Modeling a space-variant cortical representation for apparent motion

Jeremy Wurbs

Center for Computational Neuroscience and Neural Technology
Program of Cognitive and Neural Systems, Boston University, Boston, MA, USA ✉

Ennio Mingolla

Department of Speech-Language Pathology and Audiology, Northeastern University, Boston, MA, USA ✉

Arash Yazdanbakhsh

Center for Computational Neuroscience and Neural Technology
Program of Cognitive and Neural Systems, Boston University, Boston, MA, USA ✉

Receptive field sizes of neurons in early primate visual areas increase with eccentricity, as does temporal processing speed. The fovea is evidently specialized for slow, fine movements while the periphery is suited for fast, coarse movements. In either the fovea or periphery discrete flashes can produce motion percepts. Grossberg and Rudd (1989) used traveling Gaussian activity profiles to model long-range apparent motion percepts. We propose a neural model constrained by physiological data to explain how signals from retinal ganglion cells to V1 affect the perception of motion as a function of eccentricity. Our model incorporates cortical magnification, receptive field overlap and scatter, and spatial and temporal response characteristics of retinal ganglion cells for cortical processing of motion. Consistent with the finding of Baker and Braddick (1985), in our model the maximum flash distance that is perceived as an apparent motion (D_{max}) increases linearly as a function of eccentricity. Baker and Braddick (1985) made qualitative predictions about the functional significance of both stimulus and visual system parameters that constrain motion perception, such as an increase in the range of detectable motions as a function of eccentricity and the likely role of higher visual processes in determining D_{max}. We generate corresponding quantitative predictions for those functional dependencies for individual aspects of motion processing. Simulation results indicate that the early visual pathway can explain the qualitative linear increase of D_{max} data without reliance on extrastriate areas, but that those higher visual areas may serve as a modulatory influence on the exact D_{max} increase.

Introduction

The human visual system has highly nonuniform sampling of visual input in both the space and time domains. This nonuniform, or space-variant, sampling greatly affects human perception of motion from a visual scene. These space-variant factors include cortical magnification, receptive field overlap, receptive field scatter, and retinal ganglion spatiotemporal response characteristics. Despite the pivotal role that these sampling features play in the perception of motion, relatively little is known about how the individual space-variant components of the visual processing pathway affect the perception of motion. Our study teases apart the role of cortical magnification by constructing a neural, mathematical model that incorporates a number of space-variant sampling components in their simplest form. We tie the corresponding relationship between the model output and the motion percept by constraining the model to the psychophysical function D_{max}.

For a given experimental setup, the psychometric function D_{max} is a measure of the maximum stimulus displacement that yields a percept of motion. D_{min} is directly analogous for the minimal stimulus displacement. Baker and Braddick (1985) found that D_{min} and D_{max} increase nearly linearly as a function of eccentricity up to at least 10°. The cause of this linear increase, however, is not known. Larger receptive fields in the periphery might account for larger D_{max} by

accumulating broader motion signals within the required integration time to perceive motion. It could be counterhypothesized, however, that faster peripheral processing via the magnocellular pathway could decrease the integration time necessary to combine motion signals and thus decrease D_{max} in the periphery. A major question, then, is whether early visual processing (retina through V1) can account for the observed change in D_{max} across eccentricities or whether this change requires higher level visual processing.

Previous models used to explain D_max data

Previous neural models have successfully explained a number of psychophysical findings regarding D_{max}. Eagle (1996) describes a model "in which direction discrimination is based on the nearest-neighbor matching of zero-crossings in the output of a single-spatial-filter bandpass in both spatial frequency and orientation;" the author uses this model to account for D_{max} psychophysics data for spatially broadband motion patterns, arguing that D_{max} is determined by the coarsest spatial filter activated by each stimulus. Glennerster (1998) use an implementation of the MIRAGE model (Watt & Morgan, 1985) to account for data on D_{max} as a function of stimulus dot density. Morgan (1992) considered data on increasing D_{max} as a function of increasing stimulus element size and argued that such data could be explained using a model where a spatial filtering step removes fine detail in a stimulus pattern before motion processing occurs. Tripathy, Shafiullah, and Cox (2012) describe a model based on a set of Reichardt-type local detectors whose radius of their catchment areas scale proportionately with the displacement that they are tuned to detect; this model is able to account for D_{max} data where correspondence noise is a major contributing factor.

To the knowledge of the authors, however, no previous model has explained D_{max} data showing that D_{max} increases nearly linearly as a function of eccentricity. The model presented herein elucidates this data by taking into account space variant factors present in the early visual system. How the visual system processes motion across a wide range of biological visual parameters, including cortical magnification, receptive field overlap and scatter, and spatial and temporal response characteristics of retinal ganglion cells are key to understanding how D_{max} changes at different eccentricities. By using a point spread function to compute motion signals and a discriminability index based on the model cortical activity difference between continuous and discrete inputs to determine D_{max}, our model is able to show that the space-variant factors present in the early visual system

(retina to V1) are sufficient to account for increasing D_{max} as a function of eccentricity.

Space variance in the visual system

The human visual system has evolved to process light input that enters the eye in a highly nonuniform manner. Perhaps the most notable aspect of this nonuniformity is the clarity with which we perceive objects projecting to our fovea versus periphery. This difference in clarity arises from both the sampling density of retinal receptors as well as the subsequent neural processing. Motion processing is also greatly affected by the same space-variant sampling processes, but in more subtle ways. It is not so clear that the fovea should be better suited to process or perceive motion despite its greater spatial resolution; indeed, motion processing is perhaps the one case where periphery vision may be better suited than the fovea (Westheimer, 1983, Baker & Braddick 1985). Why our visual system has evolved to be so dominantly space-variant is unknown, but a back-of-the-envelope computation from Bonmassar and Schwartz (1997) estimates that for the human brain to maintain foveal resolution throughout the entire field-of-view, the brain would have to weigh upwards of a few metric tons. Additional research has indicated that space-variance in the visual system may be useful for more than simplifying computational complexity. Space variance has been a key component in systems designed for segmentation (Mishra & Aloimonos, 2009), time-to-contact estimation (Tistarelli & Sandini, 1993), and robot navigation (Engel et al., 2009).

In this study we aim to determine what, if any, aspects of the early visual system's space variance are responsible for the observed increase in D_{max} as a function of eccentricity. In the following sections we describe several space-variant properties present in the human visual system and how they relate to one another in the context of cortical sampling of the retinal image.

Cortical magnification factor

Cortical magnification (CM) is a measure of how much cortical length is dedicated to process a stimulus of a given visual angle. In most space-variant visual systems, the amount of cortical surface dedicated to processing the central visual field is far greater than for the peripheral field. In humans this variance is generally about two orders of magnitude (Daniel & Whitteridge, 1961). The quantitative measure of CM is called the cortical magnification factor (CMF) and is typically measured in millimeters of cortical length per degree of visual angle (Figure 1). The exact measure of CMF
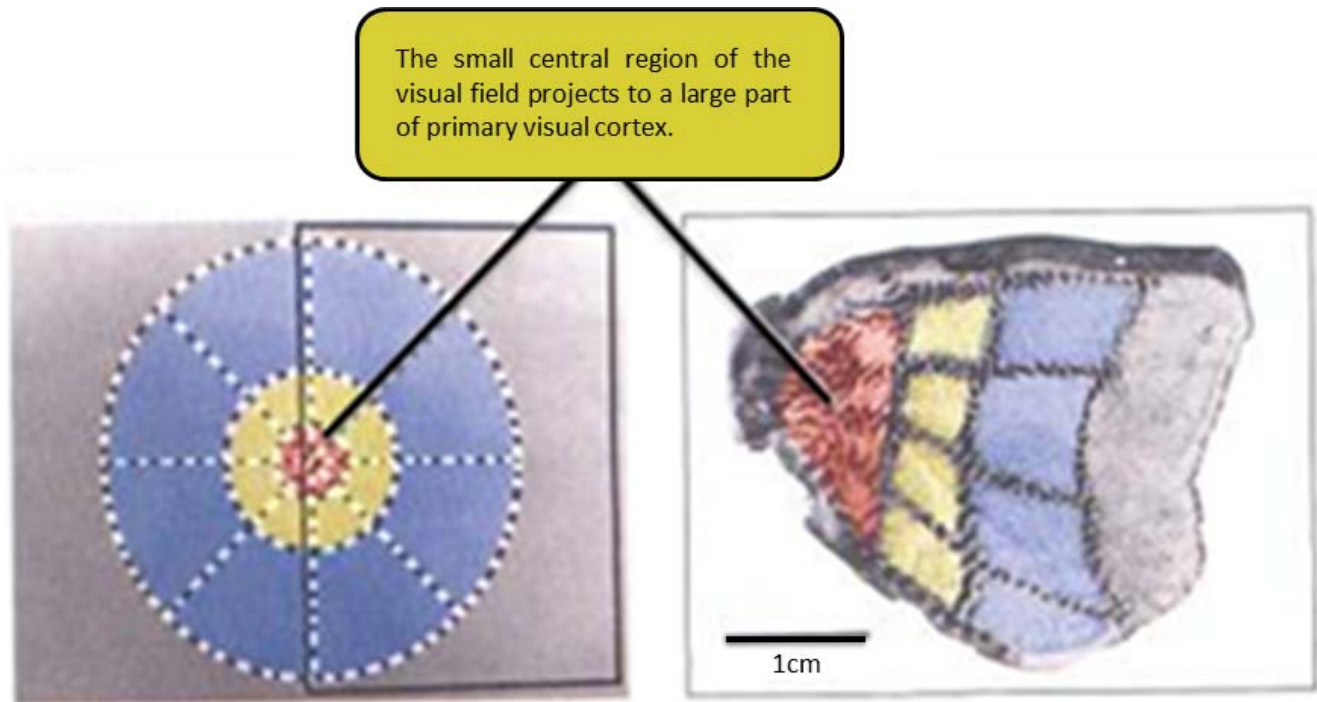
Figure 1. Two-DG staining on Macaque V1. *Left*: Color-coded visual input. *Right*: Color-coded stain of Macaque V1 after being shown the input on the left. Note that, per unit area, much more cortical area is dedicated to the fovea compared to the periphery. This allocation of visual processing resources is known as cortical magnification and is expressed through the cortical magnification factor, or the ratio of millimeters of cortical length per degree of visual angle. [Figure adapted from Tootell et al. (1982) with permission].

differs between cortical areas. Area V2, for example, has more enhanced foveal representation than V1; as much as 50% of Marmoset monkey V2 is dedicated to the central 5° of the visual field (Rosa, Fritsches, & Elston, 1997). Also, compared to V1, area V4 (which is believed to be involved in fine shape and texture analysis) has very little area dedicated to the periphery whereas other areas, such as V6 (which is believed to be responsible for analyzing self-motion), show a more gradual decline in the amount of cortical area dedicated to the periphery (Daniel & Whitteridge, 1961).

To the knowledge of the authors no model to date has incorporated CMF into determining the maximal displacement for perceived motion. The present work explores the functional role of higher CMF in the foveal region compared to the periphery. It is not trivially clear whether a greater CMF in the fovea would contribute to lower D_{max} in this region.

Receptive field size

The concept of a receptive field is used to describe individual neurons. The receptive field for a particular neuron is defined as the region in stimulus space that alters the firing of that neuron. For visual cortical neurons this correlates to the region on the retina wherein a change in the stimulus affects the firing rate of that neuron. Receptive fields have complex shapes but are often taken to be near-circular for quantitative purposes. The receptive field size for visual cortical neurons is most often measured in degrees of visual angle and increases with stimulus eccentricity. Note that this property is distinct from cortical magnification, as it is theoretically possible to have nearly any combination of cortical magnification and receptive field size for a given sheet of cortical neurons. Since receptive field sizes increase as a function of eccentricity (Figure 2) while CMF decreases, it seems reasonable to think that visual cortical architecture contains some inherent constant factor. Others have assumed that this constant factor is the product of CMF and receptive field size, that some have called point image size, to conflicting results (Hubel & Wiesel, 1974; Dow, Snyder, Vautin, & Bauer, 1981; Van Essen et al., 1984; McIlwain, 1986). More recent studies, however, have yielded additional credence to the hypothesis that the point image size in V1 is indeed constant or very slightly-increasing (Harvey & Dumoulin, 2011; Palmer et al., 2012). A related hypothesis is that receptive field overlap might be constant (Bolduc & Levine, 1998). While not mathematically equivalent, the assumption of constant receptive field overlap combined with a linear increase/decrease in receptive field size/CMF, respectively, can yield a near-constant point image size for a range of constant overlap constants. Our model
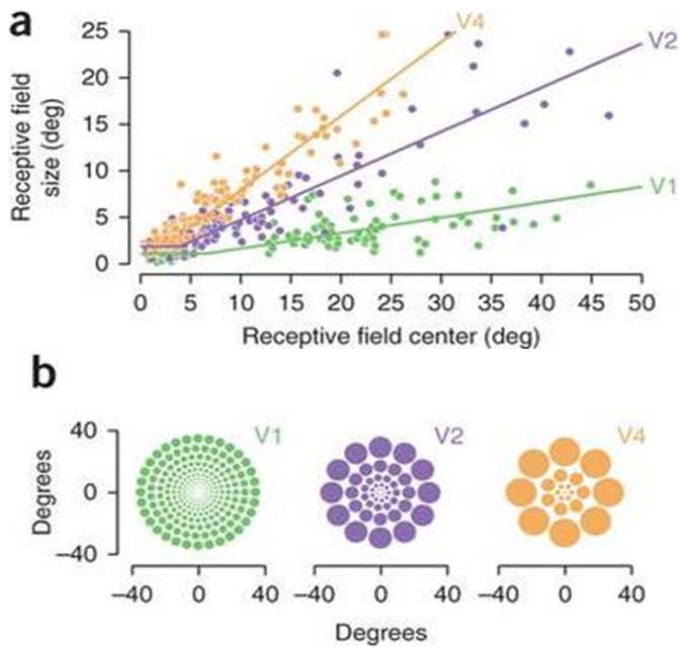
Figure 2. Schematic comparison of receptive field sizes across eccentricities for different cortical regions. (Figure taken from Freeman and Simoncelli, 2011, with permission).

uses this assumption of constant receptive field overlap in order to constrain the relationship between CMF and receptive field size.

Receptive field overlap

Receptive field overlap refers to the portion of adjacent receptive fields that share the same stimulus space. Assuming near-circular receptive fields, receptive field overlap can be computed knowing the receptive field size and centroid locations for each point in the visual field. For primates, biological evidence suggests any given region of the retina can be part of as many as 35 ganglion cell receptive fields (Braccini, Gambardella, Sandini, & Tagliasco, 1982; Bolduc & Levine, 1998). This number appears to be relatively constant throughout the entire visual field, thus yielding a large constraint that can be applied when determining the model parameter values for CMF and RF size. Our model incorporates this overlap factor in order to constrain both the point image size, as well as the model's cortical columnar structure.

Cortical receptive field scatter

The primary visual cortical surface is topographically organized. That is, physically adjacent neurons in the cortex will have adjacent receptive field centroids. This topographical organization means that as one moves through a cortical column all neurons would be expected to have the same centroid location. To a first
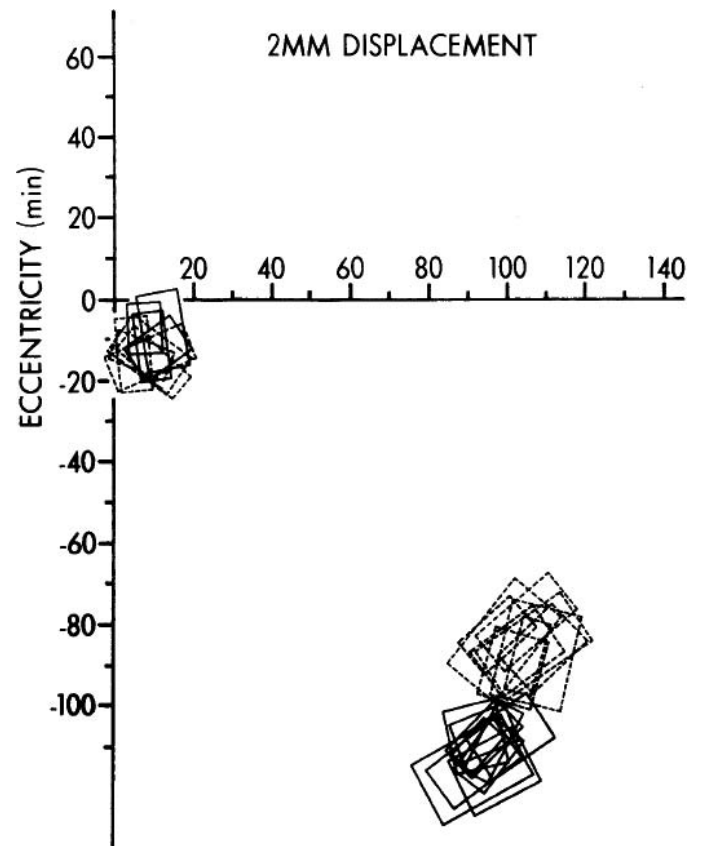

Figure 3. "Overlap of striate receptive fields recorded in pairs of vertical penetrations displaced 2 mm from one another on the cortical surface. For each pair, dashed lines represent receptive fields from one penetration; solid lines represent receptive fields from the other penetration. One pair of penetrations was made in the central foveal representation (15-foot eccentricity), the other in peripheral representation (110-foot eccentricity)." Used with permission from Dow et al. (1981).

order approximation this assumption holds true. There is, however, still a measurable deviation of receptive field centroid locations even within one column. The amount of deviation of receptive field centers as one moves down a cortical column is called cortical receptive field scatter (Figure 3).

There is debate as to whether this scatter has a functional role in refining acuity perception or is simply added noise. While the model presented here does not seek to answer what exact functional role for cortical receptive field scatter might serve, it does, however, seek to place an upper limit on the amount of added noise that such cortical scatter could place on cortical motion processing in V1.

Receptive field centroid density

Referring to a model of overlapping, circular receptive fields with well-defined centers, the receptive

field centroid density is the density of receptive fields per unit area (2D case) or per unit length (1D case) measured in retinal coordinates. As changing the RF centroid density would alter model cortical magnification values, such a change would affect D_{max}. For the purposes of this study, RF centroid density is a derived measure that can be computed from known CMF and receptive field size values, along with the additional data constraint of constant or near-constant point image size (Harvey & Dumoulin, 2011; Palmer et al., 2012) derived from a constant receptive field overlap factor.

Spatiotemporal response

The previous visual sampling parameters have all been defined in the spatial domain. The response of actual neurons, however, is also a function of time. That is, their responses are more accurately described in a spatiotemporal domain wherein the output of a neuron to a particular spatial stimulus changes over time. The exact nature of the spatiotemporal response is important when modeling speed-tuned neurons. Adelson and Bergen (1985) describe how to set up linearly separable spatiotemporal filters that are functionally equivalent to Reichardt detectors and posit a neural implementation for their filters. Our model extends this spatiotemporal description to allow each component, spatial and temporal, to vary independently as a function of eccentricity. By adding in this eccentricity-dependence we are able to more accurately track the spatiotemporal response of model neurons across V1 as the input stimuli moves across many degrees of visual space.

Apparent motion

Apparent motion describes a number of related phenomena wherein motion is perceived from one or more static images. Apparent motion can be subdivided by the manner in which the motion is perceived and the stimulus used, such as illusory motion, beta motion, and phi motion. In the case of multi-image stimuli it is common to use single flashes of light or arrays of random dots.

In the simplest case, a two-flash stimulus is used to produce a motion percept. Classically, there are three main stimulus parameters that dictate when a two-flash stimulus will yield apparent motion. These parameters were characterized by Wertheimer's student Korte in what have become known as Korte's laws, which describe the relationships between stimulus parameters over some ranges to stay at threshold for perceiving apparent motion:

1. *Separation versus Intensity.* Intensity must increase as the separation distance increases (and vice versa).
2. *Rate versus Intensity.* Intensity must increase as the flicker rate decreases (and vice versa).
3. *Separation versus Rate.* Flicker rate must decrease as the separation distance increases (and vice versa).

While the classic definition of D_{max} and D_{min} are meant to be constant for a given flicker rate and intensity, Baker and Braddick (1985) have also shown that they are both proportional to visual eccentricity.

Determining D_{max}

D_{max} and its counterpart, D_{min}, are perceptual functions. In other words, they measure the dependence of a perceptual phenomenon (the perception of motion) as a function of input spatial separation in a two-flash stimulus paradigm. Baker and Braddick (1985) measured D_{max} and D_{min} using a two-frame random dot stimulus. In this setup the subject is shown two frames of random dots wherein a single region of the image is replicated and shifted from frame to frame. The subject's task is to respond to the direction of motion of the replicated region (see Figure 4).

Baker and Braddick (1985) measured D_{max} and D_{min} and found that both measures increased linearly as a function of eccentricity (Figure 5). Psychometric experiments that are used to determine D_{min} and D_{max} must then provide a link between the perceptual phenomena they wish to discover and the psychometric data which they produce. Baker and Braddick (1985) provide this link by asking subjects to rate a perceptual measure (motion clarity) on a graded scale. Plotting this motion clarity measure versus performance yields a sigmoidal diagram. By selecting the center point of this sigmoid as a performance cutoff one can approximately define an analytic link between the performance-based measure observed in the psychometric study and the perception of motion.

The modeling study presented here continues in a similar vein; while our input stimulus follows a paradigm that can be directly linked back to apparent motion, our measurement of D_{max} follows an activity-difference metric. By measuring the Euclidean distance between model V1 network activity profiles we obtain a direct relationship between the input stimulus and the distinguishable network activity which can be used to derive a percept of motion. By defining a cutoff threshold for this Euclidean distance measure we effectively produce a performance-based metric. Aligning the model performances with that from Baker and Braddick (1985) we are able to derive a direct conversion between network model activity and the perceptual measure, D_{max}.
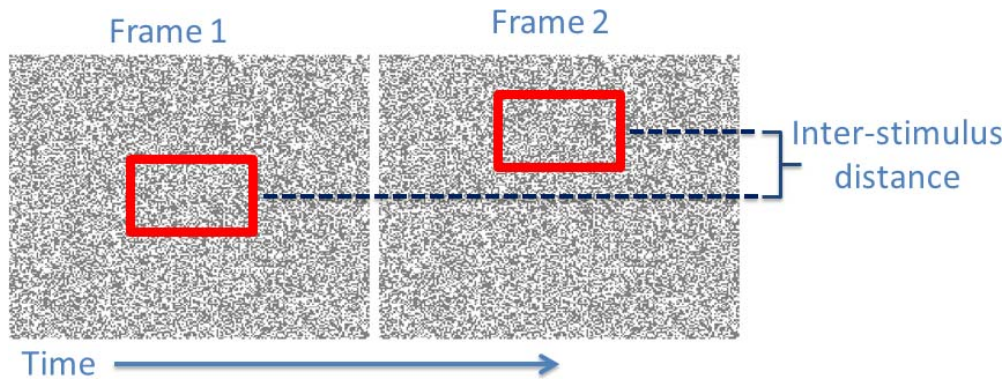
Figure 4. Input schematics for determining D_{max} psychophysically. The subject is shown the two random-dot kinematogram frames in sequence and asked which direction the replicated region moved (up or down). The maximal displacement for which the subject is able to correctly respond above a set percentage is called D_{max}. D_{min} is directly analogous for the small displacements.

What is the source of the increase in D_{max} as a function of eccentricity?

Behavioral data from the Baker and Braddick (1985) psychophysical experiments make it clear that there is an eccentricity-dependence in the visual system's processing of motion. Physiologically there are many sources of space variance in the visual system (Table 1) that could all contribute to this eccentricity dependence, and it is not trivially clear which of these physiological factors are necessary to explain the behavioral data. Because these factors interact together in a very nonlinear manner, teasing apart the role of each one individually will prove difficult to do through psychophysical and physiological methods. Instead, we construct a model of the early visual system that incorporates each space-variant factor in Table 1. We tie the model V1 activity output to D_{max} behavioral data through the use of a discriminability
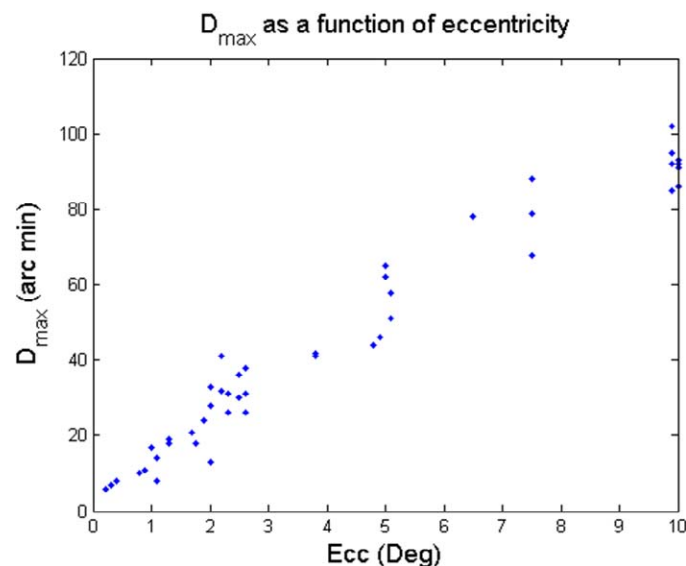

Figure 5. D_{max} data taken from Baker and Braddick (1985), plotting D_{max} as a function of eccentricity.

index that measures the difference in model V1 activity between the cases of discrete versus continuous input. The results of this modeling effort give credence to the idea that the increase in D_{max} as a function of eccentricity can be explained by early visual space-variant factors alone. Furthermore, behavioral data obtained by Todd and Norman (1995) that show an increase in D_{max} as a function of the number of input frames is *not* able to be explained by the current model (refer to the Discussion section *Role of Higher Level Visual Areas*). Thus the model also hints that the number-of-frames effect expounded by Todd and Norman (1995) might require processing from higher level visual areas (extrastriate areas) that is not present in the early visual system in order to be explained.

The rest of this paper is broken down as follows: the Methods section breaks down the major model components and how the model incorporates each of the individual space-variant factors; the Results section shows the model is able to fit D_{max} psychophysical data showing that D_{max} increases nearly linearly as a function of eccentricity while the model fails to fit D_{max} psychophysical data showing that D_{max} also increases with the number of input frames; finally a Discussion section summarizes the main findings of this paper and what the model data supports in terms of the early visual system's roles in space-variant motion processing.

Methods

The following section outlines the model, which includes three main neural correlate stages: (1) input sampling layer (corresponding to the retina), (2) shunting inhibition layer (corresponding to LGN/V1), and (3) a cortical layer (corresponding to V1).

Figure 6 gives a graphical representation of the model stages. The input stimulus (shown in Figure 8) to the model is a paired-flash stimulus, which is sampled

Sources of space-variance in the early visual system
Cortical magnification factor
Receptive field size
Receptive field overlap
Cortical receptive field scatter
Spatiotemporal response

Table 1. List of sources of space-variance in the early visual system.

by a number of spatially-offset retinal fields with different scales. The purpose of multiple, off-set retinal sampling fields is to allow the model's LGN/V1 neurons to receive input from retinal ganglion neurons with randomly-selected offsets in their receptive fields. This sampling mechanism allows the model to incorporate a measure of cortical receptive field scatter. The retinal layer is connected to a model LGN layer described by a feed-forward center-surround network which incorporates shunting inhibition (refer to Equation 3) in order to assure that the model remains normalized throughout a broad range of possible input intensities. The model V1 cortical architecture contains individual layers wherein each layer receives input from a different, offset retinal sampling field, thus incorporating cortical receptive field scatter. Model V1 averages activity across a single column, resulting in a cortical activity profile. D_{max} is computed by generating a difference measure between the activity profile

generated from a continuous input and the activity profile generated from the paired-flash stimulus (refer to the Discriminability indices section).

Model input

We simulate our network following the 1D flash stimulus paradigm shown below.

Figure 7 shows an example input to our model. The two input nodes receive dot (impulse) inputs while all other nodes receive the same, nonzero background activity.

While we tie our model's output to the psychometric-to-perception link provided by Baker and Braddick (1985), it is important to note the distinction between our input and that used in their experiment. Our stimulus is a two-flash input directly applicable to Korte's laws of apparent motion, wherein motion would be regarded as a perceptual phenomenon. The experimental setup used by Baker and Braddick (1985) consisted of random dot stimuli inputs under a performance-based psychometric task (refer to Figure 4). This experimental design removes the results from direct applicability to perceptual motion via Korte's laws in two ways: (1) the performance-based metric must be carefully calibrated to the perception of motion, and (2) there is the conflicting factor of background noise. The first factor has been discussed
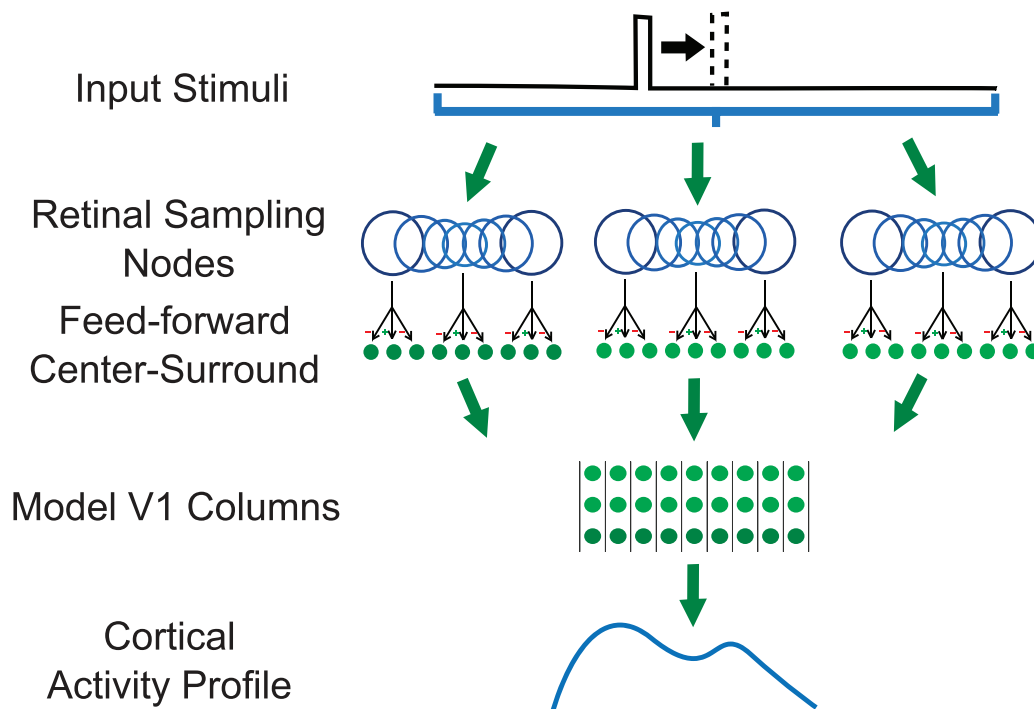

Figure 6. Model overview. The model consists of five main stages detailed in the following sections: (1) model input stimuli, (2) retinal sampling layer, (3) feed-forward center-surround layer (corresponding to LGN/V1), (4) V1 cortical columnar layer, and (5) aggregate cortical activity.
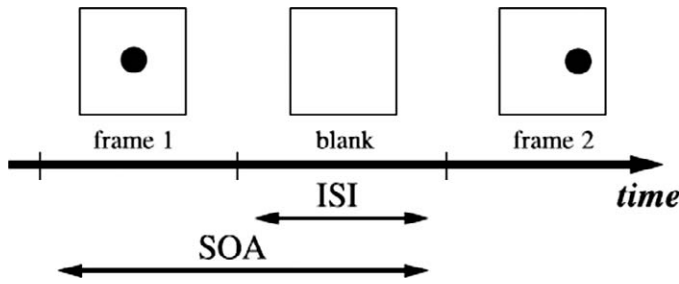
Figure 7. Two-flash stimulus paradigm. The stimulus consists of two frames separated by a blank. Stimulus onset asynchrony (SOA): the time from the onset of the first frame to the onset of the second frame; interstimulus interval (ISI): the time from the offset of the first frame to the onset of the second frame.

previously, but selecting an appropriate stimulus is important for our study for the second factor as well. Since the model is focused on early visual processing stages and does not rely on any direct motion processing, it is important that the final network activity not require any additional filtering or higher level visual processing to remove noise effects as these effects would corrupt the Euclidean difference metric used to determine the D_{max}.

Spatiotemporal receptive fields

Model retinal sampling follows the overlapping circular receptive field model described in Yamamoto, Yeshurun, and Levine (1996) (Figure 9). A central region in the retinal sampling field is defined to be the fovea. All receptive fields in this region have the same spatiotemporal properties with receptive field size r_f. The overlap constants are also held constant in this
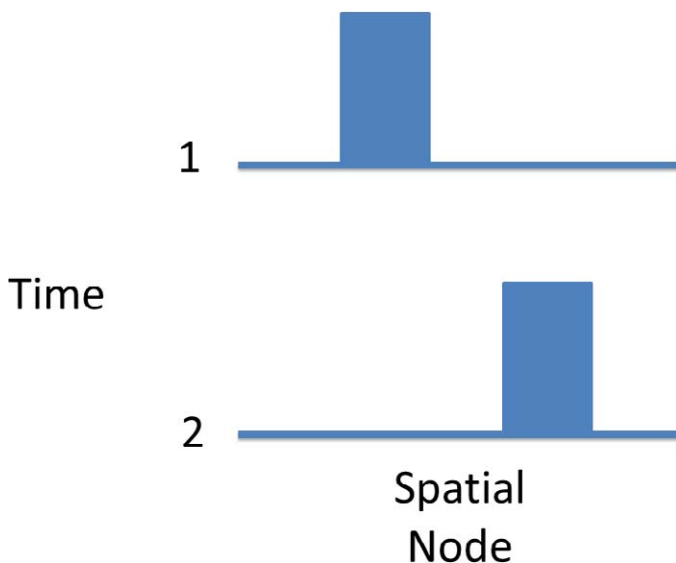

Figure 8. Example of two-flash input stimuli presented to model.

region such that the RF density and cortical magnification factor are constant. An iterative scheme is then used to create a space-variant sampling method one ring at a time using the following equation:

$$r_n(\alpha, \omega, r_n, n) =$$

$$\left(\frac{-\alpha \cdot (1 - 2\omega) - \left(\alpha^2 \cdot \left((1 - 2\omega)^2 - 1 \right) + 4 \right)^{\frac{1}{2}}}{(\alpha - 2)} \right)^n \cdot r_f,$$

(1)

$$c_n(c_{n-1}, r_{n-1}, \omega) = c_{n-1} + r_{n-1} \cdot \omega,$$

(2)

where r_n is the radius of the nth RF in the peripheral direction (note that in this notation $r_{n=0} = r_f$), c_n is the centroid location of the nth RF ($c_{n=0} = c_f$), α is a parameter that describes the ratio of RF radius to eccentricity (Figure 10), and ω is a parameter that describes the amount of overlap between RFs (Figure 11).

Cortical layers exhibit feed-forward center-surround excitation/inhibition

The output from the retinal layer feeds into the model LGN, which is described by a feed-forward competitive shunting network (Figure 12). Shunting inhibition is used to assure that the network remains normalized throughout a range of possible input intensities. The LGN is analytically described by:

$$\tau_i \frac{dx_{ij}}{dt} = -A_i x_{ij} + (B - x_{ij})$$

$$\cdot \sum_{k=1}^{n} I_k C_{ki} - (x_{ij} + D) \cdot \sum_{k=1}^{n} I_k E_{ki}, \quad (3)$$

where I is an input vector, x is the associated output activation vector, n is the number of units in both the input and output layers, t is time, B and D are shunting parameters, and C_{ki} and E_{ki} are Gaussian kernels. Note that the decay-rate parameter A_i, the rise-time parameter τ_i, and the Gaussian functions C_{ki} and E_{ki} are all a function of eccentricity. C_{ki} and E_{ki} are defined as:

$$C_{ki} = C \cdot e^{-\frac{(k-i)^2}{2\mu^2}}, \quad (4)$$

$$E_{ki} = E \cdot e^{-\frac{(k-i)^2}{2v^2}}, \quad (5)$$

where μ and v are the variances of the Gaussians, $(k-i)$ is the distance from the center of the receptive field centroid of node i to node k, and C and E are parameters. For the eccentricity-dependent parameters A_i and τ_i we tried many different functional forms and find that simple linear functions still allow enough
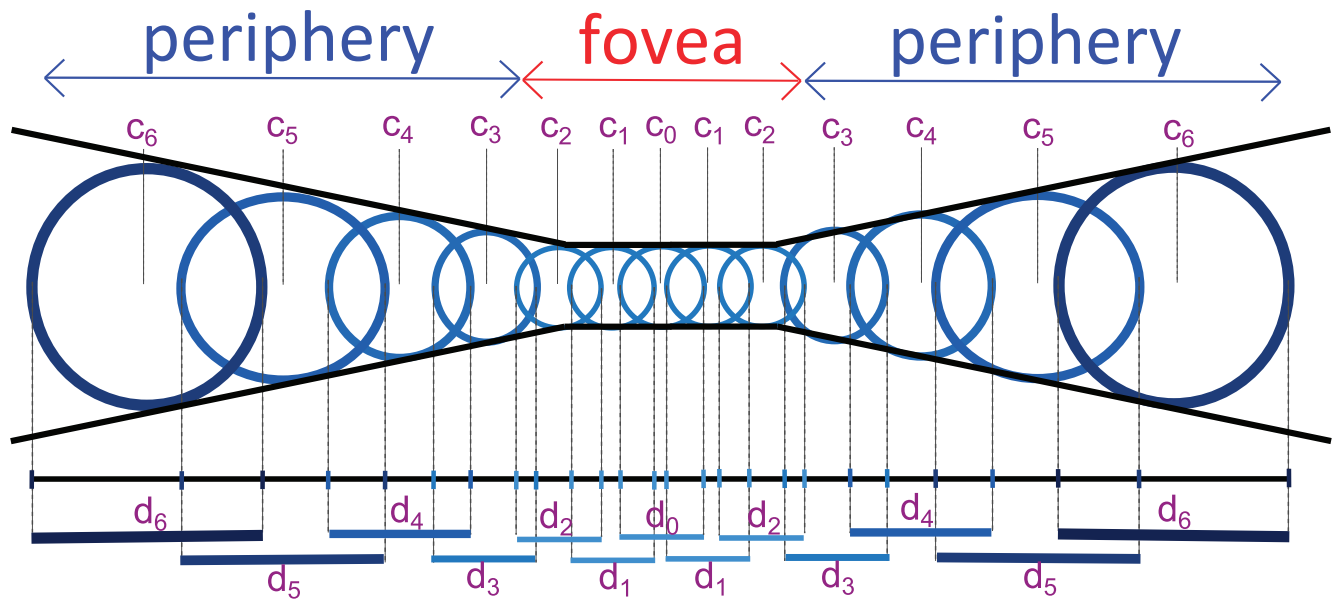
Figure 9. Model sampling. The 2D sampling paradigm from Yamamoto et al. (1996) is adapted to a 1D input space. The fovea is taken to be the central 1° of visual space; within this region receptive field characteristics are constant. Starting from the edge of the fovea moving towards the periphery, receptive field diameter increases linearly in accordance with known data (Freeman and Simoncelli, 2011).

flexibility in the model to accurately fit both the physiological and psychophysical data.

$$A_i = m_A(k - i) \cdot b_A, \tag{6}$$

$$\tau_i = m_\tau(k - i) + b_\tau, \tag{7}$$

where the individual parameter values can be found in Table 2.

Cortical columns incorporate cortical RF scatter

In order to simulate cortical scatter, multiple cortical layers are created with offset δ_j. That is, each RF in the jth retinal layer has a centroid location $c_n + \delta_j$ and RF size r_n (Figure 13).

The output of the LGN layer feeds into model V1, which sums the activity of each LGN input across all
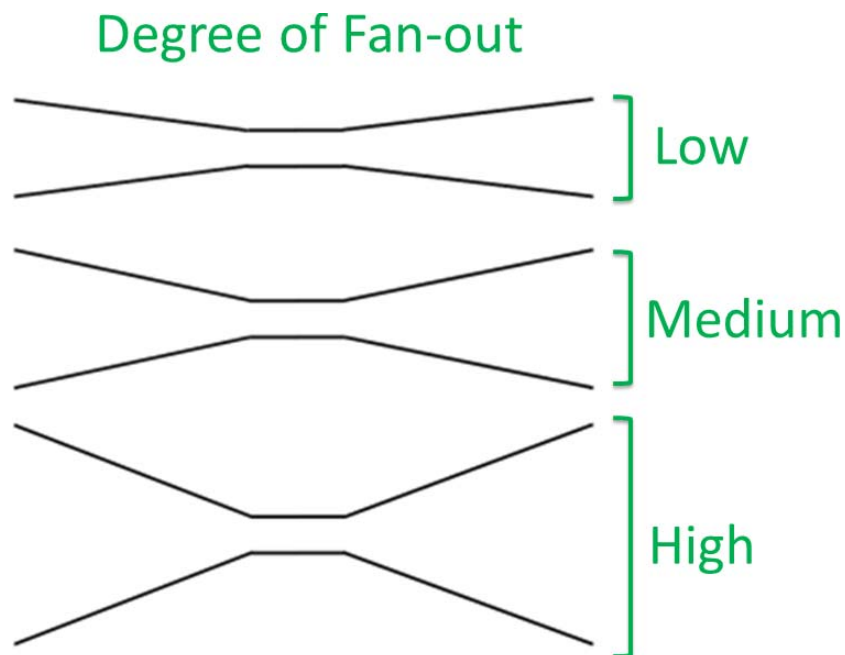

Figure 10. Displaying the degree of fan-out, determined by the parameter α.
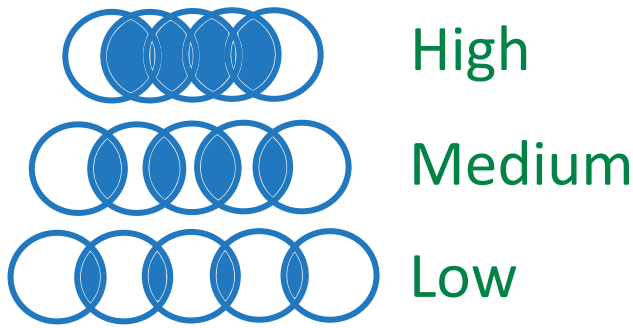
Degree of Overlap

Figure 11. Displaying the degree of overlap, determined by the parameter ω.

cortical scatter locations. That is,

$$x_i = \sum_j x_{ij}. \tag{8}$$

Cortical activity profiles

The model output to a single input stimulus is one described by "fat and fast" cortical activity in neurons with peripheral RFs and "slim and slow" cortical activity in neurons with foveal RFs (Figure 14). Because peripheral neurons have larger spatiotemporal parameter constants (dictated by m_A and m_τ) a single peripheral input stimulus will result in a broader, faster cortical response compared to cortical neurons with foveal receptive fields. While these response characteristics could be useful for spreading motion-selective neural activity, peripheral neuron activity also decays much more rapidly than their foveal counterparts. The interplay between fast, broad peripheral network activation and acute, slow foveal network activation plays a large role in determining the model D_{max}.

Discriminability indices

Once the model network activity is computed a measure of D_{max} must be derived. In order to determine D_{max} a difference measure is computed and a threshold is established to produce a value of D_{max} at varying eccentricities. The difference measure used is a straight Euclidean metric between the network activity resulting from a two-flash stimuli against the network activity resulting from a continuous stimuli input. That is, the model system is first shown an example of continuous motion and the resulting network activity is recorded. If this activity is similar enough (in a Euclidean sense) to a discrete, flash input then that input is said to produce a motion percept. If the activity difference is greater than some threshold, however, then that input is said to produce a discrete flash percept. Thus, if the network activity difference is below some threshold (selected to give an appropriate range of D_{max} values) then the network is determined to produce a motion percept. Figure 15 shows the difference measure for a wide range of interstimulus distances and for multiple eccentricities.

Results

Basic observations of the phenomenon

For our initial explorations, we sought to make sure that the model would yield eccentricity-dependent
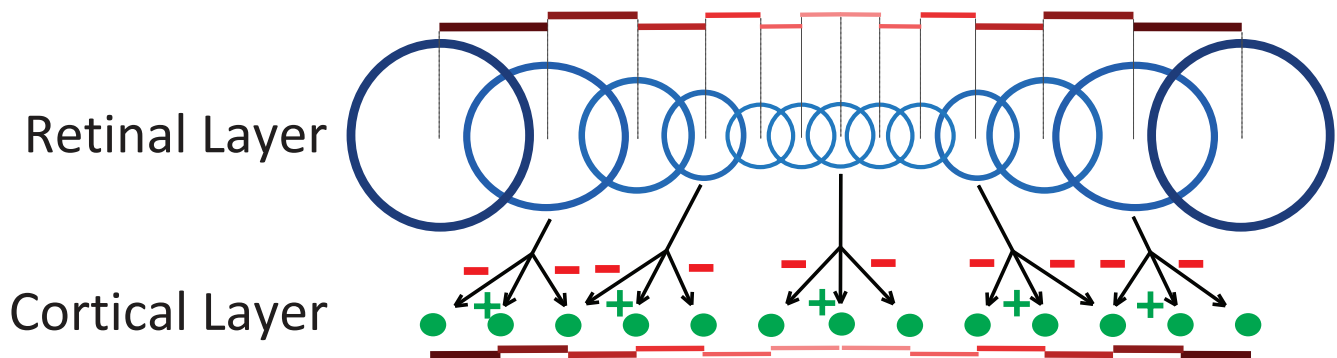

Figure 12. Diagram showing center-surround excitation/inhibition. A one-layer shunting network is used for normalization of the input space. Note that while the spatial connectivity in the cortical domain (bottom half of the diagram) is constant this connectivity is spatially skewed in the input domain (top half of the diagram. The result is that activity across the stimulus space in the far peripheral areas (in retinal coordinates) spreads wider through the model cortical layer than activity in foveal areas. Note that this effect is due solely to cortical magnification and is independent of the spatiotemporal characteristic differences between foveal and peripheral cortical neurons.

Parameter	Value
B	10
C	3
D	10
E	0
μ, v	0.1
m_A	0.2
b_A	0.01
m_τ	0.6
b_τ	0.03
α	0.25
ω	0.5
r_f	1.0
r_1	0.25
r_p	85
δ_i	0.1

Table 2. Model parameters. *Notes*: B, C, D, E: shunting parameters. μ, v: variances of the Gaussian kernels C_{ki} and E_{ki}, respectively. m_A, b_A slope: offset for retinal spatial eccentricity-dependence. m_τ, b_τ slope: offset for retinal temporal eccentricity-dependence. α: degree of receptive field fan-out. ω: degree of overlap. r_f: foveal radius. r_1: foveal receptive field size. r_p: field-of-view radius of the periphery. δ_i: scatter offset factor.

results based on standard model parameters and would not yield eccentricity-dependent results using non-variant space parameters.

The parameter α is used to determine the receptive field layout (in accordance with Equation 1); additionally for this basic condition the space variance of the network time constant (τ), decay rate (A), and Gaussian kernels (C_{ki} and E_{ki}) were all set proportional to α. That is, the slope of the τ, A, C_{ki}, and E_{ki} as a function of eccentricity is proportional to α so that when $\alpha = 0$ the aforementioned parameters are not space-variant.

We determine the exact values of D_{max} by selecting a difference measure threshold for Figure 16. Figure 17 shows D_{max} as a function of eccentricity after selecting a discriminability index of 0.5 for different alpha values (i.e., the slope of the eccentricity dependence for α, τ, A, C_{ki}, and E_{ki}). It is clear from Figure 17 that D_{max} eccentricity dependence increases as the eccentricity dependence (from all applicable parameters) increases as well.

Constant overlap factor

In the following sections we look at individual space-variant factors and their effect on the eccentricity dependence of D_{max}.

The model assumes near-constant overlap between receptive fields. That is, for any point in the retinal space there is a constant number of cortical receptive fields that contain that particular point. Bolduc and Levine (1998) use a constant overlap parameter of 35 receptive fields for each point in retinal space. We follow this precedence in determining our default model parameters.

In order to test the importance of this assumption to the model we vary the overlap parameter while maintaining its constancy across eccentricities. The results can be seen in Figure 18.

Note that changing the amount of RF overlap in the absence of other space-variant factors has no pronounced effect on the linear increase of D_{max} as a function of eccentricity (see Figure 18). For this reason we believe that above an absolute lower limit (two or three receptive fields) the exact overlap parameter is not as important as the assumption of its constancy.

Number of frames

Todd and Norman (1995) studied how the number of frames presented to an observer changes D_{max} at multiple eccentricities (Figure 19). In their experiment observers were required to identify the shapes of moving targets and discriminate regions of motion from regions of uncorrelated noise. The maximal range for which an observer could perform a particular task above a set accuracy threshold was taken to be D_{max}.

The increase in D_{max} as a function of increasing number of frames appears to be a robust phenomenon. Despite this robustness, however, little is known about its mechanism. In order to determine whether this phenomenon can be explained from lower level visual processes alone we tested the model to an increasing number of frames, finding the associated D_{max} for the last [pair of] frames in each trial. The model results can be found in Figure 20.

As seen in Figure 20, there is no observable change in model activity with an increase in number of frames shown. The cause of this result is directly related to the fact that the network activity dies down far too quickly to allow for any network activity buildup between frames. Since the temporal model parameters are fit to biological limits found in the early primate visual system it is likely that higher level processing (such as that from area MT) is required in order to obtain the number-of-frames effect on D_{max}.

Magno versus parvocellular pathways

In addition to the spatial space-variant properties of the early visual pathway the temporal characteristics are also key in determining the limits of motion perception. In order to test whether eccentricity-
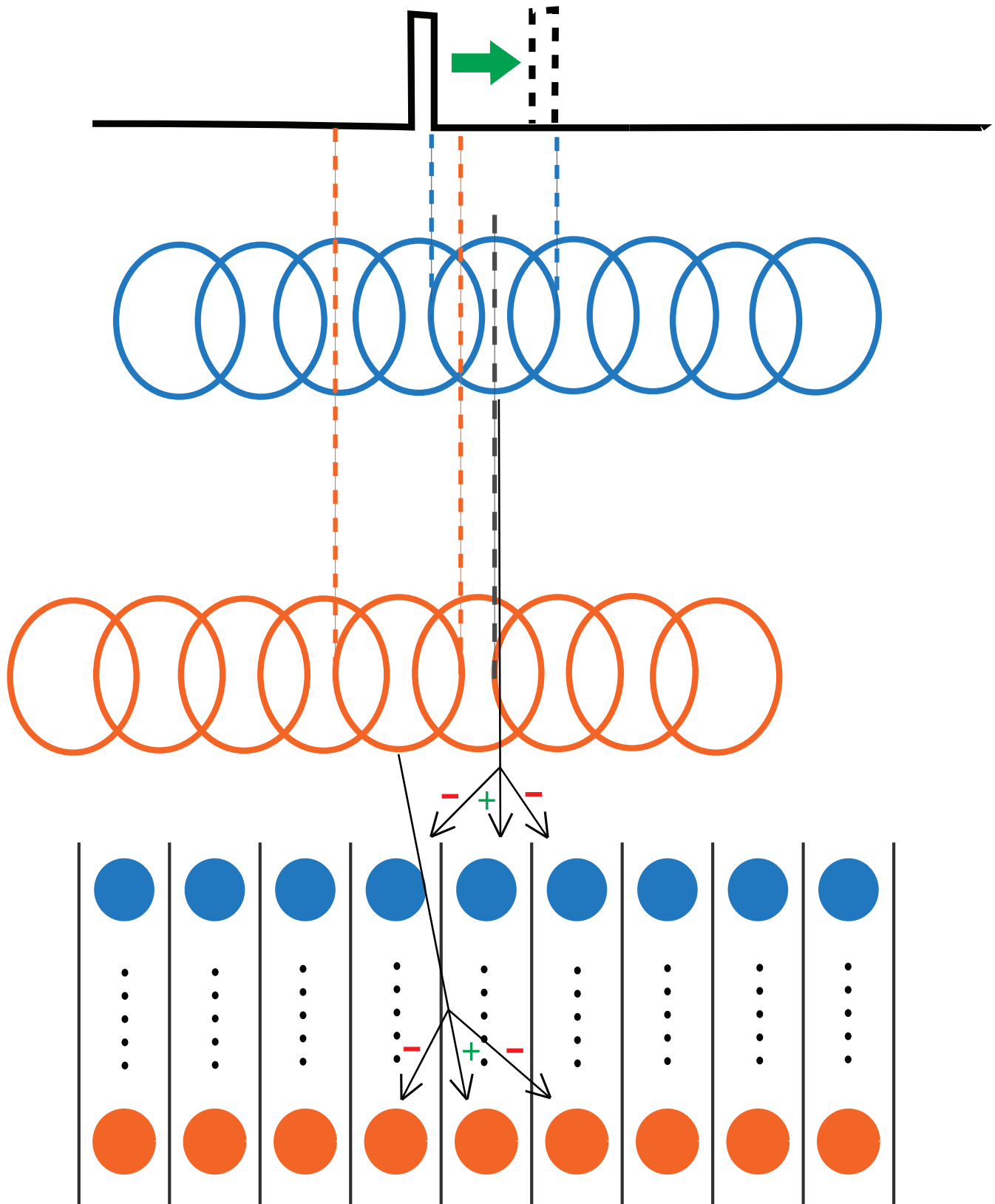
Figure 13. Model cortical scatter schematic. Model V1 receptive fields are sampled from offset retinal locations. Cortical columns are generated by combining (averaging) V1 neuron activities from neurons with RFs that are offset in retinal coordinates. All neurons within a column have a RF centroid taken in the range $c_n + \delta_j$, where c_n is the centroid location of the nth neuron and δ_j is a model parameter (Table 2).
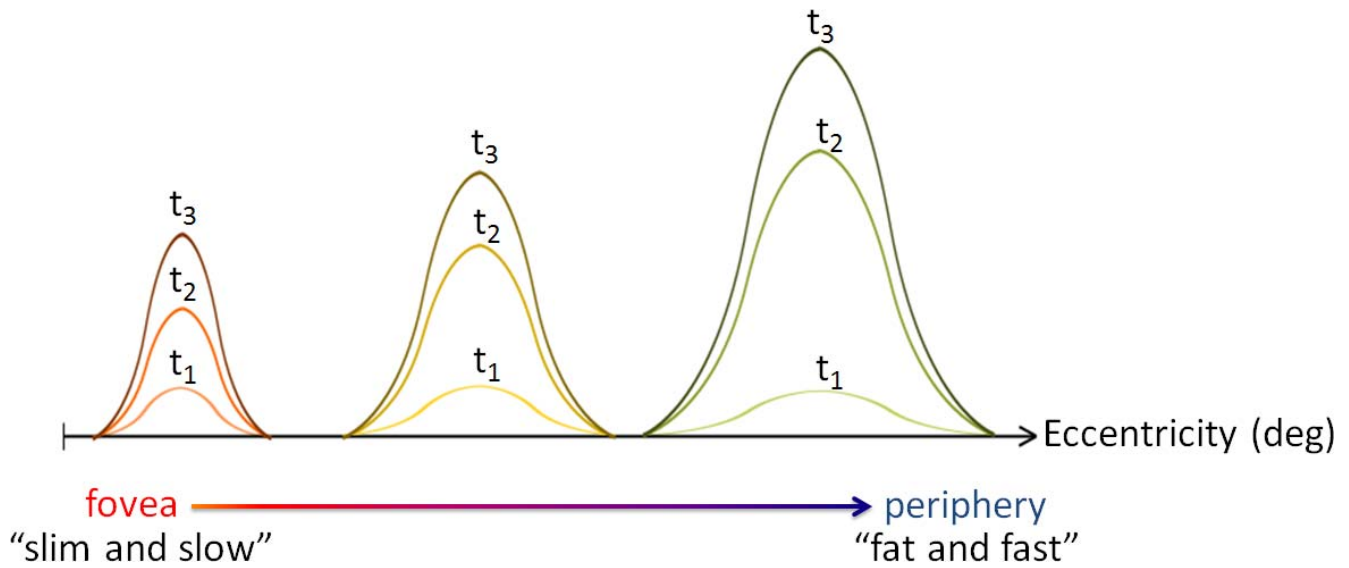
Figure 14. Schematic showing the spatiotemporal response characteristics for model cortical neurons to single, unit amplitude input stimuli at different eccentricities.

dependence of D_{max} could be related to the temporal response properties of retinal cells as a function of eccentricity, we vary the response time, τ (Figure 21).

The data show that as τ increases to higher values D_{max} increases until it reaches a plateau, after which D_{max} ceases to increase further. The concavity of the eccentricity-dependence curves for varying τ changes as
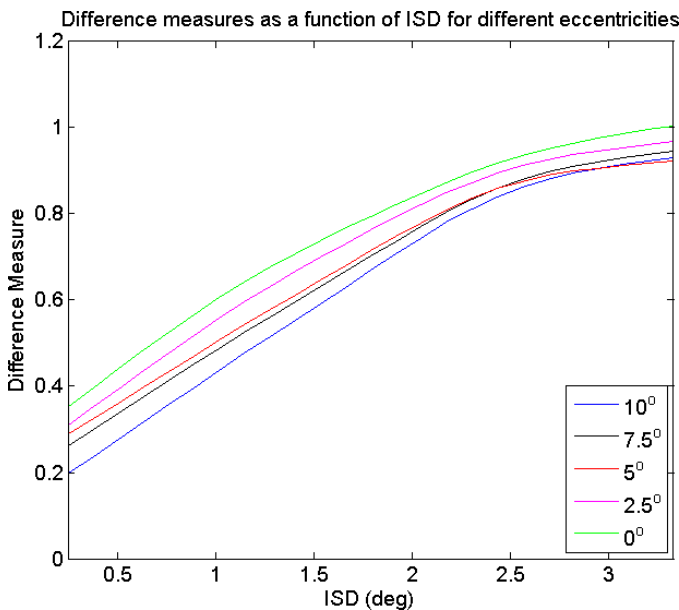

Figure 15. Schematic showing the discriminability measure (normalized Euclidean distance) over a range of interstimulus distances (ISD) for multiple eccentricities. In order to fit psychophysical measurements from Baker and Braddick (1985) an appropriate cutoff is selected to produce an appropriate range for D_{max}.

well. This result shows that the network loses the ability to distinguish high speeds of motion for large τ.

Discussion

The goal of this paper is to explore the importance of individual space-variant visual features in forming cortical motion percepts. These key features include retinal spatiotemporal sampling characteristics, cortical magnification, constant receptive field overlap, and number of input frames. By constructing and testing a model of the early human visual system that incorporates these features we are able to determine a set of minimal requirements for increasing D_{max}/D_{min} as a function of eccentricity and the potential role of higher level visual processes motion processing.

Minimum requirement for the increase in D_{max} as a function of eccentricity

Baker and Braddick (1985) showed that D_{max} and D_{min} increase linearly as a function of eccentricity. The neural substrates responsible for this increase, however, have remained elusive despite the many space-variant aspects of the visual system that are known. One major question that arises is whether the space-variancy present in motion-processing areas, such as MT, are directly responsible for increase in D_{max}, or if the perceptual phenomena can be entirely accounted for in the early visual stream.

In the model presented here we show that the combination of faster magnocellular processing, larger
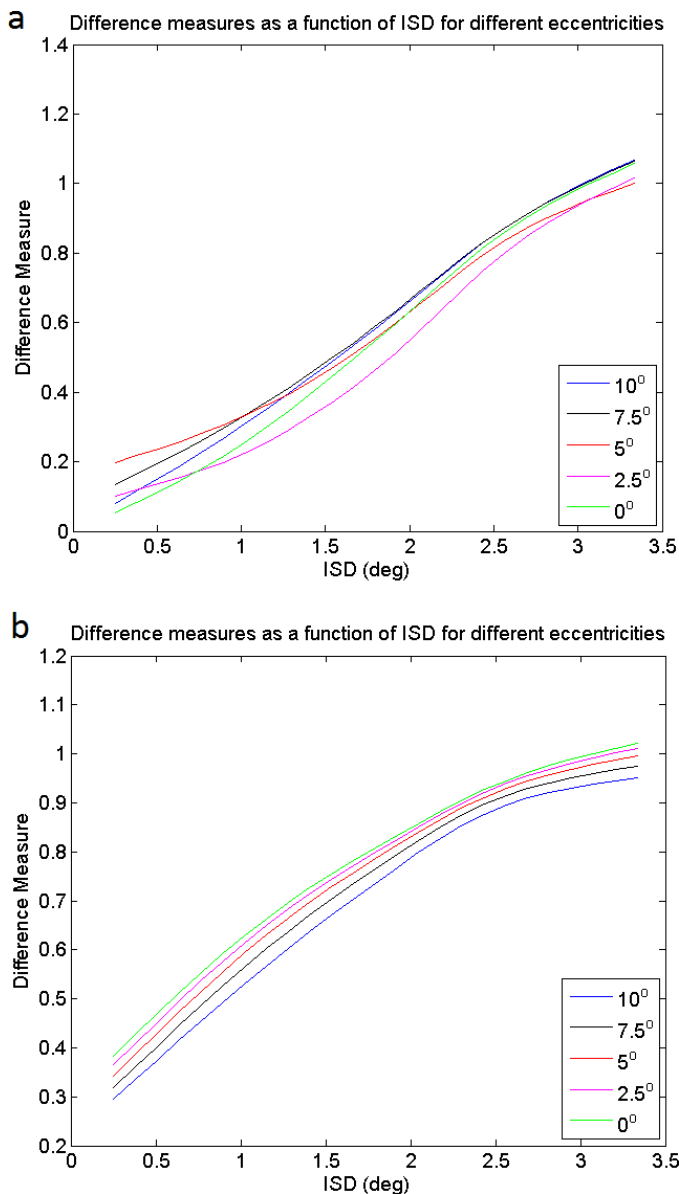
a

b

Figure 16. Model discriminability difference measures as a function of eccentricity for difference ISDs for (a) $\alpha = 0$ and (b) $\alpha = 0.6$. (a) The Euclidean distance measure varies erratically for a given ISD at multiple eccentricities in the absence of any space-variant factors. (b) When cortical magnification is present, however, a "layered" structure of the Euclidean difference measure forms across increasing eccentricity as one holds ISD constant. While the exact parameter value chosen for the discriminability index is arbitrary this Figure shows that regardless of the exact discriminability parameter value (and thus the exact values of D_{max}) the qualitative monotonicity of the eccentricity dependence will be present only in the space-variant case.

peripheral receptive field sizes, and increasing cortical magnification factor is enough to account for current D_{max} psychometric data. That is, this subset of the early visual stream is sufficient to account for the
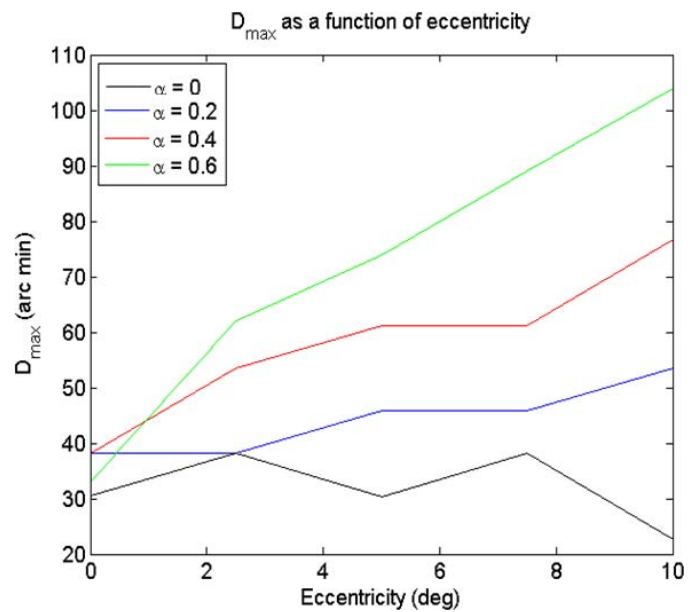

Figure 17. D_{max} as a function of eccentricity for the D_{max} threshold 0.5.

observed D_{max} increase as a function of eccentricity in its most basic qualitative form.

Does cortical receptive field scatter play a role in D_{max}?

The role of receptive field cortical scatter has been debated. Classically, this scatter is simply considered to be a source of noise to the visual system. Recent research, however, speculates that it could be used to refine visual measurements (Engel et al., 2009, Mishra & Aloimonos, 2009, Tistarelli & Sandini, 1993). Difficulty in determining the exact contribution of cortical RF scatter is due to the complex columnar connections through V1 and the limited data detailing the phenomenon. Despite these difficulties, we are able to determine an order-of-magnitude approach geared toward ascertaining whether RF scatter is important in determining D_{max}. Under the assumption that cortical receptive field scatter functions as a low pass filter on the retinal input we use a simple columnar model to show that the relative contribution towards D_{max} due to receptive field scatter is minimal. Further work would be needed to determine whether cortical scatter can serve a refinement role for visual features.

Role of higher level visual processes

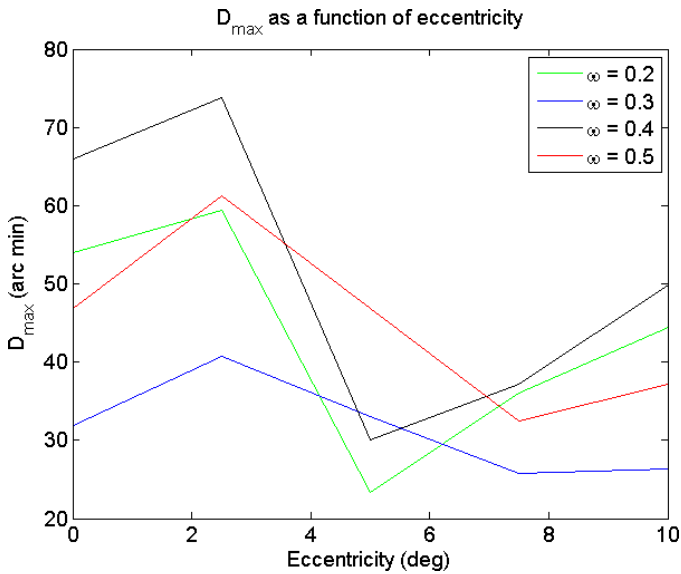One requirement found from the model in order to obtain increasing D_{max} as a function of eccentricity is

Figure 18. Displaying D_{max} as a function of eccentricity for varying overlap factor. Note that there is no obvious increase in D_{max} as a function of increased overlap factor.

the cortical magnification factor. Cortical magnification, however, does not stop at V1. It is highly possible, then, that cortical magnification to higher visual areas (V1 to V2, etc.) may play a significant role in determining the exact value of D_{max} at any given eccentricity. It is still not clear, however, what role, if any, motion-specific areas such as MT play on D_{max}. While it may not be requisite that they be involved in order for a linearly-increasing D_{max} as a function of eccentricity, the current model can only place con-
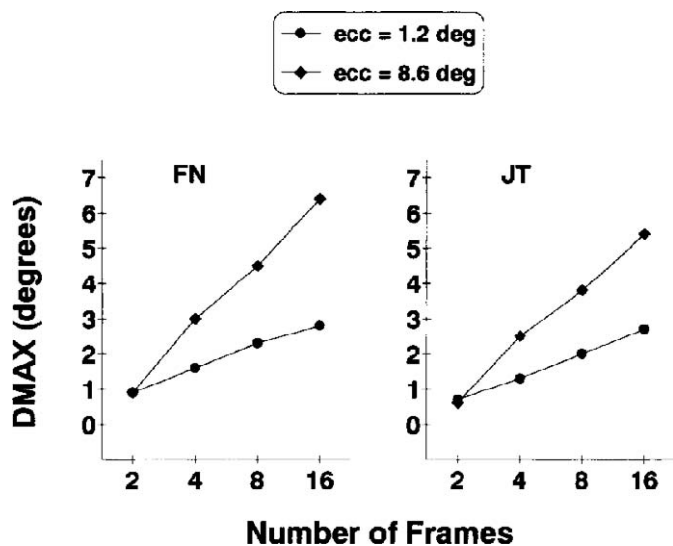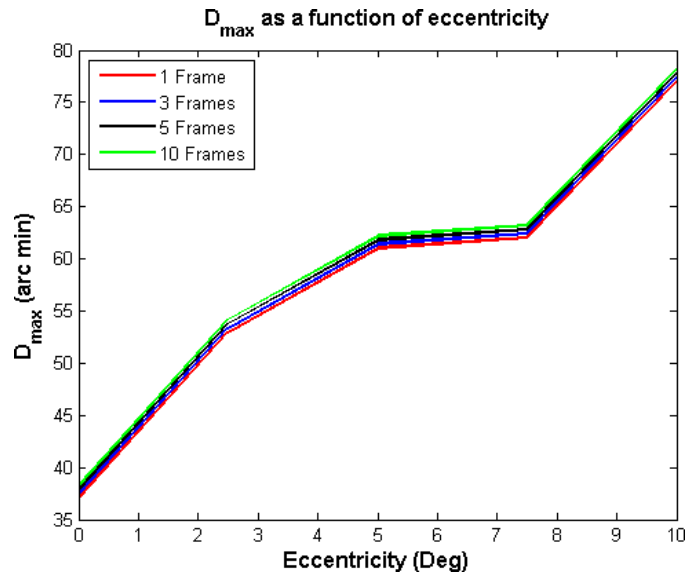

Figure 20. Model D_{max} as a function of the number of input frames. Note that there is negligible change in D_{max} as a function of the number of frames. As network activity builds up in the model increasing frame-pairs more easily excite the network above threshold, thus increasing D_{max}. This effect is offset, however, by the rapid decay of network activity. As shown above, in model simulations using biologically plausible parameters, the rapid decay of network activity was by far the more prominent effect. There was only a very slight increase in D_{max} as a function of number of frames.

straints on the sufficiency of the early visual system, not in the exclusion of any higher visual area.

Todd and Norman (1995) found that increasing the number of frames increases D_{max}. While the model presented here cannot account for this data, one
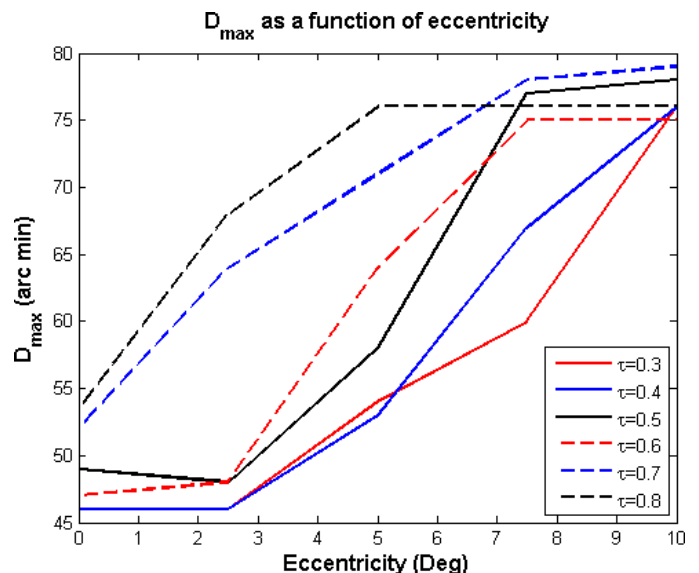

Figure 19. D_{max} as a function of the number of frames for two observers. Figure taken from Todd and Norman (1995) with permission.

Figure 21. D_{max} as a function of eccentricity while varying τ.

hypothesis is that extrastriate areas may be requisite to explain the increase of D_{max} as the number of presented frames increases. Our model, bound by parameter limits found in the early visual system, shows a negligible increase in D_{max} as a function of the number of frames. That is, D_{max} stays virtually the same even as more frames are shown to the model. One possibility we propose that would account for the model results is that higher level visual areas may be able to pool visual input across larger time scales than can be pooled in the early visual system. Larger integration times can arise due to additional low pass filtering in the temporal domain due to additional synaptic delays in interareal connections or from larger temporal constants present in extrastriate cellular responses.

Conclusion

D_{max} and D_{min} are the maximum and minimum displacements, respectively, to which apparent motion can be observed for a particular stimulus. Thus it follows that while D_{max} might be the result of long-range motion processing D_{min} is much more apt to be an acuity measurement. Baker and Braddick (1985) posited that D_{min} could be the result of the early visual pathway while D_{max} could be the result of higher level visual processing. The picture drawn here, however, would seem to indicate a more complex arrangement. We posit that the basic D_{max} eccentricity dependence may well be found entirely within the early visual system, but that the exact discriminability may be modulated by higher visual areas operating on additional features (e.g., motion).

Keywords: cortical magnification factor, space-variant vision, apparent motion, receptive field overlap, receptive field scatter, neural model, D_{max}, D_{min}, primary visual area, extrastriate visual area, Korte's laws

Acknowledgments

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America, 2*(2), 284–299.

Baker, C., & Braddick, O. (1985). Eccentricity-dependent scaling of the limits for short-range apparent motion perception. *Vision Research, 25*(6), 803–812.

Bonmassar, G., & Schwartz, E. L. (1997). Space-variant fourier analysis: The exponential chirp transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*(10), 1080–1089.

Bolduc, M., & Levine, M. (1998). A review of biologically motivated space-variant data reduction models for robotic vision. *Computer Vision and Image Understanding, 69*(2), 170–184.

Braccini, C., Gambardella, G., Sandini, G., & Tagliasco, V. (1982). A model of the early stages of the human visual system: Functional and topological transformations performed in the peripheral visual field. *Biological Cybernetics, 44,* 47–58.

Carrasco, M., McElree, B., Denisova, K., & Giordano, A. (2003). Speed of visual processing increases with eccentricity. *Nature Neuroscience, 6,* 699–700.

Daniel, P. M., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *Journal of Physiology, 159,* 203–221.

Dow, B., Snyder, A., Vautin, R., & Bauer, R. (1981). Magnification factor and receptive field size in foveal striate cortex of the monkey. *Experimental Brain Research, 44,* 213–228.

Eagle, R. (1996). What determines the maximum displacement limit for spatially broadband kinematograms? *Journal of the Optical Society of America, 13*(3), 408–418.

Engel, G., Greve, D., Lubin, J., & Schwartz, E. (1993). Space-variant active vision and visually guided robotics: Design and construction of a high-performance miniature vehicle. *Proceedings of the 12th IAPR International Conference on Pattern Recognitio,* Vol. 2 –Conference B, pp. 487–490. Jerusalem, Israel: Computer Vision & Image Processing.

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience, 14,* 1195–1201. doi:10.1038/nn.2889.

Glennerster, A. (1998). d$_{max}$ for stereopsis and motion in random dot displays. *Vision Research, 386,* 925–935.

Grossberg, S., & Rudd, M. (1989). A neural architec-

ture for visual motion perception: Group and element apparent motion. *Neural Networks, 2,* 421–450.

Hartmann, E., Lachenmayr, B., & Brettel, H. (1979). The peripheral critical flicker frequency. *Vision Research, 10,* 1019–1023.

Harvey, B., & Dumoulin, S. (2011). The relationship between cortical magnification factor and population receptive field size in human visual cortex: Constancies in cortical architecture. *Journal of Neuroscience, 31,* 13604–13612.

Hubel, D., & Wiesel, T. (1974). Uniformity of monkey striate cortex: A parallel relationship between field size, scatter, and magnification factor. *Journal of Comparative Neurology, 158,* 295–306.

Kronauer, R. E., & Zeevi, Y. Y. (1985). Reorganization and diversification of signals in vision. *IEEE Transactions on Systems, Man, and Cybernetics, 15,* 97.

Lee, B., Pokorny, J., Smith, V., & Kremers, J. (1994). Responses to pulses and sinusoids in macaque ganglion cells. *Vision Research, 34,* 3081–3096.

Licktenstein, M. (1963). Spatio-temporal factors in the cessation of smooth and apparent motion. *Journal of the Optical Society of America, 53,* 302–306.

McIlwain, J. (1986). Point images in the visual system—new interest in an old idea. *Trends in Neuroscience, 9,* 354–358.

Mishra, A. K., & Aloimonos, Y. (2009). Active segmentation. *International Journal of Humanoid Robotics, 6,* 361–386.

Morgan, M. (1992). Spatial filtering precedes motion detection. *Letters to Nature, 355,* 344–346.

Nakayama, K. (1985). Biological image motion processing: A review. *Vision Research, 255,* 625–660.

Ogawa, T., Bishop, P., & Levick, W. (1966). Temporal characteristics of responses to photic stimulation by single ganglion cells in the unopened eye of the cat. *Journal of Physiology, 29,* 1–30.

Palmer, C., Chen, Y., & Seidemann, E. (2012). Uniform spatial spread of population activity in primate parafoveal V1. *Journal of Neurophysiology, 107,* 1857–1867.

Perry, V., Oehler, R., & Cowey, A. (1984). Retinal ganglion cells that project to the dorsal lateral geniculate nucleus in the macaque monkey. *Neuroscience, 12*(4), 1101–1123.

Rosa, M., Fritsches, K., & Elston, G. (1997). The second visual area in the marmoset monkey: Visuotopic organisation, magnification factors, architectonical boundaries, and modularity. *Journal of Comparative Neurology, 387,* 547–567.

Stone, J. (1965). A quantitative analysis of the distribution of ganglion cells in the cat's retina. *Journal of Comparative Neurology, 124,* 337–352.

Thompson, P., & Hammett, S. (2004). Perceived speed in peripheral vision: It can go up as well as down. *Journal of Vision, 4*(8):83, http://www.journalofvision.org/content/4/8/83, doi:10.1167/4.8.83. [Abstract]

Tistarelli, M., & Sandini, G. (1993). On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15,* 401–410.

Todd, J., & Norman, J. F. (1995). The effects of spatiotemporal integration on maximum displacement thresholds for the detection of coherent motion. *Vision Research, 35*(16): 2287–2302.

Tootell, R. B., Silverman, M. S., Switkes, E., & De Valois, R. L. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science, 218,* 902–904, doi:10.1126/science.7134981.

Tripathy, S., Shafiullah, S., & Cox, M. (2012). Influence of correspondence noise and spatial scaling on the upper limit for spatial displacement in fully-coherent random-dot kinematogram stimuli. *PLOS One, 7*(10), e42995.

Van Essen, D., Newsome, W., & Maunsell, J. (1984). The visual field representation in striate cortex of the macaque monkey: asymmetries, anisotropies, and individual variability. *Vision Research, 24,* 429–448.

Wässle, H., Grünert, U., Röhrenbeck, J., & Boycott, B. (1989). Cortical magnification factor and the ganglion cell density of the primate retina. *Nature, 341,* 643–646.

Watt, J., & Morgan, M. (1985). A theory of the primitive spatial code in human vision. *Vision Research, 25,* 1661–1678.

Westheimer, G. (1983). Temporal order detection for foveal and peripheral visual stimuli. *Vision Research, 23,* 759–764.

Yamamoto, H., Yeshurun, Y., & Levine, M. D. (1996). An active foveated vision system: Attentional mechanisms and scan path convergence measures. *Computer Vision and Image Understanding, 63,* 50–65.