patch (the "test spot") in the center of an homogeneous grey background reflectance. The figure shows the ratio of the network outputs in a ternary plot. The plots shown are sub-spaces of the full graph in which the ratios between each of the outputs and the sum of all three outputs range from 0 to 1. Each of the three corners of the full triangular graph represents activity in a single channel only, as would result from a monochromatic light stimulating one color channel exclusively. Because we are using broadband reflectances and illuminants and because the cone spectral sensitivity functions overlap significantly, the network outputs are concentrated in the center of the graph. In particular, the $OR/(OR+OG+OB)$ and $OG/(OR+OG+OB)$ ratios are usually close to 0.5 because of the large amount of overlap in the R and G cone sensitivities.

The squares mark the eigencolor outputs. The outputs to the aperture colors are shown with solid circles. As explained in the previous section, these are almost completely non-constant, and represent the physical change in the color signal. The open circles represent the test stimuli outputs, the responses to each of the reflectances under each of the illuminants under natural viewing conditions. The network responses to the test stimuli are significantly closer to the eigencolor outputs than the aperture color outputs are to the eigencolor outputs, demonstrating color constancy.

(b) Larger Illuminant Changes

Because natural illuminants are relatively similar to each other, the constancy shifts required in the previous sections were rather small. To see how the network would handle larger shifts, more strongly colored illuminants were tried. These illuminants were a linear combination of a spectrally flat illuminant and one of three illuminants with Gaussian spectral distributions which peaked at 440nm, 560nm, and 660nm. Some examples are shown in figure 7. Solid black circles represent aperture color outputs, and open circles represent test stimuli outputs. Eigencolor outputs are shown by squares.

To see the difference in the system's color constancy abilities in natural versus artificial illuminants, compare figure 6 to figure 7. Although the percent constancy achieved with natural and artificial illuminants is similar (average shift for all stimuli was approximately 50%), the total distances, caused by the physical color change, for which compensation is needed, are much smaller for the natural illuminants than for the artificial illuminants. Therefore, the variance in the test stimuli outputs is much smaller for the natural illuminant conditions. Because the biological system evolved to deal with natural
illuminants, it should not be surprising that the amount of constancy achieved in laboratory conditions with artificial illuminants and viewing conditions is relatively small.

iii. Color Constancy with Complex Backgrounds

The network implementation of this color constancy algorithm allows complex backgrounds to be handled as easily as simple backgrounds. This is not to say that the response to a complex background is the same as the response to a spatially weighted average of the background (the equivalent surround hypothesis: [49]). As has been demonstrated in psychophysical experiments [7, 27], local features in the background region of the image, such as low contrast edges, can affect the constancy and induction results for the test spot. In the simulation, color and luminance edges in the background influence responses of the local operations in the retinal layers of the network. For example, the spectrally opponent cells enhance luminance edge responses. These enhanced responses are then spatially integrated and used by the spectrally specific contrast operation in the final stage of the network.

The network was tested using Mondrian images, in addition to the grey homogeneous background condition described in the previous section. Some examples are shown in figure 8. Again, solid circles represent aperture color outputs and open circles represent the test stimuli outputs. Squares represent the outputs for those reflectances with the Mondrian background and a neutral illuminant. Therefore, the squares represent the color constant output, but do not eliminate color induction effects from the Mondrian background reflectances.

The triangles mark the eigencolor outputs for these two reflectances. These outputs represent each of the reflectances under neutral illumination with a homogeneous neutral background. The shifts from the neutral background, eigencolor outputs to the square symbols, show the color induction due to the Mondrian background. The colors of the patches in the Mondrian used for the results shown here were chosen at random and had a yellow average chromaticity. Therefore, independent of the influence of the illuminants, there is a shift in the outputs away from red and green and toward blue. The amount of induction depends on the amount of color contrast between the center test spot and the background. Color induction will be discussed in more depth in section v.

iv. Gradients in Illumination

Most previous computational algorithms for color constancy required that the illuminant be constant across space (e.g., [5], [8], [14]) or at least that the illuminant vary slowly relative to the reflectance changes [15, 31, 42]. This assumption was necessary in
order to make the computations tractable and to distinguish changes in illuminant from changes in reflectance. However, nearly all natural scenes have gradients in the illumination from multiple light sources, varying distance to the light source, and shadows. The neural network's distributed representation of the input image does not have this requirement. We tested the network using simulated images (test spot plus uniform grey background) with chromatic gradients in the illuminant.

There is little psychophysical data on the effects of gradients, either in luminance or color, on the perception of color. Therefore, we tested our network on simple stimuli which would have an easily predictable perceptual effect. We chose to use circularly symmetric images with symmetric illuminant gradients. The linear gradients were either horizontal or vertical across the image and were pivoted about the center of the image so that the gradient was positive on one side of the central test spot and negative on the other side. Such a stimulus would be expected to have no net effect on the perception of the color at the center of the image.

The spatially homogeneous illuminants used above are simulations of illuminants passed through spatially homogeneous filters. The illuminant gradients were effectively a simulation of these same illuminants, but passed through one or more filters whose density increases linearly across the image. The gradient filter has the same average density as the spatially homogeneous filter used above. The activation levels of cones of one type were multiplied by \((1 + (a \cdot d))\), where \(a\) is the "amplitude" of the gradient and \(d\) is the distance (either positive or negative) from the center of the image along either the horizontal or vertical axis. The distance was measured as the distance from the center of the image to the center of the cone's receptive field as a fraction of the distance from the center of the image to the edge of the image. A constant value was added or subtracted uniformly across the entire image for some of the stimuli to accommodate a DC offset.

The results are shown in figure 9. Each group of data points represents a different offset value. The horizontal axis shows the color channel that contained the gradient, and the amplitude of the gradient multiplication factor at its highest point (the outer edge of the image). As the figure shows, the offsets produced a shift in the R/G ratio of the outputs, but the gradients showed no effect at all on the output of the network, either for chromatic or luminance gradients. This result would be expected intuitively for human perception as well because the background was uniform and the gradients were symmetric. The increase in activity on one side of the image is canceled by the decrease in activity on the other side of the image when the large surrounds of the V4 network stage integrate their inputs. Had this image contained a more complex background, which was not spatially symmetric in the chromaticities of the reflectances, then a chromatic or
luminance bias could have been introduced by the gradient. The important point here is that the network enables the computation of color in the presence of gradients.

v. Color Induction

In their implementation of the retinex algorithm, Moore et al. [39] were concerned about the existence of color induction in their output images. Indeed, if the induced shift in color output is too great, or in the wrong color direction, then color induction is undesirable. However, if we wish to imitate human color vision, our system must have color inducing behavior. In addition, color induction is helpful if we wish to take advantage of the increase in contrast that color induction creates in order to do object detection or surface segmentation.

The current network simulation produces color induction through the same mechanisms which produce color constancy. Color induction results are shown in figure 10. The size of the induction shift increases as the size of the surround increases and decreases as the width of the gap between the inducing surround and the center test spot increases. The size of the induction shift also depends on the color difference between center and surround as can be seen in the different induction amplitudes for the bluegreen center, blue surround stimulus and the purple center, red surround stimulus. This difference in amplitude is because of the dependence of the final stage of the network on the spectrally specific contrast between center and surround. If the contrast between center and surround is small, then the push-pull spectrally-specific contrast mechanism will not be activated. On the other hand, if the difference between center and surround is large, then one or more of the color channels for either the center, the surround, or both, is likely to be responding in the nonlinear region of its response curve where the gain is small, making the induction ineffective.

vi. Color Assimilation

Another practical consideration for color applications is color assimilation, which is the opposite of color induction. Assimilation results in the blending of colors, a decrease rather than an increase in color contrast. Assimilation is usually seen when a fine pattern is placed on a colored background. For example, thin white stripes over a colored background will make the background appear lighter. Thick white stripes, however, will cause induction and make the background appear darker.

Color assimilation occurs in the simulation as a result of the spectrally and spatially opponent cells in the second layer of the network. As explained in section iv of "Network Architecture", the opponent stage enhances luminance edges, but diminishes
chromatic edges. This is caused by the difference in spectral sensitivities between the center and surround of these cells' receptive fields. All three cone types have responses which are highly correlated to a change in luminance, so the spectrally opponent sensitivity of the cell is not a factor at a luminance edge. A high spatial frequency change in color, however, often causes decorrelated responses between cone types, leading to the decrease in cell responses near a chromatic border when the excitatory center receives different input than the inhibitory surround. For example an R+G- cell whose center is stimulated by a red region in an image will have a smaller response if its receptive field is close enough to the edge of a green region so that the green inhibitory surround is stimulated by the green region in the input image.

The response of the R on-center opponent layer to an assimilation stimulus is shown in figure 11. In this example, a red spot and an open red square are shown with a yellow background. When the spot is large and the lines of the square are wide so that both the excitatory center and the inhibitory surround portions of the opponent cell receptive fields are stimulated by the same color region, the color gain of the network is high. However, when the spot is small and the lines of the square are narrow so that the center and the surround regions of the opponent cells each receive different color inputs, the color gain of the network is low. The difference between the network response to the pattern and the response to the background is greater for the low spatial frequency stimulus than for the high spatial frequency stimulus. This is consistent with human color perception. As with color induction, color assimilation must be considered if one wishes to make a system which predicts human color perception.

DISCUSSION

We have shown that by using mechanisms similar to those found in the human visual system, one can create an artificial system which has good color constancy, flexibility in the types of images that it can handle, and very few assumptions about the images. Both simple and relatively complex backgrounds can be used and the color constancy behavior of the network simulation is not affected by gradients in the illuminant. In addition, because the network calculates its own reference level, it does not require the "grey world assumption", which is that the average reflectance chromaticity is the same for all images. As is true with the human visual system, the network does not produce perfect color constancy but only a shift towards constancy. Exact imitation of the human visual system, however, will require consideration of additional complex features, as will be discussed below, and necessitate more complex neural circuitry and processing stages not available in the present network.
Color induction is produced by the network in addition to color constancy. Although this could be interpreted as undesirable for some systems because it affects the accuracy of the color output, in many applications this could be an advantage because it enhances color contrast and therefore aids in object detection. In addition, because the magnitudes and directions of the color shifts are consistent with human color psychophysical data, the color induction behavior of the network would be needed in a system intended to mimic human color perception. Another aspect of human color perception, color assimilation, is also produced by the simulation in the spectrally opponent layer of the network. The human visual system also displays luminance assimilation, although the magnitudes and spatial scales of color and luminance assimilation are somewhat different [1]. Luminance assimilation is not produced in the current network because of the enhancement, rather than diminishment, of high spatial frequency luminance edges. In the biological system, luminance assimilation may be caused by factors not included in the current simulation such as light scatter and excitatory lateral neural connections such as cone-cone gap junctions.

This model has several benefits in overcoming restrictions associated with the computation of color constancy. The concept of spectrally-specific contrast, included in the current model, is similar to the independent "lightness" channels of the Retinex algorithm [31, 36, 38, 48]. Several "lightness" algorithms which have been proposed for color constancy, have been shown to be mathematically equivalent to a local spatial derivative plus a normalization term [23]. Some have argued (see review [26]) that the Retinex is similar to Von Kries adaptation in that each is a renormalization of color channel activities relative to some white reference. In that sense, the adaptation stage and the spectrally specific contrast stage in this simulation are also similar. However, there are several important differences which allow the operations described here to each cover different stimulus conditions and to cooperate in their contributions to color constancy.

First, the cone adaptation stage has a permanent white reference set by the midpoint of the range of possible threshold values. The adaptation stage also has a long-term adaptation reference which is usually close to neutral because it is established through exposure to many different stimuli over a long period of time. Faster, more localized adaptation effects are deviations from this long term reference. The reference for the spectrally specific contrast stage is the activity of the "local reference" cells. The spectrally specific contrast reference is not fixed, but instead changes with each new image. This reference is also not usually neutral. The cone specific contrast mechanism described here is also different from most "lightness" algorithms in that the normalizing
reference is measured locally, rather than a globally. The local reference is used in a push-pull mechanism with a spatially weighted global measure of spectrally specific contrast.

Second, the spatial profiles of the adaptation and V4 mechanisms are different. The effect of localized adaptation depends more heavily on the central test spot, while the effect of the large surrounds in the spectrally specific contrast operation are more affected by the background stimulus. The effect of the spectrally specific contrast operation is increased by the spectral and spatial opponency of the preceding stage which enhances color and brightness contrast.

There are several improvements to be made to the network. An obvious next step is to use "real world" images. There are two additions that need to be made to the network in order to allow it to handle real images appropriately. The first is spatial scale invariance. In the human visual system, the color of an object does not depend on the size of that object, provided that the surrounding objects are also scaled proportionately. In the current network simulation, however, the size of the test color patch must be appropriate for the size of the receptive fields of the cells in each of the network layers. One possible solution would be to incorporate dynamic receptive fields into the network which would change to match the size of each uniform color region in the image.

Another possible solution to the scale invariance problem would also incorporate the second aspect of real world imagery which remains to be addressed, that of surface segmentation and depth perception. Color and brightness perception are known to depend on the perceived depth plane of the object of interest relative to its surrounding surfaces, with color constancy and induction effective only within a single depth plane [45, 55]. Feedback from a surface segmentation process which assigns a relative depth to each surface, such as that described by Finkel and Sajda [16] could enhance or inhibit the color inducing effects of each region within the final stage of the network.

Each of the stages in the network contributes to the overall color constancy behavior. The final stage is a novel mechanism based on the chromatic properties of cells found in cortical area V4. The explicit calculation of a local reference and the use of a push-pull mechanism to incorporate the contrast responses in the final output, in addition to producing color constant responses for most images, allows the network to give accurate color responses for images which contain little chromatic contrast. The adaptation stage increases the dynamic range of the system, increases its flexibility to incorporate different viewing conditions, and also enhances the color constancy abilities of the system. The middle stage, the spectrally and spatially opponent mechanism, serves to enhance contrast, and therefore constancy, particularly at high spatial frequency luminance
edges. Together, the stages of the model comprise a system which produces good color constancy while maintaining flexibility.

**Acknowledgements:** This work was supported by ONR grants N00014-93-1-0681, by AFOSR grants NL 91-0082 and 92-J-0316, and by grants from The Whitaker Foundation, and The McDonnell-Pew Program in Cognitive Neuroscience.
References


Figure Captions

FIG 1 Detail of the cortical (top) and retinal (bottom) stages of the simulation. White triangles represent on-center cells, shaded triangles are off-center cells, and large circles are interneurons. Synapses are shown by small ovals, in white for excitatory, black for inhibitory. The silent surrounds in the cortical layers have an exponential synaptic weighting function as is shown inside the interneuron providing input to each contrast cell. In the retinal layers, the adaptation mechanism receives Gaussian weighted input from nearby cones. The adaptation cell then provides feedback to the cone at the center of its receptive field, changing that cone's threshold. On- and off-center cells in the retina have a difference of Gaussians synaptic weighting profile, as is shown inside each cell.

FIG 2 The Vos-Walraven cone spectral sensitivity functions. Figure adapted from [52].

FIG 3 Response curves for cones in the simulation under a range of values for the adaptation threshold. (a) Shows the sigmoidal limits of the adaptation range. The luminance level of the adapting stimulus was increased linearly, but the threshold values reach an asymptote at both ends of the range. (b) Same as (a) in log-linear coordinates.

FIG 4 Illustration of the simulated method for producing "aperture colors". The aperture color illuminant condition has the standard neutral illuminant on every part of the image except the test spot. The test spot is illuminated exclusively by the test illuminant, "illuminant 2". In the natural viewing condition, there is a single source which illuminates the entire image.

FIG 5 The power spectra of several natural illuminants [56]. The solid line is a full radiator of color temperature 2856K, similar to incandescent lighting. The others are all phases of daylight. Long dashes represent direct sunlight; short dashes, natural daylight averaged over all times of day; dotted line, overcast skylight; long odd dashed, natural daylight; and short odd dashed, clear blue sky. The illuminants are all very similar in the middle wavelengths but vary in their relative amounts of power in the long and short wavelengths.

FIG 6 Color constancy shifts made by the network for three Munsell reflectances (2.5yr7/10, 10b6/10, 2.5bg6/10) with the illuminants shown in figure 6. The axes show the relative proportions of each of the three network outputs. The plots shown are sub-
spaces of the full graph in which the ratios between each of the outputs and the sum of the other two range from 0 to 1. Each of the three corners of the full triangular graph represents activity in a single channel only, as would result from a monochromatic light stimulating one color channel exclusively. The squares represent the eigencolors of the reflectances, the response of the network with neutral illumination. The solid circles are the non-constant responses to aperture colors as explained in "Results: i. measuring constancy". The displacements of these symbols, from the square in each graph, represent the physical change in the color signal as the illuminant is changed. The aperture color closest to the $O_B$ corner of the graph is for the blue skylight illuminant, and the aperture color closest to the $O_R$ corner of the graph is for the incandescent illuminant. Responses for all other illuminants fall between these two extremes. The open symbols show the test stimulus responses of the network in natural viewing conditions. The open symbols show a shift toward constancy, away from the aperture colors and toward the eigencolors. Arrows show the direction of the shift.

**FIG 7** Color constancy shifts for Munsell reflectances 2.5bg6/10, 2.5g7/6, 2.5yr7/10, and 5.0r5/12. The illuminants used here are the combination of a spectrally flat illuminant and one of three illuminants with Gaussian spectral distributions peaking at 440nm, 560nm, and 660nm. The physical color shifts produced by these illuminants (distance from eigencolor to aperture color) are larger than those produced by natural illuminants in figure 7. Therefore, although the percent color constancy shifts are similar for the natural and artificial illuminants, the variances for the test stimuli outputs are smaller for the natural illuminants. As in figure 7, solid symbols mark aperture colors and open symbols mark test stimuli responses. Eigencolors are shown by the squares. Arrows are shown to indicate the direction of shift the reflectance-illuminant pairs.

**FIG 8** Color constancy with a complex background, a Mondrian, for two reflectances, 10gy7/10 and 10b6/10 which were placed at the center of the Mondrian. Responses are shown for incandescent light (A, shown in fig. 5) for both reflectances, for direct sunlight (B) with reflectance 10gy7/10 and for blue skylight (D75) with reflectance 10b6/10. Solid and open symbols are as in figs. 6 and 7. The squares are the neutral illuminant, Mondrian background outputs and so represent the color constant response, but do not eliminate the color induction effects of the Mondrian background. The triangles mark the eigencolors, the responses of the network under neutral illumination with a neutral homogeneous background. The differences between the neutral illumination, Mondrian background outputs and the eigencolor responses show the induction effect of the...
Mondrian background, independent of illumination changes. Because the test responses were obtained with the Mondrian background, the constancy shifts are toward the neutral illuminant, Mondrian background responses.

**FIG 9** Change in output ratios of the network with various offsets and gradient amplitudes in the input image. The results show a shift toward $R_{out}$ for a positive $R$ offset and a shift toward $G_{out}$ for a positive $G$ offset, as expected. Because all operations in the network are circularly symmetric, there is no effect from the addition of a symmetric linear illuminant gradient to the stimulus.

**FIG 10** (a) Simulation image used to measure the spatial properties of color induction. Width of the gap and of the surround can be varied. The center patch is 3x3 input units, the same size as the centers of the V4 receptive fields. (b) Results for two examples of color induction. The y-axis shows the size of the induction shift measured as a change in the ratio of two of the outputs due to the presence of the inducing surround. For the circles, the center reflectance is blue-green and the surround is blue. For the diamonds, the center is purple and the surround is red. As the gap between center and surround is increased, the induction effect decreases. At 4 units separation, the surround of the stimulus is outside of the silent surrounds of the V4 cells. (c) With no gap, the width of the surround is varied. As the width of the surround increases, the amount of induction increases. The amplitudes of the induction shifts for the two images are dependent upon the contrast between the center and surround colors.

**FIG 11** Demonstration of color assimilation in the spectrally opponent layer of the network. The input stimulus for each plot is a red spot and a red square with a yellow background. The two stimuli vary only in the width of the lines comprising the square and the size of the central spot. On the right, the coarse pattern (the wider lines and larger spot) results in a higher contrast in the response of the network layer cells. With the fine pattern (left), the contrast is reduced. This can also be seen in the cross-sectional profiles of the response of the network layer. The fine pattern is shown by the dashed lines and the coarse pattern by the solid lines. The background near the square has a higher $R$ response (looks more red) and the pattern itself has a lower $R$ response (looks more like the yellow background).
Table 1: Each of the most significant parameters in the simulation is presented along with the criteria used to determine that parameter's value. In parentheses are the values used for the results presented here and the range of possible values.

<table>
<thead>
<tr>
<th>parameter</th>
<th>description</th>
<th>factors in choice of parameter value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_{ij}$</td>
<td>connection strength between cells</td>
<td>spatial weighting chosen to create receptive field shapes found physiologically, amplitude chosen to keep all stages within linear range of response function</td>
</tr>
<tr>
<td>$\sigma_i$</td>
<td>threshold of cell $i$</td>
<td>chosen so that most inputs fall in middle of cell's response range, cone threshold changes with adaptation state</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>slope of linear portion of cell's response function</td>
<td>chosen in combination with $\sigma_i$ to give the appropriate dynamic range for each processing stage</td>
</tr>
<tr>
<td>$\theta$</td>
<td>width of adaptation weighting function</td>
<td>small value for simulation of fixation or short presentation time of image, large value for viewing conditions with free eye movements ($\theta=3.0$, relatively small compared to cortical silent surrounds, large compared to center of spectrally opponent receptive fields; $0&lt;\theta&lt;\text{diameter of image}$)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>fraction of total long term adaptation achieved</td>
<td>dependent upon length of viewing time ($\alpha=0.2$; $0&lt;\alpha&lt;1$)</td>
</tr>
<tr>
<td>$c_1$, $c_2$</td>
<td>coefficients for push-pull mechanism</td>
<td>chosen together with $a$ to give a total average constancy shift of 20% or greater for nearly all reflectance-illuminant combinations ($c_1=c_2=0.25$; $0\leq c_1 \leq 1$; $0\leq c_2 \leq 1$)</td>
</tr>
</tbody>
</table>
Cortical mechanisms

- Interneurons
- Spectrally and spatially opponent cells
- On-center cells
- Off-center cells
- Positive contrast
- Negative contrast
- Center input
- Surround input
- Reference
- Output

Retinal mechanisms

- Spectrally and spatially opponent cells
- Cones
- Change of Threshold (Adaptation)
- On-center
- Off-center
- Spatio-temporal integration
Cone Responses with Various Threshold Values

(a)

Cone Responses with Various Threshold Values

(b)
Aperture viewing condition

Natural viewing condition, for test stimuli
Power Spectra of Six Natural Illuminants

- full radiator, 2856K
- overcast, 5503K
- direct sunlight, 4874K
- natural daylight, 6504K
- average daylight, 6774K
- blue sky, 7504K
Color Constancy with Natural Illuminant Changes

Yellow-Red Reflectance

Blue Reflectance

Blue-Green Reflectance

Legend:
- ● yr aperture
- ○ yr test
- ■ yr eigencolor

Legend:
- ● b aperture
- ○ b test
- ■ b eigencolor

Legend:
- ● bg aperture
- ○ bg test
- ■ bg eigencolor
Color Constancy with Larger Illuminant Changes

Blue-Green Reflectance

Green Reflectance

Yellow-Red Reflectance

Red Reflectance

- bg aperture
- bg test
- bg eigencolor

- g aperture
- g test
- g eigencolor

- yr aperture
- yr test
- yr eigencolor

- r aperture
- r test
- r eigencolor
Color Constancy with a Mondrian Background

Green-Yellow Reflectance and Illuminants A, and B

Blue Reflectance and Illuminants A, and D75

- gy aperture
- gy test
- gy neutral illum, mondrian bkgd
- gy eigencolor

- b aperture
- b test
- b neutral illum, mondrian bkgd
- b eigencolor
Changes in Output due to Chromatic Gradients and Offsets in Illuminant

Gradient Color and Amplitude
(a) Diagram showing different regions: background, surround, gap, center.

(b) Graph with axes labeled 'surround width (input units)'. Two lines: one for bluegreen center, blue surround; another for purple center, red surround.

(c) Graph with axes labeled 'gap width (input units)'. Two lines: one for purple center, red surround; another for bluegreen center, blue surround.
Assimilation in Fine Color Patterns

On-center R response to Fine Red/Yellow Pattern

On-center R response to Coarse Red/Yellow Pattern