2021

Mechanisms of Sensory Adaptation in the Primate Visual System

Boris Isaac Peñaloza Rojas

Mechanisms of Sensory Adaptation in the Primate Visual System

_____

A Dissertation

Presented to

the Faculty of the Daniel Felix Ritchie School of Engineering and Computer Science

University of Denver

_____

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

_____

by

Boris Isaac Peñaloza Rojas

August 2021

Advisor: Haluk Öğmen, Ph.D.

Author: Boris Isaac Peñaloza Rojas
Title: Mechanisms of Sensory Adaptation in the Primate Visual System
Advisor: Haluk Öğmen, Ph.D.
Degree Date: August 2021

Abstract

Under ecological conditions, the luminance impinging on the retina varies within a dynamic range of 220 dB. Stimulus contrast can also vary drastically within a scene, and eye movements leave little time for sampling luminance. In addition, the amount of information reaching our visual system far exceeds the brain's information processing capacity. Given the limited dynamic range of its neurons and its limited capacity in processing visual information in real-time, the brain deploys both structural and functional solutions that work in tandem to adapt to the surroundings. In this work, employing visual psychophysics and computational neuroscience, we study the mechanisms by which the brain adapts to the sensory signals that it encounters in the natural environment. We found that the processes underlying motion perception in ecological vision are mediated by an adaptive center-surround mechanism that trade-offs spatiotemporal resolution for signal enhancement when the signal is weak. We proposed a new dynamic neural network that can account for adaptive properties of motion integration and segregation under various luminance and contrast conditions. Finally, in order to clarify the implications of attentional mechanisms deployed to select inputs according to the brain's processing resources, we tested the predictions of a neural model and showed that competitive interactions between parvocellular and magnocellular systems can explain the differential effects of attention on spatial and temporal acuity. This dissertation's results contribute to

the important work of reverse-engineering the human brain, with potential clinical and engineering applications.

Acknowledgments

I am thankful to many individuals, institutions, and circumstances. However, I start by expressing my eternal gratitude to Dr. Haluk Öğmen for his encouragement, support, guidance, and constructive criticism throughout my process. I feel very grateful and privileged to have had such a knowledgeable and humble advisor. It is a role model that I will follow during my scientific career and life.

I would also like to extend my gratitude to all the professors who gave me a word of encouragement and advice in my doctorate studies. Special thanks to Dr. Michael Herzog and my committee members, Dr. Timothy Sweeny, Dr. Mohammad Mahoor, and Dr. Kimon Valavanis. I would also like to acknowledge the Panamanian government for its initial financial support. My sincere gratitude also goes to the Universidad Tecnológica de Panamá, the University of Houston, the University of Denver, and to all the professors and teachers I had during my education.

Finally, I would like to thank my wonderful wife Nelva and my son Mateo for their unconditional love and support. Special thanks to my mother, father, sisters, mother-in-law, father-in-law, and extended family and friends. During these years, they all contributed to my success and helped me grow and mature.

This has been a wonderful journey that I will never forget.

To Mateo and Nelva with love

Table of Contents

List of Figures

Chapter 1. Introduction

We live in a highly dynamic environment. Daily the human brain needs to adapt to the different challenges posed by our surroundings. For instance, when arriving with a friend at a loud soccer game, our auditory system needs to recognize our friend's voice even though the background noise was not as strong as inside the stadium just a few moments ago. Another instance would be spending two hours watching a film in a dark movie theater and then walking out of the theater. Here, the visual system must be able to adapt to the new background luminance quickly. Furthermore, regardless of background luminance, the visual system must flexibly adapt to variations in luminance and contrast changes within the visual field due to shading, occlusions, camouflage, type of light sources, and many others. Additionally, on a daily basis, the brain is bombarded with an enormous amount of visual information. However, the primate brain has limited resources to process these data. Thus, it needs a selective process to help it optimize its limited processing capacities. In other words, our survival and successful interaction with the environment depends critically on the brain's ability to process, select, and extract sensory information under a wide range of ecological conditions. Therefore, the general purpose of this project is to investigate and model the potential mechanisms by which the brain adapts to the sensory signals that it encounters in the natural environment.

In this work, we focus on the visual system. We selected the visual system because it is the most important sensory modality for humans and the most well-known part of the

brain (Gallace & Spence, 2009; Hutmacher, 2019). In particular, this dissertation concentrates mainly on the parts of the visual system specialized in processing motion information. The processing of motion signals, coming from our movements or the motion of objects in the world, is a fundamental goal of the visual system. Thus, our understanding of the mechanisms underlying motion processing under different environmental conditions is of great importance.

To study these brain mechanisms, we used two techniques, namely, visual psychophysics and computational neuroscience. On the one hand, visual psychophysics refers to the quantitative study of the relationship between physical (visual) stimuli and the sensations and perceptions they generate. For example, in a typical visual psychophysical experiment to determine the threshold for contrast detection, the experimenter will first vary the stimulus contrast presented to the observer. Then, using a psychophysical measuring method, the contrast threshold will be estimated. Thus, a link between physical stimulus (contrast) and perception (contrast threshold) is established. On the other hand, computational neuroscience is the branch of neuroscience that employs (typically mathematical) abstractions of the brain to understand the principles that govern its functions.

The body of this dissertation is composed of three projects that address the following question: how does the visual system process and extract a wide range of ecological sensory information, given its intrinsic limitations? Here we consider two intrinsic limitations: The limited *dynamic range* of its constituent elements (neurons) and the limited *capacity* of its processing systems. The first two projects address the dynamic-

range problem whereas the last project addresses the effects of attentional selection in order to deal with capacity limits of sensory processing and memory storage.

One possible strategy used by the visual system to tackle the leading question of this work is an adaptive center-surround mechanism. For instance, in a psychophysical study, Tadin et al. (2003) found that motion-direction discrimination in humans increases with increasing stimulus size for low-contrast stimuli. Surprisingly, however, the ability to discriminate motion direction decreased at high-contrast conditions.

The authors proposed that this is a perceptual correlate of a center-surround antagonism, whereby a *spatial suppressive mechanism* processes highly visible objects and weak low-contrast sensory inputs are processed by a *spatial summation mechanism*. This motion-discrimination experiment was conducted using retinotopic coordinates. Retinotopy refers to the spatial organization principle by which neighboring points in the visual field project to neighboring points in the retina and, from there, they project to higher visual areas in the brain (Engel et al., 1997; Tootell et al., 1997). Under ecological viewing conditions, however, motion perception occurs predominantly in non-retinotopic coordinates (i.e., non-retinotopic mechanisms are required to achieve a stable perception of the world because the retinal images constantly change as a product of the motion of our eyes, head, body, or external objects) (Johansson, 1973, 1974; Öğmen, 2007; Ternus, 1926). Therefore, in project 1 (Chapter 2), we conducted visual psychophysical experiments to investigate whether or not a similar adaptive center-surround antagonism, found in retinotopic motion-perception (Tadin, 2015; Tadin et al., 2003; Tadin et al., 2011), also modulates the mechanisms of non-retinotopic motion perception.

In project two (Chapter 3), we approached the dynamic-range problem of sensory systems in the brain from a computational perspective. Here we propose a new dynamic and biologically plausible neural model built upon well-established canonical neural mechanisms. We show how the properties of this model, which emerge from its nonlinear dynamics, can account for adaptive properties of motion integration and segregation under various luminance and contrast conditions. We demonstrate the explanatory scope of our model by fitting its results to several selected psychophysical and neurophysiological data.

In order to solve the dynamic range problem and overcome its limitations in processing capacity, the brain allocates a significant amount of resources and deploys multiple solutions that work in synergy. One group of solutions can be called *structural* solutions. For instance, two structurally different photoreceptor types (rods and cones) are used to sense light at low and high background levels. A second group of solutions can be called *functional* solutions, viz., solutions based on *emergent* properties of nonlinear neural networks, such as background adaptation, contrast normalization stemming from spatial/temporal integration, and attentional modulation. Each solution, nevertheless, comes with its trade-offs. For instance, when the environmental conditions require spatial/temporal integration to improve the signal-to-noise ratio, this comes at the cost of deteriorating the spatial/temporal resolution. One solution is a parallel structure, viz., two distinct parallel pathways with complementary spatiotemporal sensitivities (parvo- and magnocellular streams) that can provide a joint optimization in terms of spatiotemporal sensitivity and resolution. *Attention*—the selective process that helps the brain filter the critical information and optimize its limited information-processing resources—can modulate spatial and temporal sensitivities to enhance signals of interest according to the

subject's goals and task's demands. Attention has been found to positively affect spatial resolution while negatively affecting temporal resolution (Yeshurun & Carrasco, 1999; Yeshurun & Levy, 2003). One possible explanation for these attentional effects on spatiotemporal acuity is an attentional mechanism based on interactions between the parvocellular and magnocellular systems (Ogmen, 2005; Yeshurun, 2004; Yeshurun & Levy, 2003; Yeshurun & Sabo, 2012), providing, therefore, a synergy between structural and functional solutions to help the visual system optimize its limited processing resources. Therefore, in our third project (Chapter 4), we tested the predictions of a computational model of retino-cortical dynamics (RECOD model) that incorporates interactions between Parvo and Magno systems (Öğmen, 1993; Ogmen et al., 2003). By validating our model's predictions with behavioral data, we showed how our model explains the modulatory effects of spatial attention in the processing of the spatial and temporal characteristics of the visual world.

In summary, in this work, we found that an adaptive center-surround mechanism also mediates the non-retinotopic processes underlying motion perception in ecological vision. The flexible spatial integration mechanisms described in this work might be how vision deals with the wide variability of motion signals found in the natural environment. We also showed how a biologically plausible neural model could account for adaptive properties of motion integration and segregation under various luminance and contrast conditions. Finally, we tested and analyzed the predictions of a neural model that explains the role of spatial attention in the solution to the processing capacity limitations of biological visual systems. Taken together, the results of this dissertation contribute to the efforts of reverse-

engineering the human brain. In particular, we addressed the problem of how the brain adapts to the wide range of signals coming from the environment.

Chapter 2. Non-retinotopic adaptive center-surround modulation in motion processing[1]

2.1. Introduction

The early visual system is organized retinotopically i.e., neighboring points in the visual field are mapped onto neighboring photoreceptors in the retina. This *retinotopic* encoding principle is maintained in the early visual areas. However, motion perception occurs usually in *non-retinotopic coordinates*. For example, consider the perceived trajectory of a reflector located on the wheel of a moving bicycle. With eyes fixed, the reflector's motion on the retina is a curtate cycloid (Fig. 2-1A). However, perceptually, the reflector appears to move on a circular orbit (Fig. 2-1B) because the horizontal motion of the bicycle is subtracted from the curtate motion. The curtate cycloid is invisible (but can still influence decisions, Lauffs et al., 2018). Likewise, Öğmen et. al. (2006), Otto et. al. (2006) and Boi et. al. (2009) showed that perception of *form* and *motion* is influenced not only by retinotopic stimulation but also by the establishment of perceptual groups. Perceptual groups as well as perceptually organized form (Tadin et al., 2002) establish reference-frames in which motion is computed. Even though the importance of non-retinotopic

[1] Presented at the VSS 2019 conference (Peñaloza et al., 2019), published in peer-reviewed journal (Peñaloza et al., 2020).

processes in motion perception has been established (e.g., Johansson, 1973, 1974; Pikler, 1917; Ternus, 1926), very little is known about the neural mechanisms that mediate them.

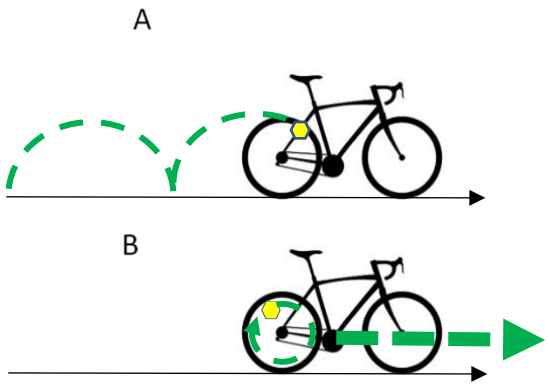

Fig. 2-1. When the eyes are fixed, the yellow reflector on the rear wheel of the bicycle traverses a curtate cycloid trajectory as shown in (A). Hence, the perceived trajectory of the reflector should be a curtate cycloid if perception were to follow retinotopic coordinates. However, as shown in (B), the perceived trajectory of the reflector is circular. Thus, we perceive the reflector relative to the horizontal motion of the bicycle, i.e., according to a non-retinotopic reference-frame set by the horizontal motion of the bicycle.

Using a drifting Gabor, Tadin and colleagues (Tadin et al., 2003) demonstrated that increasing the spatial size of a high-contrast stimulus renders the discrimination of motion-direction more difficult. The opposite effect occurs with a low-contrast stimulus. The authors proposed that this is a perceptual correlate of a center-surround antagonism, whereby highly visible objects are processed by a *spatial suppressive mechanism* and weak low-contrast sensory inputs are processed by a *spatial summation mechanism*. In other words, the interaction between size and contrast may be explained by an *adaptive center-surround mechanism*. Tadin et. al.'s (2003) experimental design did not distinguish between retinotopic and non-retinotopic processes. The goal of this study was to use the Ternus-Pikler display to test the hypothesis that similar adaptive center-surround

mechanisms operate in non-retinotopic coordinates. This generalization is of interest because, as mentioned above, motion perception in natural environments is essentially non-retinotopic.

The Ternus-Pikler display (Pikler, 1917; Ternus, 1926)[2] is an experimental paradigm that offers the possibility to pit against each other retinotopic and non-retinotopic reference frames. A conventional Ternus-Pikler display consists of three elements that are spatially shifted by one inter-element distance from frame to frame (Fig. 2-2). Depending on the Inter-Stimulus Interval (ISI) between frames, two different percepts occur (Pantle & Picciano, 1976): Element motion for short ISIs (0 ms in this experiment) or group motion for long ISIs (266 ms in this experiment).

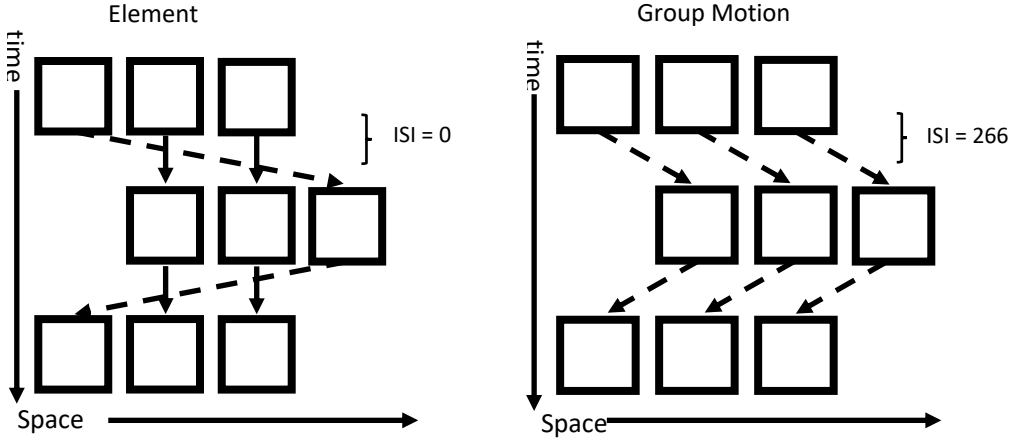

Fig. 2-2. The Ternus-Pikler display. The conventional Ternus-Pikler display consists of three elements that are spatially shifted by one inter-element distance from frame to frame. Depending on the time interval between frames (ISIs), two different percepts are obtained: Element motion (left panel) and Group motion (right panel). In the **Element Motion (0ms ISI):** the two central elements are perceived as stationary (indicated by the solid lines) while the outermost squares are perceived to alternate from left to right (dashed line). **Group Motion (266ms**): all three squares appear to move in tandem as a group from left to right (dashed lines).

[2] There are extensive studies examining how various stimulus parameters affect the perceived motion in Ternus-Pikler displays (e.g., Pantle & Picciano, 1976; Pantle & Petersik, 1980; Breitmeyer & Ritter, 1986; Dawson & Wright, 1994; He & Ooi, 1999; Kramer & Rudd, 1999; Scott-Samuel & Hess, 2001; Alais & Lorenceau, 2002; Hein & Moore, 2012; review: Petersik & Rice, 2006).

When *element motion* prevails, the two central elements are perceived as stationary (indicated by the solid lines in Fig. 2-2), while the outermost elements are perceived to move back and forth from left to right (dashed line in Fig. 2-2). In the *group motion* percept, the elements appear to move in tandem from left to right in a trajectory indicated by the dashed lines in Fig. 2-2.

To test the hypothesis of this study, we used a modified version of the Ternus-Pikler display. We embedded Gabor patches inside the squares that make up the Ternus-Pikler display and varied the spatial size and the contrast of the Gabor patches. We measured phase-shift thresholds for discrimination of motion-direction in retinotopic and non-retinotopic tasks.

2.2. General Methods

2.2.1. Participants

Six observers (6 males) including one of the authors participated in this study. The same observers participated in all experiments. Participant's age ranged from 23 to 35 years and all participants had normal or corrected-to-normal vision. Experiments followed a protocol approved by the University of Denver Institutional Review Board for the Protection of Human Subjects. Each observer gave written consent before the experiments.

2.2.2. Apparatus

Observers viewed the stimulus on a NANAO F2-21 monitor driven by a ViSaGe card (Cambridge Research Systems). The stimulus was displayed at a resolution of 1280 x 960 with a refresh rate of 60 Hz. The background luminance was set to 30.5 cd/m^2 as measured with a Minolta (LS-110) luminance meter. The room was dimly illuminated. A head/chin

rest was used. Observers reported their responses by pressing the arrow keys of a keyboard. The viewing distance was set to 1.07 m.

2.2.3. Stimuli

Our version of the Ternus-Pikler display consisted of three drifting Gabor patches embedded in black square frames. The square width and length was set to 4.5 degrees of visual angle (deg). The square lines had a thickness of 0.15 deg. The horizontal center-to-center separation of the Ternus-Pikler squares was 5 deg. The spatial frequency of the Gabor elements was fixed to 1 cpd. In all experimental conditions, the size of the Gabor patch was 2σ, where σ is the standard deviation of the stationary Gaussian envelope that windowed a vertical sine grating. The contrast varied from 2.8% to 90% Michelson contrast.

Motion correspondences in the Ternus-Pikler display establish a reference frame underlying computations of form, motion, etc. (Öğmen & Herzog, 2010). In the element motion condition (Fig. 2-2 left), the central elements are perceived at their retinotopic positions (depicted by solid arrows) and hence any stimulus positioned in these elements is processed following a *retinotopic* reference frame. In group motion, all three elements are perceived to move left-and-right in tandem, thereby establishing a *non-retinotopic* reference frame depicted by dashed arrows in Fig. 2-2 right. In sum, the Ternus-Pikler display generates retinotopic (element motion, solid arrows in Fig. 2-2 left) and non-retinotopic (group motion, dashed arrows in Fig. 2-2 right) reference frames. We embedded Gabor patches in the Ternus-Pikler display elements (Fig. 2-3) and the observers' task was to report the drift direction (up or down) of Gabor patches. The drift of the Gabor was

implemented by adding a coherent spatial phase-shift from frame to frame (integer multiples of δ; see Fig. 2-3).

This phase-shift can be added according to retinotopic or non-retinotopic coordinates. Adding a spatial phase-shift according to retinotopic coordinates means that the phase-shift is added to a Gabor that remained in the same spatial (retinotopic) position from frame-to-frame (Fig. 2-3A and 2-3B). As shown by solid arrows in Fig. 2-2A, there are two such possible cases. Here we used the case that starts with the *central* element of the first frame (see the yellow arrows in Fig. 2-3A). A spatial phase-shift according to non-retinotopic coordinates means that the phase-shift was added from frame-to-frame to Gabors at different spatial (retinotopic) positions following the non-retinotopic reference frame shown by dashed arrows in Fig. 2-2B. Among the three possible choices, we used the case that starts with the *central* element of the first frame (see the yellow arrows in Fig. 2-3D).

Hence, we obtained a 2x2 design, with reference frame as one factor with two levels (retinotopic, non-retinotopic) and Gabor drift as the other factor with two levels (retinotopic, non-retinotopic) as shown in Figure 3. The arrows in the display indicate the *perceived* motion correspondences of the Ternus-Pikler elements (the squares) from frame to frame. The rest of the pathways in the displays were filled with mainly ambiguous motion, i.e. whenever possible we added 180^{o} phase-shifts from frame to frame (shown in red). The purpose of this was to eliminate the possibility of participants picking up motion signals from pathways other than the one containing the coherent drift information. This was empirically verified through a control experiment.

2.2.4. General Procedures

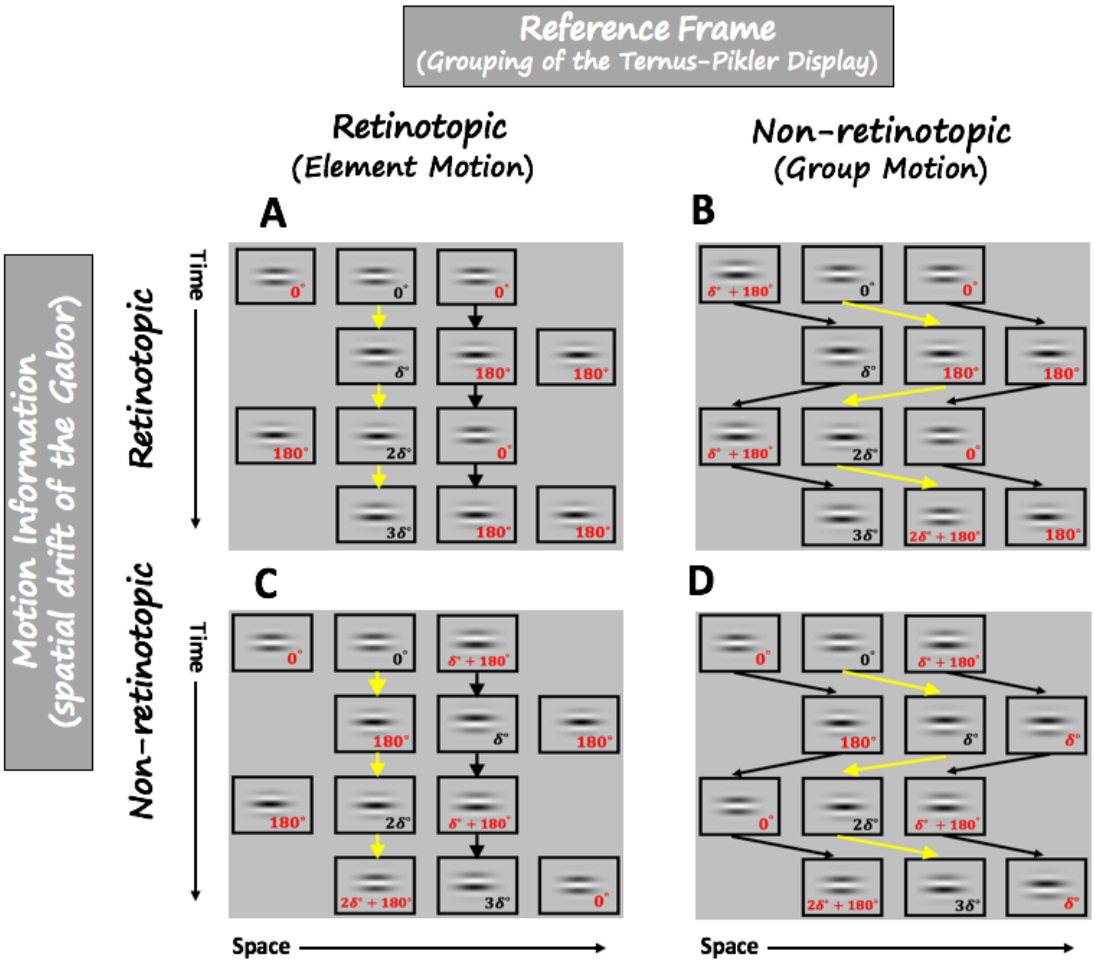

Fig. 2-3. Schematic of the 2x2 factorial design. The left column (panels A and C) shows the Element-motion condition (i.e., retinotopic reference-frame, ISI = 0 ms) with retinotopic (panel A) and non-retinotopic (panel C) drift-information, respectively. In these two conditions, the central elements of the display appear to be stationary indicated by the yellow and black arrows. Note that in the retinotopic reference-frame and retinotopic drift condition (panel A), coherent phase-drift (represented by integer multiples of δ) is placed in the same retinotopic location whereas for the retinotopic reference-frame and non-retinotopic drift condition (C), coherent phase-shift is in non-retinotopic coordinates (the integer multiples of δ shown in black bold font). The right column (panels B and D) represents the Group-motion condition (i.e., non-retinotopic reference-frame, ISI = 266.67 ms) with retinotopic (panel B) and non-retinotopic (panel D) drift. In these two conditions, the Ternus-Pikler elements(rectangular frames) appear to move in tandem in a trajectory indicated by the black and yellow arrows. As before, coherent phase-shift δ is placed in either retinotopic (B) or non-retinotopic coordinates (D).

At the beginning of each session, participants were adapted for 4 minutes to the background luminance of the display. Each trial started with a fixation cross in the middle

of the screen presented for 1000-ms. After that, the stimulus sequence was presented. The stimulus sequence consisted of 4 Ternus-Pikler frames (Fig. 2-3) each displayed for 133.33-ms. Between these frames, a blank screen was shown for an Inter Stimulus Interval (ISI) duration of either 266.67 (group motion) or 0 ms (element motion). After stimulus presentation, a blank screen appeared until observer's response. Throughout the experiment observers were instructed to attend to one of the Gabor elements of the display (indicated by the yellow arrows in Fig. 2-3) while maintaining fixation on the center of the screen. The *task* was to report the perceived direction of drift (upward/downward) of the attended Gabor element as accurately as possible by pressing the up or down arrows of the keyboard. The next trial started 1000 ms after the observer gave a response.

We measured the *threshold phase-shift* δ, required for the observers to accurately identify the motion direction of the drifting Gabor stimulus. The threshold was set to 80% correct performance. We used the Quest+ adaptive method (Watson, 2017) to obtain the thresholds. Data was fitted with a Weibull function. The lapse level was fixed to 0.01. In some experimental conditions, observers were at chance and never reached 80% accuracy. In those cases, we report % correct performance. As the focus of our study was primarily on measuring performance *changes* as a function of contrast and size rather than identifying optimal performance, we did not provide feedback to the participants.

One important observation is the potential interference of eye movements with non-retinotopic processing. For example, if subjects follow the non-retinotopic stimulus with eye movements by foveating it, the non-retinotopic stimuli will be projected at a retinotopically fixed position (fovea) even though it occupies different spatial positions in

the external world. In several previous studies (Boi et al., 2011; Boi et al., 2009; Lauffs et al., 2018; Lauffs et al., 2019; Lauffs et al., 2017; Thunell et al., 2016) we addressed this point in detail by measuring directly eye movements during experiments using the Ternus-Pikler display. The results showed that observers can and indeed do maintain a steady fixation with negligible eye movements. Thus, we have strong experimental evidence that eye movements do not play a role in perceptual effects associated with the Ternus-Pikler displays.

An experimental block consisted of 51 trials. The direction of drift was randomized across trials. During a given session, the order of experimental condition and the combination of Gabor patch-width and contrast level were randomized. Each observer ran a total of 165 blocks. Typically, a block was completed in about 1.5 minutes, which resulted in approximately 6 sessions of 40 minutes each.

2.3. Experiments, Their Rationales and Predictions

As discussed above and as depicted in Fig. 2-3, depending on the match between the reference frame (retinotopic/element-motion or non-retinotopic/group-motion) and the placement of the coherent drift information δ (placed in either retinotopic or non-retinotopic coordinates), four main experimental conditions are generated.

Condition 1: Retinotopic reference-frame and retinotopic drift-information (Fig. 2-3A):
In this condition, the element-motion percept establishes a retinotopic reference-frame and the coherent drift-information (represented by integer multiples of δ) is added in retinotopic coordinates. Thus, there is a *match between the reference-frame and the drift information*, both in retinotopic coordinates similar to the design of Tadin et al. (2003). Therefore, for

high-contrast Gabors we expect a decrease in performance as a function of size whereas for a low-contrast stimulus we expect an improvement in performance.

Condition 2: Non-retinotopic reference-frame and retinotopic drift-information (Fig. 2-3B): Here, group-motion establishes a non-retinotopic reference-frame (depicted by the black and yellow arrows in Fig. 2-3B), whereas the drift direction is perceived retinotopically. Hence, in this case, there is a *mismatch between the reference frame and the drift information*. Because of this mismatch, we predict that observers will not reach the threshold level set at 80% correct. In fact, their performance should be about chance level (50% correct).

Condition 3: Retinotopic reference-frame and non-retinotopic drift-information (Fig. 2-3C): The outermost elements are perceived to move on the trajectory shown by the black dashed lines in Fig 2-3C, whereas the central elements are perceived as stationary. Drift information is added in non-retinotopic coordinates. Based on the *mismatch* between the reference frame and the drift information, as in the previous case, we predict that observers will be at chance.

Condition 4: Non-retinotopic reference-frame and non-retinotopic drift-information (Fig. 2-4): In this condition, the reference-frame is non-retinotopic and the drift information is added in non-retinotopic coordinates. Hence, there is a *match* between the reference-frame and the drift information. Here we will test the main hypothesis of the study: If indeed, adaptive center-surround mechanisms, as observed by Tadin et al. (2003) is general, we expect to replicate their results not only when there is a match in retinotopic coordinates (Condition 1), but also when there is a match in non-retinotopic coordinates. Accordingly, similar to Condition 1, we expect to observe a decrease in performance with

size for high-contrast stimuli and an increase in performance for low-contrast stimuli. Preliminary pilot data showed that the lowest contrast value to perform the non-retinotopic task of Condition 4 above chance level was 5.5%. Therefore, in order to keep consistency in the stimulus parameters, we used this minimum value for *all experiments* except for Condition 1, where we set the low-contrast value to 2.8% in order to match closely the parameters used by Tadin et. al. (2003) in his experiment.

In all these experiments, in order to measure how observers perceive stimuli according to different reference-frames, observers were asked to attend an element according to the reference-frame prevailing for that condition. Hence, in Conditions 1 and 3 observers attended to a retinotopic location (see the yellow arrows in Figures 4 and 7), whereas in Conditions 2 and 4 they attended along the non-retinotopic pathway (see the yellow arrows in Figures 6 and 8). It may be also of interest to ask what happens if the observers' attention is directed to an element outside the reference frame. We studied the dissociation between attention and the reference frame in Lauffs et al. (2019). That study showed that, whereas tracking by focal attention can generate an attention-based reference-frame, the effect is rather small compared with reference-frames generated by motion-based grouping, as in Ternus-Pikler displays. Hence, in the experiments reported in this paper, attention and reference-frame were congruent in order to generate strong effects upon which we could investigate center-surround modulation effects.

Control Condition: No Ternus-Pikler motion (Fig. 2-11): Across all experiments, the path along which the drift information is introduced has only integer multiples of δ from frame to frame. Other task-irrelevant paths contain mainly ambiguous motion (180^{o} phase shift), but there are still some frames containing multiples of δ. For example, consider the

phase shifts added to the Gabors placed in the left-square of the display in Fig. 2-11A. The phase shift from the first frame to the second frame (from 0^o to 180^o) is ambiguous, since it is equally likely to result from an upward or downward drift. Similarly, the transition from the third to the fourth frame is ambiguous (from 2δ to $2\delta + 180^o$). However, the transition from the second to the third frame (from 180^o to 2δ) is not ambiguous and observers may be able to use the transition in this frame alone to determine the direction of drift. To test the possibility that this "artefactual drift information" can be used effectively, we ran a control condition in which we removed the leftmost and rightmost elements of the Ternus-Pikler display as depicted in Figure 11A. If observers are at chance in this experiment, it indicates that the artefactual drift information (e.g., in the transition from the second to the third frame) is not sufficient to accomplish the task. This could be simply due to the fact that this transition is temporally preceded and succeeded by ambiguous counter-phase (180^o) frames. On the other hand, if observers' performance turns out to be above chance in this experiment, this would necessitate a redesign of how phase-shift information is placed in the Ternus-Pikler display so that sufficient drift information exists *only* in the paths tested in specific experiments. Since the interpretation of Conditions 1 to 4 assume that observers will be at chance in this control condition, we ran this condition first to test this fundamental hypothesis. The demos included with this manuscript illustrate each of these stimuli.

2.4. Results

2.4.1. Experiment 1. Condition 1: Retinotopic reference-frame and retinotopic drift-information:

2.4.1.1 Procedure

In this condition we evaluated two contrast values (2.8% and 90%) for each Gabor patch size (0.6, 1.4, and 2.7 deg) matching the values in Tadin et. al. (2003). Observers were instructed to fixate on a fixation cross placed in the middle of the screen for 1000 ms and to *attend* to the element marked with the yellow arrow in Fig. 2-4A.

2.4.1.2 Results and discussion

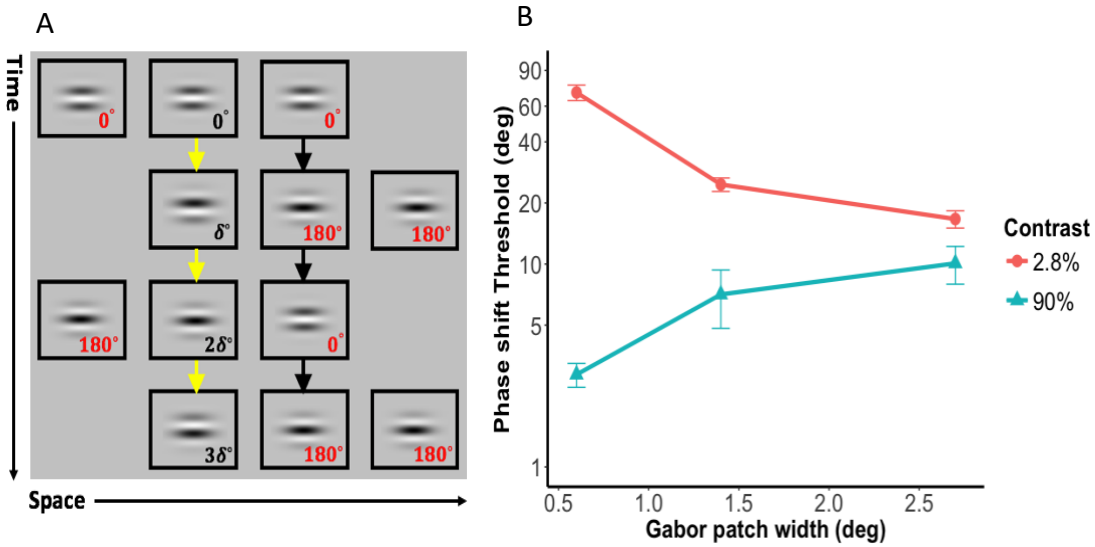

Fig. 2-4. Stimulus for the retinotopic reference-frame and retinotopic drift information condition. The arrows represent the perceived trajectory of the Ternus-Pikler elements (squares) across frames. In this case the two central elements are perceived as stationary. The yellow arrow highlights the element to be attended in the display (the yellow arrows were not displayed in the actual stimulus; see demos). Coherent phase-shift is represented by successive integer multiples of **δ** across frames. The rest of the pathways contained mainly ambiguous motion. (B) Results: Effect of contrast on phase-shift threshold as a function of Gabor patch size. The data represent the average phase-shift threshold across observers (N=6) for low and a high-contrast Gabors. Error bars are ± S.E.M.

A two-way repeated-measures ANOVA (Greenhouse-Geisser corrected where appropriate) showed a statistically significant main effect of size ($F(1.1, 5.5)$, $p = 0.001$, $\eta_p = 0.89$) and contrast (($F(1, 5)$, $p < 0.001$, $\eta_p = 0.93$)). Likewise, there was a statistically

significant interaction between size and contrast ((*F*(2, 10), $p < 0.001$, $\eta_p = 0.97$)). For the low-contrast stimulus, thresholds decreased with increasing patch size suggesting that a *spatial summation mechanism is at work* known to occur for low contrast stimuli (Anderson & Burr, 1991; Kapadia et al., 1999; Watson & Turano, 1995). To the contrary, for the high-contrast stimulus, thresholds increased as stimulus size increased. This effect is consistent with a *spatial suppression mechanism* (Tadin et al., 2006, 2003; Tadin & Lappin, 2005, 2012).

For comparison purposes, in Fig. 2-5 we plot the average data from Tadin's study (dashed lines) alongside our results (solid lines) for the three stimulus sizes. The data from the two studies are similar, even though there were differences in the stimuli. For instance, in Tadin's experiment, observers identified the motion direction of a Gabor patch that abruptly shifted in phase in the middle of a 100 ms presentation interval. Whereas in our study, the stimulus was more complex, containing multiple squares and Gabors, of which observers had to attend to one. The Ternus-Pikler display introduces additional motion compared to Tadin et al's stimulus. The squares used as part of the Ternus-Pikler display provide a strong horizontal motion signal thereby making the resulting reference-frame salient. To test that our results hold even in the absence of these squares, we ran a control experiment on one subject and obtained similar results (see Appendix A). Overall this experiment replicated Tadin et. al.'s (2003; Fig. 3a) results in the Ternus-Pikler stimulus.

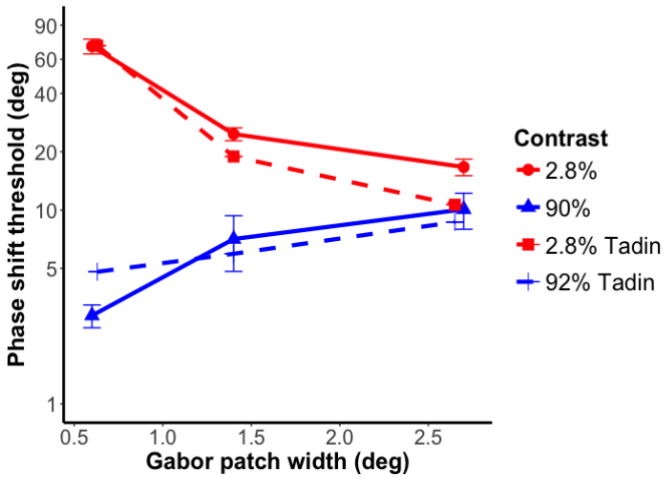

Fig. 2-5. Comparison of Tadin's data (Tadin et al, 2003; Fig. 3a) with the results of our Experiment 1. Solid lines represent the results of Experiment 1 while dashed lines show Tadin's data. Red lines correspond to the low-contrast stimulus and blue lines represent high-contrast stimulus. Error bars are ± S.E.M.

2.4.2. Experiment 2. *Condition 2:* Non-retinotopic reference-frame and retinotopic drift-information:

2.4.2.1 Procedure

Observers were instructed to fixate on a fixation cross placed in the middle of the screen for 1000-ms at the beginning of each trial and to attend to the central element of the display (yellow arrow in Fig. 2-6A). In this condition two contrast values (5.5% and 90%) and three Gabor patch sizes were tested.

2.4.2.2 Results and discussion

It could be argued that, even though the Ternus-Pikler display establishes a non-retinotopic reference-frame, the drift information could be read out from retinotopic

coordinates. For instance, by inspecting Fig. 2-6A, it can be noted that drift information in retinotopic coordinates may be used by the participants to make a decision on the task.

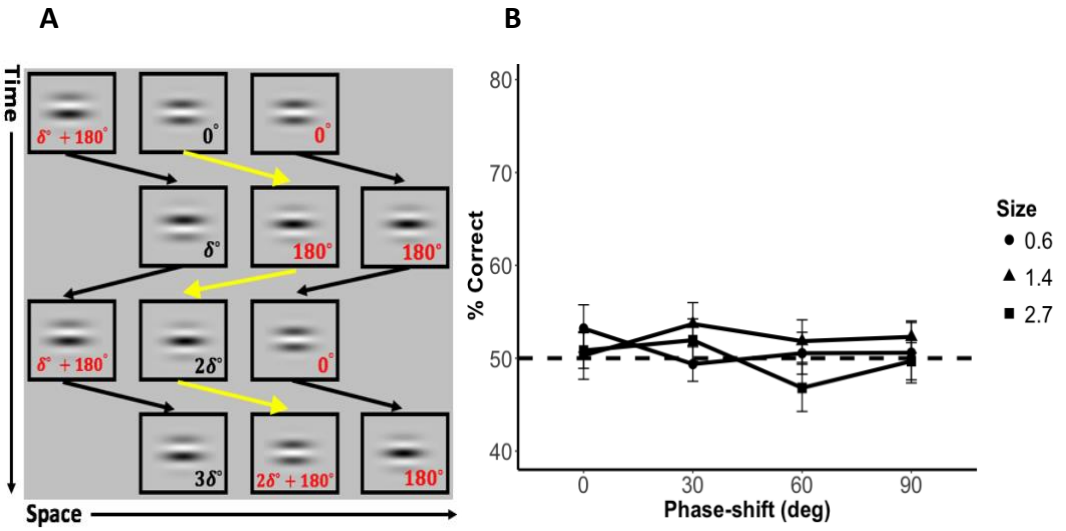

Fig. 2-6. (A) In the group motion percept the three elements appear to move from left to right in a trajectory marked by the black and yellow arrows . The yellow arrows highlight the element to be attended in the display. Coherent phase-shift is inserted in retinotopic coordinates (follow the successive integer multiple values of **δ** across frames). The rest of the motion pathways contained mainly ambiguous motion. (B) Results: Observers were at chance level regardless of stimulus contrast and size. Panel B shows the psychometric functions averaged across observers (N=6) for a stimulus set to 90% contrast (see Appendix for the low-contrast results). Error bars are ± S.E.M.

Nevertheless, the results in Fig. 2-6B show that regardless of phase-shift magnitude (δ) the observers were at chance performance. Participants' anecdotal reports after the experiment stated that the task was impossible to perform. This was the case *regardless* of contrast or size. The data in Fig. 2-6B show the high contrast data. The plot for the low-contrast case can be found in the Appendix. The reason for the poor performance is that in the group motion percept the feature attributes of the elements in one frame of the display are perceptually linked with the elements of subsequent frames in a manner highlighted by the arrows in Fig. 2-6A. Thus, if the trajectory of the yellow arrow in Fig. 2-6A is followed, it can be seen that it possesses mostly ambiguous motion which then explains the

performance. In total, this result shows that information put in retinotopic coordinates does not inform motion direction when the reference system is non-retinotopic.

2.4.3. Experiment 3. *Condition 3:* Retinotopic reference-frame and non-retinotopic drift-information:

2.4.3.1 Procedure

Fig. 2-7A shows the stimuli for Experiment 3. In this case, multiples of are added in non-retinotopic locations whereas the element-motion percept establishes a retinotopic reference-frame. Observers were asked to attend to the central element in the display

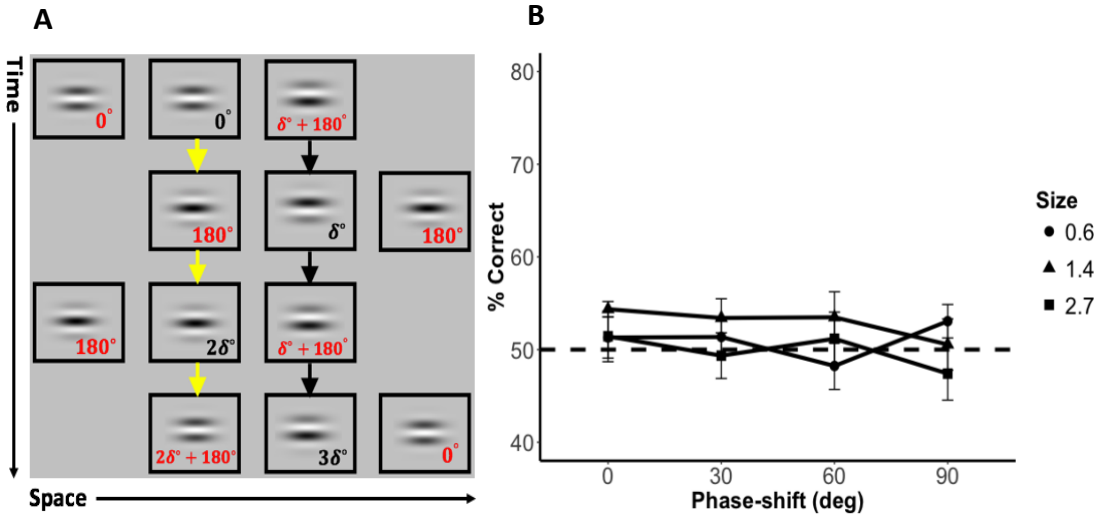

Fig. 2-7. (A) Stimulus for the retinotopic reference-frame and non-retinotopic drift information condition. The display configuration is similar to the control condition, but this time the outermost element of the Ternus-Pikler display were included. This configuration of the Ternus-Pikler display only allows the formation of retinotopic representations (regardless of ISI). Motion information is inserted in non-retinotopic coordinates while perception is established in a retinotopic reference system. Observers were instructed to attend to the element highlighted with the yellow arrow and to report the perceived direction of drift. (B) Results: Average percent correct across observers (N=6) is shown as a function of phase-shift **δ** value. The data come from the 90% contrast stimulus evaluated at all sizes. As predicted, the result shows that observers were at chance performance. Error bars are ± S.E.M.

marked by the yellow arrow in Fig. 2-7A while maintaining fixation in the center of the display on a fixation cross that appeared for 1000-ms at the beginning of each trial. Once

again, we evaluated two contrast values (5.5% and 90%) for each of the three Gabor patch sizes.

2.4.3.2 Results and discussion

From Fig. 2-7B it is clear that observers in the high contrast conditions are at chance level in this experiment. The results for the low-contrast conditions also show chance performance (Appendix B).

This result and the result of the previous experiment indicate that a match between the reference frame and the motion information is essential for detecting the direction of motion. In addition, comparing this stimulus to that in Experiment 5 (cf. Figs 2-7A and 2-11A), one can see that adding or removing the outermost Ternus-Pikler elements (squares) do not influence how motion is perceived since the same retinotopic reference-frame is maintained in both stimuli.

2.4.4. Experiment 4: *Condition 4:* Non-retinotopic reference-frame and non-retinotopic drift information:

2.4.4.1 Procedure

A schematic of the stimulus is depicted in Fig. 2-8A. This time both the reference frame and the coherent phase-shift for motion drift are in non-retinotopic coordinates. That is, the motion correspondences of the elements in the display coincide with the placement of integer multiples of δ values, that together create a coherent drift along non-retinotopic coordinates. Based on previous studies (Agaoglu et. al., 2017; Boi et. al., 2009; Clarke & Herzog, 2013; Clarke et. al., 2015; Grossberg et. al., 2011) and based on the theoretical concept of reference-frame match, we expected that observers will be able to extract this

non-retinotopic drift information and perform at a higher than chance level on the experiment.

As in the other conditions, observers were instructed to fixate on a fixation cross located in the center of the screen for 1000-ms at the beginning of each trial and to attend to the central element in the Ternus-Pikler display (yellow arrows in Fig. 2-8A). In order to capture a transition from spatial suppression to spatial summation, if it exists, five contrast values ranging from low to high were tested (5.5%, 9.16%, 15%, 55%, and 90%).

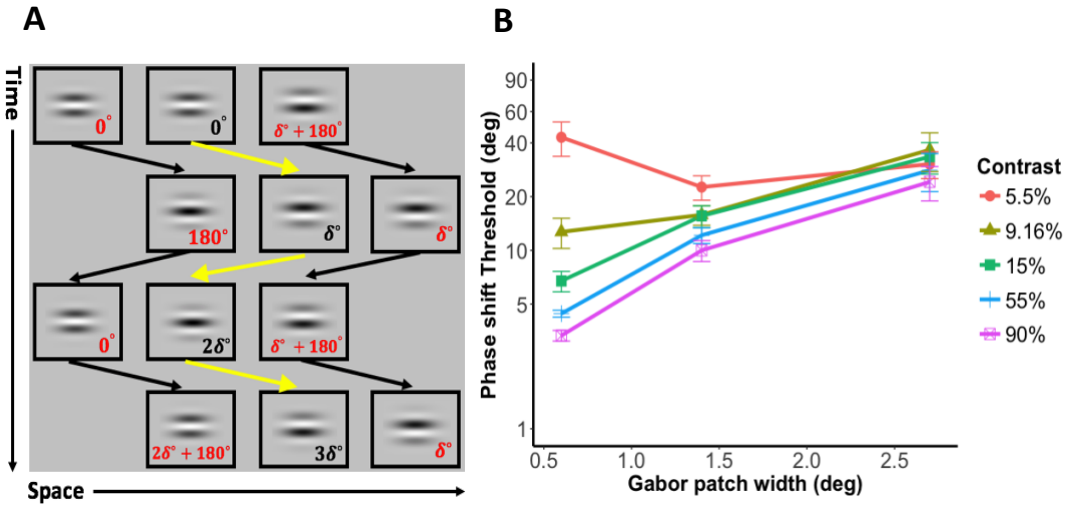

Fig. 2-8. (A) Stimulus for the non-retinotopic reference-frame and non-retinotopic drift information condition. There is agreement between the reference frame and the location of coherent phase-shifts, as both are in non-retinotopic coordinates. The yellow arrow highlights the element to be attended in the display. Observers were instructed to attend this element and report its perceived direction of drift. The rest of the motion pathways contained mainly ambiguous motion. (B) Results. A statistically significant interaction of size and contrast was found. We observe a decrease in performance as a function of size for contrast values up to 15% Michelson contrast. For the low-contrast stimulus, there is an improvement in performance with size. These results suggest that the non-retinotopic processes may be mediated by a similar adaptive center-surround mechanism as the one found in retinotopic representations. Error bars are ± S.E.M.

2.4.4.2 Results and discussion

The results are presented in Fig. 2-8B. A statistically significant interaction between size and contrast (($F(2.55, 12.78)$, $p = 0.006$, $\eta_p = 0.59$)) was found. The main effects of contrast (($F(1.38, 6.88)$, $p = 0.002$, $\eta_p = 0.81$)) and size (($F(1.08, 5.38)$, $p = 0.04$, $\eta_p = 0.58$))

were also statistically significant. It is observed that thresholds increase as a function of stimulus size for high-contrast stimuli. This suggest the presence of a spatial suppression mechanism. Also, note that threshold curves for contrast values between from 90% to 15% run parallel with relatively small improvement in performance as contrast is increased. For low contrast stimulus (5.5%), the threshold decreases as a function of size. This result in turn, signals the transition from spatial suppression to a spatial summation mechanism.

Fig. 2-9 presents the results of Experiment 4 replotted next to Tadin et al's data (Tadin et al., 2003; Fig. 3a). One major quantitative difference is the range of threshold values: Quantitatively, Tadin et al's results are more "compressed" on the threshold axis compared to ours. On the other hand, qualitatively, the general profile of both results is similar. For example, for the lowest- and highest-contrast stimuli, there is improvement of performance as a function of size and deterioration of performance with size, respectively.

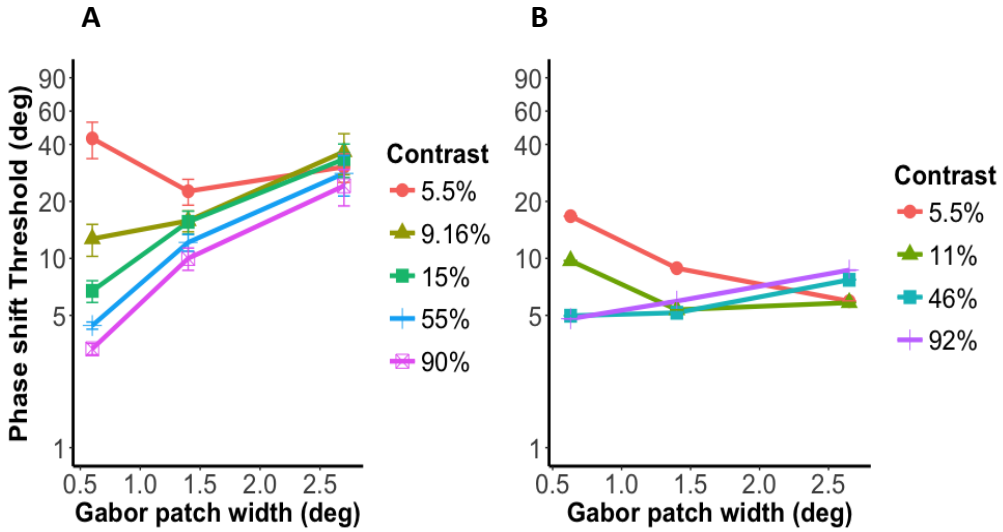

Fig. 2-9. (A) Replot of the results for Experiment 4. (B) Data obtained from Tadin et. al. (2003), Fig. 3a. We extracted the data points for the three stimulus sizes evaluated in our study and for the contrast values examined in that experiment. Error bars are ± S.E.M.

Also, in both results there is a progressive transition from *spatial summation* to *spatial suppression* for contrast values between 5.5% to 15%, which is within the range of contrast

values where spatial grouping transitions to spatial segregation (Takeuchi, 1998). Finally, for contrast values higher than 46%, we observe a plateauing in performance.

On possible reason for the better performance in Tadin et al's results is that our study contains a long ISI between frames while their study did not. Another reason may be the higher complexity of our stimulus in comparison to Tadin et al's. Finally, the threshold differences in this experiment may be also due to differences in the reference-frames: Let us recall that Experiment 4 is a non-retinotopic motion-direction task whereas Tadin's data come from a retinotopic motion-direction task. In other words, the reference-frame in our study was non-retinotopic whereas it was retinotopic in Tadin et al's study.

In order to compare more directly the effects of retinotopic vs non-retinotopic reference-frames, we can compare the results of Experiment 1 (retinotopic task) to those of Experiment 4 (non-retinotopic task) in the present study. In this way, we compare the performances of retinotopic and non-retinotopic reference-frames using the same stimulus paradigm and the same group of observers and hence, eliminating the differences between our and Tadin et al's studies. In Fig. 2-10, we replot the data from Experiment 1 (dashed lines) together with the results from the lowest and highest contrast values from Experiment 4 (solid lines). Data look similar. Notice that for low-contrast stimuli (red lines) there is a decrease in threshold with size. The opposite effect is observed for the high-contrast stimuli (blue lines). Also observe that in some conditions the performance for Experiment 1 is slightly better than the performance for Experiment 4. This may be due to the long ISI included in Experiment 4. In other words, the experiments had different temporal conditions. These results suggest that the nature of the reference-frame, whether it is retinotopic or non-retinotopic, does not affect strongly the computation of motion

information. In Appendix A, we also show that similar results are obtained in the absence of the squares that are part of the Ternus-Pikler display.

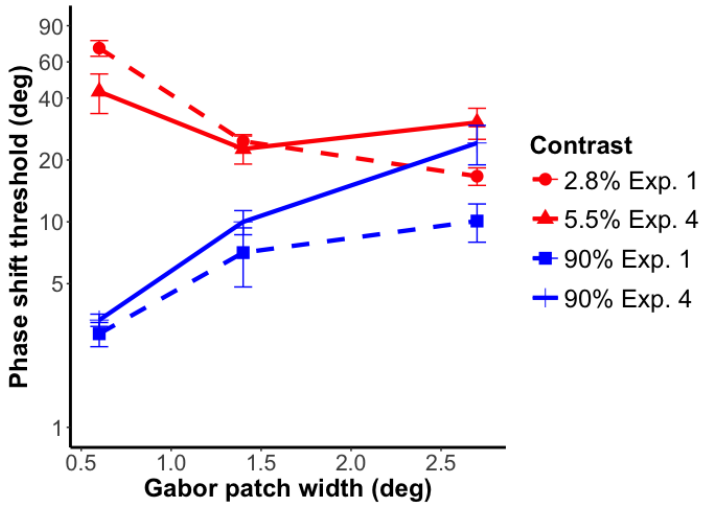

Fig. 2-10. Comparison of results from Experiment 1 and Experiment 4. Dashed lines and solid lines correspond to Experiment 1 and Experiment 4, respectively. The low-contrast stimuli are shown in red while the high-contrast stimuli are shown in blue. Error bars are ± S.E.M.

Taken together, these findings provide support for the main hypothesis of this study, viz., non-retinotopic computation of motion is mediated by an adaptive center-surround mechanism similar to that found in retinotopic coordinates.

2.4.5. Experiment 5. *Control condition:* No Ternus-Pikler motion (Fig. 2-11):

2.4.5.1 Procedure

Participants were instructed to attend to either the left element or the right element (Fig. 2-11) and report the direction of drift. The Gabor patch size and the contrast were 2.65^{o} and 90%, respectively. The critical point was to assess whether the non-ambiguous phase shifts (i.e., phase shifts other than integer multiples of 180^{o}) were used by the subjects to produce above chance performance. These non-ambiguous phase-transitions

occur in the transition from frame two (180°) to frame three (2δ) in the left side of the display, as well as in the transition from frame three to frame four in the right side of the display (from δ + 180° to 3δ).

2.4.5.2 Results and discussion

For each of the four phase-shifts, participants were at chance level (Fig 11B). The presence of multiples of δ, i.e., non-ambiguous phase shifts, in some frames did not allow observers to successfully report the direction of drift. Hence, even though our stimulus is not perfectly ambiguous with respect to the direction of drift (shifts of 180° are ambiguous whereas shifts by δ degrees are not), this control experiment shows that the "artefactual drift information" is unnoticed.

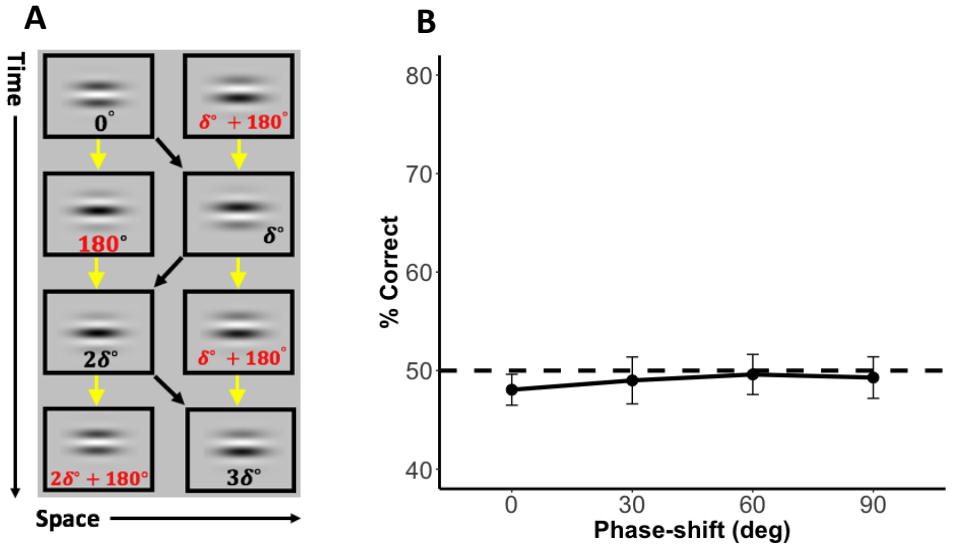

Fig. 2-11. (A) Control condition. The outermost elements of the Ternus-Pikler display were removed. The black arrows indicate the non-retinotopic trajectory of the coherent motion. The yellow arrows show the retinotopic trajectories. Observers were instructed to attend to one of the elements in the display and to report the perceived direction of drift. (B) Results of the control condition. Average percent correct across observers (N=6) as a function of phase-shift δ. Observers are at chance level. Error bars are ± S.E.M.

2.5. General discussion

Motion perception occurs usually in non-retinotopic coordinates. However, the neural correlates underlying non-retinotopic motion processing are still elusive. The major goal of the study was to investigate whether the adaptive center-surround mechanisms reported in Tadin et al. (2003) generalize to non-retinotopic reference-frames. Our results support the hypothesis that adaptive center-surround mechanisms are used in the computation of non-retinotopic motion.

Spatial suppression is thought to be the psychophysical correlate of neurophysiological surround suppression processes (e.g., Tadin, 2015) Moreover, correlational and causal evidence suggest that spatial suppression of background motion signals is critical for the rapid segregation of moving objects (Tadin et al., 2019). More generally, surround-inhibition may be a solution to the noise-saturation problem (Grossberg, 1988): The system needs to amplify weak signals to improve the signal-to-noise ratio. However, when signals become strong, the same amplification may lead to saturation. How can a system be sensitive to weak inputs while avoiding saturation for strong inputs? In other words, how can an adaptive gain-control mechanism be used to match the dynamic range of the processing units (neurons) to that of input signals, given that the latter tends to be much larger than the former. At high input levels, the surround suppression can reduce the net signal thereby preventing saturation. On the other hand, for weak input signals, the surround suppression would be detrimental since it will weaken further the signal burying it in noise. The adaptive strategy reduces the effectiveness of the surround for weak signals (low contrast) thereby leading to summation by the center of the receptive field. The summation in turn amplifies weak signals so that they can be reliably detected in the

presence of noise. Given this adaptive dynamic-range property, it is not surprising to find antagonistic center-surround (or similar) receptive-field profiles throughout the visual system, starting from the retina. However, it remains to be seen at which stage(s) this mechanism operates when it comes to motion detection. In the primate visual processing, visual motion detection starts in the cortex and involves center-surround receptive fields found in V1 (Cavanaugh et al., 2002a), MST (Eifuku & Wurtz, 1998) and in V5/MT (Pack et al., 2005). Correspondences between the characteristics of psychophysical spatial-suppression and the properties of suppressive center-surround receptive fields in area MT suggest this area as the main locus of these adaptive mechanisms (Liu et al., 2016; Schallmo et al., 2018; Tadin & Lappin, 2005, 2012; Tadin et al., 2003; Tadin et al., 2011). For instance, the dependency of spatial suppression on contrast accords with contrast dependency of a population of MT neurons (Pack et al., 2005); MAE which is linked to MT mechanisms, is reduced with large high-contrast stimulus (Tadin et al., 2003); isoluminant moving stimuli are unable to produce spatial suppression effects, a result consistent with weak MT responses to isoluminant gratings (Gegenfurtner et al., 1994; Tadin et al., 2003). Therefore, it is proposed that spatial suppression is, at least in part, a behavioral correlate of surround suppression in cortical area MT (Tadin, 2015). The above arguments apply to spatial suppression generated with stimuli encoded in retinotopic coordinates. The results of Experiment 5 support the involvement of similar mechanisms in non-retinotopic motion computation. In an fMRI study, also using the Ternus-Pikler display, Thunell et. al. (2016) found that the average blood-oxygen-level (BOLD) activation in areas V1, V2 and V3 correlated with retinotopic percepts but not with non-retinotopic percepts. On the other hand, the human motion processing complex (hMT+)

was active with both retinotopic and non-retinotopic encoding which suggested the hMT+ as the first visual area encoding non-retinotopic percepts. Thus, the available evidence about the location of non-retinotopic encoding in the brain also suggests area MT as a potential candidate. Given the similarities in the use of adaptive center-surround mechanisms in retinotopic as well as non-retinotopic processing, both in terms of behavioral performance and potential neural correlates, it remains to determine whether adaptive center-surround mechanisms are "inherited" from one computation to another (e.g., from retinotopic motion computation to non-retinotopic motion computation) or they are independently implemented at each stage.

To our knowledge, this is the first study to provide evidence for the existence of adaptive center-surround mechanisms in non-retinotopic encoding. The flexible spatial integration mechanisms described in this work might be a way in which vision deals with the wide variability of motion signals found in the natural environment. Even though the early visual system is organized retinotopically (Tootell et. al. ,1998), this retinotopic organization is insufficient to support perception under natural viewing conditions (Boi et al., 2009). Additionally, our results add further evidence that most visual features are processed in non-retinotopic coordinates, including motion (Boi et al., 2009, Thunell et al., 2016), attention (Boi et al., 2009; Boi et al., 2011), fine spatial detail such as vernier offsets and feature fusion (Ogmen et al., 2006; Scharnowski et al., 2007), backward masking (Noory et al., 2015), but not adaptation in the tilt aftereffect (Boi et al., 2011).

Fig. 2-12 shows the "two-stage" model that we proposed for non-retinotopic processing (Ogmen & Herzog, 2010). According to this model, the first stage of processing consists

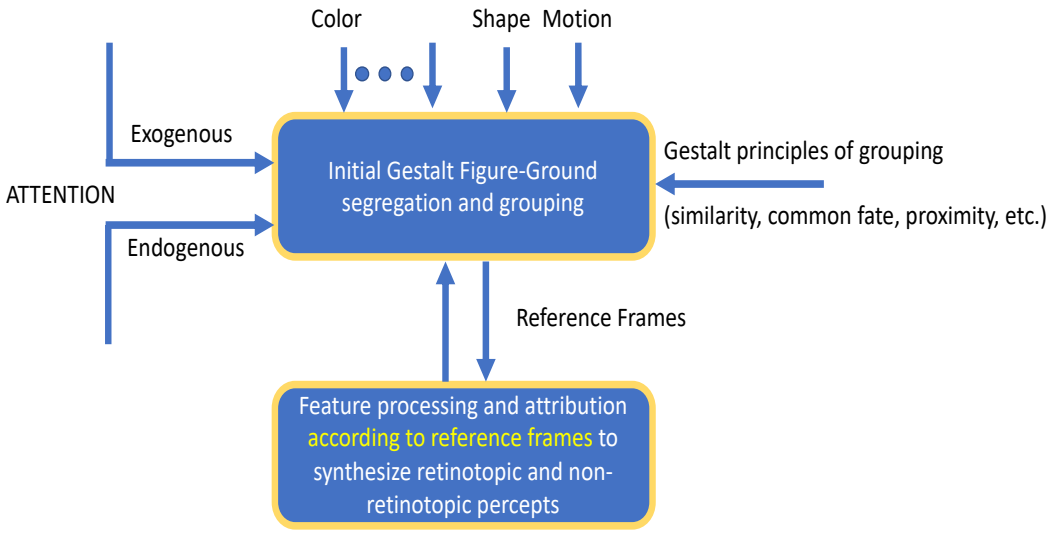

Fig. 2-12. The "two-stage" model for non-retinotopic processing. The model assumes that the primitive building blocks of perceptual processing are groups, as proposed by Gestalt psychologist. Hence, the first stage of visual processing consists of separating figure from ground and grouping stimuli according to Gestalt principles of grouping. These groups, in turn, provide reference-frames that are used to synthesize retinotopic and non-retinotopic percepts.

of grouping stimuli using Gestalt principles of grouping. This grouping process generates reference-frames, which in turn are used in a second stage where non-retinotopic processes are carried out according to these reference-frames. For example, different motion signals on the bicycle of the example in Fig. 2-1 are grouped together (e.g., by the principle of "common fate") and a "common motion vector" is selected as reference-frame (e.g., Johansson, 1973, 1974; Agaoglu et al., 2015, 2017). This reference-frame is passed on to the following stage where it provides coordinates to judge relative motion according to an exocentric (non-retinotopic) reference-frame. In this study, we used the Ternus-Pikler display to generate retinotopic and non-retinotopic reference frames that are pitted against each other due to overlapping elements between different frames. Changing the geometry and/or timing of Ternus-Pikler elements allows the generation of experimentally-controlled reference frames. The key however is not the specifics of the stimulus that

generates the reference frame. It could be the bicycle in a natural environment as in Fig. 2-1. Regardless how it is generated, the resulting reference-frame is passed on to the second stage where it controls how stimuli are processed and integrated (Fig. 2-12). Future work will determine where in this model the adaptive center-surround organization may be operating.

Chapter 3. Adaptive Mechanisms of Visual Motion Discrimination, Integration, and Segregation[3]

3.1. Introduction

The evolutionary success of species depends critically on their ability to process and extract sensory information under a wide-range of luminance levels. For example on a normal day, the luminance of the environment can change from 10^8 cd.m^{-2} (sunlight) to 10^{-3} cd.m^{-2} (starlight) (Graham, 2011; Kalloniatis & Lu, 2005; Stockman & Sharpe, 2006), i.e. a dynamic range of 220 dB. This imposes a big challenge to the sensory organs due to the limited dynamic range of its neurons (Publio et al., 2012). Even with a fixed background luminance, the stimulus *contrast* can vary significantly across the visual field due to shading, camouflage, etc. Given this fundamental problem, the brain allocates a significant amount of resources and deploys multiple solutions that work in synergy. One group of solutions can be called *structural* solutions. For example, the visual system uses two types of structurally different sensors, rods and cones, to deal with different ranges of luminance. A second group of solutions can be called *functional* solutions, i.e., properties that result from nonlinear dynamics of neural networks. Among these functional solutions are adaptation, spatial and temporal integration, and attentional selection. The brain seems to use these multiple solutions in synergy because each solution comes with its own trade-

[3] Presented at the VSS 2020 conference (Peñaloza et al., 2020), published in peer-reviewed journal (Peñaloza et al., 2021).

offs. For example, while spatial (temporal) integration can improve signal-to-noise ratio, this improvement comes at the cost of a loss in spatial (temporal) resolution (see Appendix A.1). A parallel structural solution, viz., two distinct parallel pathways with different spatio-temporal sensitivities (parvo- and magno-cellular streams) can provide a joint optimization in terms of spatio-temporal sensitivity and resolution. We believe that the functional properties are built upon a few canonical but flexible mechanisms that exploit the adaptive and nonlinear properties of neural networks. Here, we demonstrate how these mechanisms can provide solutions to these problems. A key consideration in our approach is spatio-temporal dynamics. Natural viewing is highly dynamic: in addition to the movements of objects and the observer, we make three to four saccades every second. During saccades, perception is strongly suppressed, leaving fixation times to receive and process information. Fixation durations vary according to the task demands, with mean durations ranging from 260-330 ms for scene perception and 180-275 ms for visual search (Rayner, 2009). Natural viewing conditions strongly suggest that transient characteristics of information processing are essential for understanding the function of the brain in its ecological context (Ogmen, 1993). In fact, in addition to the ecological behavior of biological systems, noise considerations also highlight the importance of understanding the transient characteristics of a real-time system operating in noisy environments. Intrinsic noise (e.g., random voltage fluctuations in neurons), particularly when it is high, constantly pushes the system out of its steady-state equilibrium point, and the system operates under its transient dynamics in its tracking of the desired equilibrium point. Yet many experiments and models characterize the *steady-state* behavior of the nervous system. Here, we build a dynamic neural network model and compare its predictions to recent data

highlighting how spatial and temporal integration can address the dynamic-range problem. The model's novelty is not based on the canonical mechanisms it uses but rather how these mechanisms are integrated to achieve the desired adaptive behaviors, much like in an electronic circuit built upon canonical components like transistors, operational amplifiers resistors, etc. Whereas these circuits are built using similar components, they can fulfill drastically different functions depending on their architecture and parameter choices.

The data that we use reveal some of the adaptive properties of the human motion-processing system. In a series of experiments, Tadin and colleagues (Tadin et al., 2003; Tadin & Lappin, 2005) showed that motion-direction discrimination in humans increases with increasing stimulus size for low-contrast stimuli (suggesting spatial summation) (Anderson & Burr, 1991; Anderson et al., 1985) and the reverse occurs at high-contrast conditions, i.e., the accuracy to discriminate motion paradoxically decreases as stimulus size increases (suggesting spatial suppression). That is, for motion signals with a low signal-to-noise ratio, the visual system *summates* information over space, whereas for motion signals with high signal-to-noise ratios, the visual system appears to inhibit spatial integration to improve spatial resolution. Hence, the spatial summation is flexible and adapts to varying stimulus conditions.

How does the visual motion system achieve this? A possible candidate is an antagonistic center-surround mechanism. Antagonistic center-surround receptive-field organization is a pervasive neural mechanism in the brain (Allman et al., 1985). In primate motion processing, center-surround organization starts at the retina and is found throughout the visual system, including areas V1 (Tsui et al., 2010), MST, and especially the key motion area MT (Churan et al., 2008; Gautama & Van Hulle, 2001; Tadin, 2015; Tadin et

al., 2011). In particular, nonlinear center-surround interactions have been associated with normalization processes in the brain (Grossberg, 1988; Carandini & Heeger, 2012). Normalization offers a powerful nonlinear mechanism that serves a variety of visual phenomena, including light adaptation (Sperling & Sondhi, 1968), contrast and automatic gain control (Grossberg, 1988; Wilson & Kim, 1998), decorrelated encoding of natural images (Schwartz & Simoncelli, 2001), among others. In fact, Carandini and Heeger (2012) proposed that normalization should be considered a canonical neural computation.

In the literature, other models have been proposed to account for this adaptive spatial-integration mechanism (Tadin & Lappin, 2005; and Betts et al., 2012). However, these models are limited in several respects. For example, they lack spatial representation (as they compute the neural response at a single point in space). Further, the models' responses are computed by arbitrary subtractive or divisive operations that do not emerge from the architecture of the models. Finally, previous models do not include the dynamics of the system and assume a steady-state behavior. Taken together these restrictions make previous models less biologically realistic. Thus, to overcome these limitations, in this work, we propose a three-layer dynamic and biologically plausible neural model that may explain how the visual system computes (1) contrast normalization and (2) adapts to stimulus conditions of varying intensities. First, we show that the model is able to normalize one-dimensional contrast inputs to the same values regardless of background intensity as a result of its nonlinear dynamics. Second, we simulate behavioral data that show that, depending on stimulus strengths, the visual system transitions from spatial integration to spatial suppression. Importantly, these properties emerge from the dynamics of our architecture. Our model's design, properties, and architecture allows us to make

general predictions of its behavior; however the selected data provided us with a focused specific area where we tested our predictions. Thus, our approach is a combination of prediction and postdiction. Our results demonstrate that the model accounts well for this adaptive mechanism revealed by behavioral data. Finally, via a minimal modification to the network, we show that the model also captures the dynamics that characterize spatial suppression of background motion signals and the rapid segmentation of moving objects (Tadin et al., 2019).

3.2. Description of the model

3.2.1. Model Architecture and the Canonical Equation

As depicted in Fig. 3-1, the model is feed-forward and consists of two layers. The first layer implements contrast-normalization whereas the second layer implements adaptive center-surround organization. The canonical equation used in our modeling is derived from the Hodgkin-Huxley electrical-circuit model for a membrane patch. We call it canonical because this general "standard form" is used to model all neurons in our model.

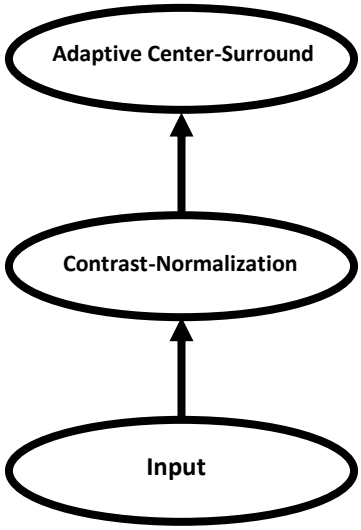

Fig. 3-1. General schematic of the model. The model consists of an input layer, a contrast normalization layer and an adaptive center-surround layer.

Different types of functional behavior are obtained by different choices of the parameters in this model. Fig. 3-2 shows the electrical-circuit equivalent for a membrane patch proposed by Hodgkin and Huxley. The variable V_m represents the membrane potential, C_m the capacitance of the membrane, g_{Na} , g_K, and g_l correspond to sodium, potassium, and leak conductances, respectively. E_{Na} , E_K , E_l are the Nernst, or reversal, potentials corresponding to each of these channels. The differential equation corresponding to this circuit can be derived as follows:

$$C_m \frac{dV_m}{dt} = -(E_l + V_m)g_l + (E_{Na} - V_m)g_{Na} - (E_K + V_m)g_K \qquad (1)$$

where *t* is time. This model, which describes how the membrane potential is controlled across a small membrane patch, had been generalized to represent an entire neuron. For

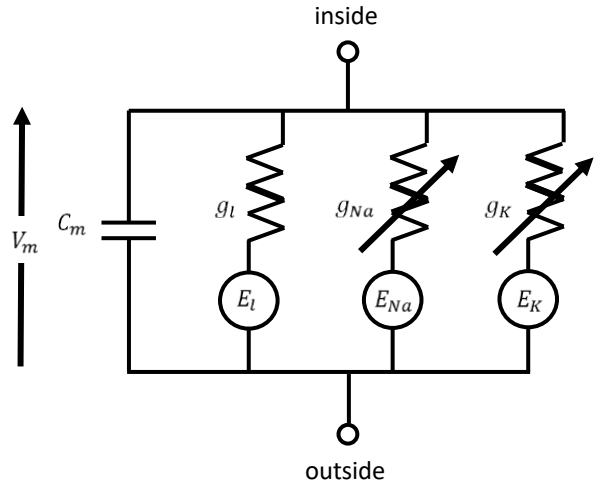

Fig. 3-2. Equivalent electrical-circuit for the Hodgkin-Huxley equation.
See main text for a detailed description.

this purpose, the potential difference across the membrane patch V_m is replaced by a variable, x_i, that represents the membrane potential of the *i*th neuron, rather than just the voltage difference across a small membrane patch. Let then $x_i = E_l + V_m$. Here x_i denotes the membrane potential for the neuron i, shifted by a constant E_l for mathematical convenience. Instead of specific ionic labels for the conductances, they are grouped into three categories: (i) those for which an increase in conductance leads to depolarization from the resting potential (cf. g_{Na} in Eqn. 1), (ii) those for which an increase in conductance leads to hyperpolarization from the resting potential (cf. g_K in Eqn. 1), and (iii) passive, i.e., fixed conductances (cf. g_l in Eqn. 1). Note that the generalized model includes ligand-gated (ionotropic) channels converting synaptic inputs into post-synaptic depolarization (excitatory post-synaptic potentials, EPSPs) or hyperpolarization (inhibitory post-synaptic potentials, IPSPs). Depolarizing and hyperpolarizing conductances are then represented by excitatory and inhibitory inputs to the neuron, I_{exc} and I_{inh} , respectively. By making these substitutions into Eqn 1 above, we obtain:

$$\frac{dx_i}{dt} = -Ax_i + (B - x_i)I_{exc} - (D + x_i)I_{inh} \qquad (2)$$

where $A = \frac{1}{C_m} g_l$, $B = E_{Na} + E_l$, $I_{exc} = \frac{1}{C_m} g_{Na}$, $D = E_K + E_l$, and $I_{inh} = \frac{1}{C_m} g_K$

Note that *A, B, D,* I_{exc} and I_{inh} are all non-negative.

This equation is also known as, for example, the Hodgkin-Huxley model, shunting model, and multiplicative model of membrane potential. In this manuscript, we will use

the term multiplicative membrane equation. The term multiplicative is used because the inputs are multiplied by the membrane potential on the right-hand side of the equation (Grossberg, 1988). This is in contrast to the additive model of membrane potential (also known as the leaky integrator model), which in itself cannot implement nonlinear operations like normalization (see Appendix D for details).

3.2.2. Contrast-normalization Layer

If we calculate the steady-state of Eqn. 2 by setting $\frac{dx_i}{dt} = 0$, we obtain:

$$x_i = \frac{BI_{exc} - DI_{inh}}{A + I_{exc} + I_{inh}} \qquad (3)$$

From this equation, we can see that the weighted inhibitory input is subtracted from the weighted excitatory input in the numerator. This is *subtractive inhibition.* We also see that the inhibitory input also appears in the denominator providing *divisive inhibition.*

To show that this equation limits the activity x_i within a bounded interval, consider the limits when the excitatory and inhibitory inputs go to infinity:

$$\lim_{I_{exc} \to \infty} x_i = B \text{ and } \lim_{I_{inh} \to \infty} x_i = -D. \qquad (4)$$

Hence the activity $x_i \in [-D, B]$.

This layer is designed so that its activity represents contrast values of its input. Note that the layer normalizes its inputs and hence what is normalized depends on the information carried by the inputs. If, for example, the input carries luminance information, then the network extracts luminance contrast. If the input represents motion information, then the network extracts motion contrast.

Since the activity is bounded, we used a bounded measure of contrast, viz., Michelson contrast. Michelson contrast M_c is defined by:

$$M_c = \frac{L_{max} - L_{min}}{L_{max} + L_{min}} \qquad (5)$$

where L_{max} and L_{min} are non-negative maximum and minimum, respectively, values of the variable for which contrast is computed. M_c is bounded in the interval [0,1]; it is minimum, i.e., zero when $L_{max} = L_{min}$ and maximum, i.e., 1, when $L_{max} \gg L_{min}$. Comparing Eqns (3) and (5), one can see that we can choose *D=0* and *B=1* to bring x_i to the same range as M_c . Accordingly, Eqn (2) becomes:

$$\frac{dx_i}{dt} = -Ax_i + (1 - x_i)I_{exc} - x_i I_{inh} \qquad (6)$$

with steady-state:

$$x_i = \frac{I_{exc}}{A + I_{exc} + I_{inh}} \qquad (7)$$

Finally, a neuron can receive inputs from a single to thousands of other neurons. The organization of excitatory and inhibitory inputs forms the receptive-field of the neuron. The first layer shown at the bottom of Fig. 3-1 projects the inputs through on-center off-surround connections to the contrast normalization stage. Thus, incorporating this receptive-field organization into Eqn. 6, we obtain:

$$\frac{dx_i}{dt} = -Ax_i + (1 - x_i) \sum_{j=i-n}^{i+n} K_{j-i}^{center} \cdot I_j - x_i \sum_{j=i-n}^{i+n} K_{j-i}^{surround} \cdot I_j \qquad (8)$$

with $0 \leq i \leq n$, where *i* is the spatial index denoting retinotopic positions. The size of the layer is set to *n+1* neurons. Here, for simplicity, we use one-dimensional space, with a single index *i* reflecting the spatial position of the neuron in the network. The variable x_i represents the activity, i.e., the membrane potential, of the neuron at the spatial location *i.*

$K^{center/surround}$ are the center and surround spatial kernels representing synaptic weights converging on neuron *i*. and I_i is the input signal at spatial location *i*.

As in Eqn. 3, one can calculate the steady-state of x_i by setting $\frac{dx_i}{dt} = 0$ and solving for x_i:

$$x_i = \frac{\sum_{j=i-n}^{i+n} K_{j-i}^{center} \cdot I_j}{A + \sum_{j=i-n}^{i+n} K_{j-i}^{center} \cdot I_j + \sum_{j=i-n}^{i+n} K_{j-i}^{surround} \cdot I_j} \quad (9)$$

Inspection of this equation shows that at steady-state the activity is proportional to the excitatory input (numerator) divided by a sum of excitatory and inhibitory inputs and a constant *A* (denominator). This division effectively normalizes the signal with values within a range [0,1]. Variants of this static equation have been used to model well various normalization data (Carandini & Heeger, 2012). We highlight here that, as shown above, normalization results from the multiplicative property of the Hodgkin-Huxley model as opposed to additive networks, like integrate-fire-networks (Burkitt, 2006; Feng, 2001) (See Appendix D).

An unbalanced kernel *K* was used so as to maintain a non-zero output for the uniform regions of the motion information signal, i.e., the spatial kernel *K* was selected to have low-pass characteristics. Otherwise, if a balanced kernel had been used, then only the signal's edges would have been detected while the uniform portions of the signal would have been suppressed. The contrast normalization stage consisted of 3600 neurons arranged into a spatially one-dimensional network. The rectangular kernel *K* was composed of 3600 neurons of which five neurons were excitatory and the rest of the neurons were inhibitory:

$$K(x)^{center} = 2 \; for \; |x| \leq 2; \; 0 \; otherwise$$

(10)

$$K(x)^{surround} = 0.5\ for\ |x| \leq 4;\ 0.009\ otherwise$$

3.2.3. Adaptive Center-surround Layer

In an ideal situation, the output of the contrast-normalization stage would be sufficient for dealing with dynamic-range variations. However, like any physical system, the nervous system has intrinsic noise, which limits signal reliability when neural activities are weak. Inspection of Fig. 3-3B shows that the output of the neuron is low for low-contrast signals. Hence, if one were to add random noise to neural activities, the signal-to-noise ratio for low-contrast stimuli would be low. The *spatial integration* strategy suggests that an improvement can be obtained by increasing the size of the spatial summation area when contrast is low, as evident in the experiment by Tadin and colleagues (2003; see Appendix A.1). To implement the adaptive spatial integration strategy, we used an equation identical to Eqn. (2) with the exception of the choice of receptive-field parameters and nonlinearities:

$$\frac{1}{\tau}\frac{dy_i}{dt} = -A'y_i + (B' - y_i)\sum_{j=i-n}^{i+n} G_{j-i}^{center} \cdot f\left(x_j\right) - (D' + y_i)\sum_{j=i-n}^{i+n} G_{j-i}^{surround} \cdot g(\,x_j) \quad (11)$$

where y_i is the membrane potential of a neuron at location i, the constants A', B', and D' have the same interpretation as in Eqn. (2). Function G represents the center and surround Gaussian receptive fields and τ is a time-scaling parameter that controls the time response of the neuron. A key assumption of the model that allows contrast-based integration is that

the nonlinearities for the center receptive-field $f(x)$ and the surround receptive-field $g(x)$ are different. A depiction of the nonlinearities as a function of the normalized membrane potential is shown in Fig. 3-3A. These sigmoid nonlinearities represent the thresholding (lower plateau of the function), the gain, and the saturation of the output signals. As one can see from the figure, the excitatory outputs corresponding to the center of the receptive-field have a higher gain and higher saturation point compared to the inhibitory outputs that feed the surround of the receptive field.

The nonlinearities for the model were generated with Eqn. (12) with parameters $[Amp_c = 7,\ b_c = 75.25, a_c = 0.261,\ Amp_s = 6.5,\ b_s = 80.79, a_s = 0.268]$, where $b_{c/s}$ and $a_{c/s}$ represent the slope and the center of the sigmoidal for the center and surround non-linearities, respectively. We used 3600 neurons arranged into one-dimensional spatial representation for the adaptive center-surround stage. Further, the simulation parameters for the constant decay and the Nernst potentials were set to $A' = 1, B' = 100, D' = 50$.

$$f(x)_{center} = \begin{cases} 0.23 & for\ x\ < 0.22 \\ \dfrac{Amp_c}{1 + e^{-b_c * (x - a_c)}} & for\ x \geq 0.22 \end{cases} \tag{12}$$

$$g(x)_{surrond} = \begin{cases} 0.25 & for\ x\ < 0.23 \\ \dfrac{Amp_s}{1 + e^{-b_s * (x - a_s)}} & for\ x \geq 0.23 \end{cases}$$

Finally, the Gaussian receptive fields were generated with the following equation:

$$G^{center/surround} = A_{c/s} * e^{-\frac{(x - x_0)^2}{{\sigma_{c/s}}^2}} \tag{13}$$

The parameters $A_{c/s}$, $\sigma_{c/s}$, x_0 denote the amplitude, the spread, and the mid-point location of the Gaussian for the center and surround receptive fields, respectively. The Gaussian parameters used for the simulations were: $[A_c = 2, \sigma_c = 67,\ A_s = 1, \sigma_s = 300]$. Each receptive field received input from all the other 3599 neurons.

3.3. Simulation Methods

3.3.1. Input stimulus

In this work, the input stimulus is motion strength. We used rectangular signals that varied in width and height. Fig. 3-3A (bottom subplot) illustrates how the stimulus was represented as input to the network. In order to determine stimulus intensities to produce specific Michelson contrast values, one can re-write Eqn. (5) as follows:

$$M_c = \frac{\Delta I}{\Delta I + 2L_0} \qquad (14)$$

where $L_0 = L_{min}$ and $\Delta I = I_{max} - I_{min}$. Solving for ΔI, one obtains:

$$\Delta I = \frac{2M_c L_0}{1 - M_c} \qquad (15)$$

M_c represents the Michelson contrast, L_0 stands for the background value, and the delta increment is designated as ΔI. Throughout the simulation the background luminance was set to 60.5 cd m^{-2}. Each layer in the model contained 3600 neurons, representing a one-dimensional retinotopic map and the input is defined on this retinotopic map with 3600 spatial samples. All results in this manuscript are based on one-spatial dimension with one exception (Fig. 3-4) where we used a 512x512 grey-scale image for a 2D demonstration.

3.3.2. Summary of Model Equations and Parameters

Here we summarize the equations and the parameters (Table 3-1) used in the simulations for completeness. Additionally, we performed a sensitivity analysis to

determine which parameters are critical to the model's predictions. Details can be found in Appendix A.3 and the supplementary materials.

Contrast-normalization layer:

$$\frac{dx_i}{dt} = -Ax_i + (1 - x_i) \sum_{j=i-n}^{i+n} K_{j-i}^{center} \cdot I_j - x_i \sum_{j=i-n}^{i+n} K_{j-i}^{surround} \cdot I_j \qquad (8)$$

$$K(x)^{center} = 2 \; for \; |x| \leq 2; \; 0 \; otherwise$$

$$K(x)^{surround} = 0.5 \; for \; |x| \leq 4; \; 0.009 \; otherwise \qquad (10)$$

Adaptive center-surround layer:

$$\frac{1}{\tau}\frac{dy_i}{dt} = -A'y_i + (B' - y_i) \sum_{j=i-n}^{i+n} G_{j-i}^{center} \cdot f(x_j) - (D' + y_i) \sum_{j=i-n}^{i+n} G_{j-i}^{surround} \cdot g(x_j) \qquad (11)$$

$$f(x)_{center} = \begin{cases} 0.23 & for \; x < 0.22 \\ \dfrac{Amp_c}{1 + e^{-b_c * (x - a_c)}} & for \; x \geq 0.22 \end{cases} \qquad (12)$$

$$g(x)_{surrond} = \begin{cases} 0.25 & for \; x < 0.23 \\ \dfrac{Amp_s}{1 + e^{-b_s * (x - a_s)}} & for \; x \geq 0.23 \end{cases}$$

$$G_{\frac{surround}{s}}^{center} = A_{\frac{c}{s}} * e^{-\frac{(x - x_0)^2}{\sigma_{\frac{c}{s}}^2}} \qquad (13)$$

Table 3-1. Model's parameters.

PARAMETER	DESCRIPTION	VALUE	FIXED PARAMETER?	SENSITIVITY ANALYSIS
A	Decay rate	0.5	Yes	No
B	Nernst potential for depolarization	1	Yes	No
D	Nernst potential for hyperpolarization	0	Yes	No
K^{center}	Spatial kernel for the center RF	4	Yes	No
$K^{surround}$	Spatial kernel for the surround RF	0.5	Yes	No
A'	Decay rate	1	Yes	No
B'	Nernst potential for depolarization	100	Yes	No
D'	Nernst potential for hyperpolarization	50	Yes	No
A_c	Amplitude of the Gaussian center RF	2	Yes	No
A_s	Amplitude of the Gaussian surround RF	1	Yes	No
τ	Time-scaling parameter	1	Yes	No
σ_c	Spread of the center-Gaussian receptive field	67	No	**Yes**: this parameter determines the spatial center summation extent of the model
σ_s	Spread of the surround-Gaussian receptive field	300	No	**Yes**: this parameter determines the spatial surround summation extent of the model
a_c	Sigmoidal midpoint for the center non-linearity	0.261	No	**Yes**: this parameter controls the rightward/leftward shifts of the center non-linearity
a_s	Sigmoidal midpoint for the surround non-linearity	0.268	No	**Yes**: this parameter controls the rightward/leftward shifts of the surround non-linearity
b_c	Slope of the center non-linearity	75.25	No	**Yes**: this parameter controls the gain of the center non-linearity
b_s	Slope of the surround non-linearity	80.79	No	**Yes**: this parameter controls the gain of the surround non-linearity
Amp_c	Saturation of the center non-linearity	7	No	**Yes**: this parameter controls the saturation value of the center non-linearity
Amp_s	Saturation of the surround non-linearity	6.5	No	**Yes**: this parameter controls the saturation value of

		the surround non-linearity

3.3.3. Model simulations

The system of ODEs given in the previous section was numerically solved using the dynamic programing language Julia (version 1.4.1). We used the Backward Differentiation Formula (BDF) solver from the SUNDIALS suit (https://computing.llnl.gov/projects/sundials). The BDF is a classic method for stiff equations.

3.3.4. Model fitting

The model had a total of eight free parameters consisting of the six parameters of the center-surround nonlinearities and the two standard deviations of the center and surround receptive fields (See Table 3-1). We used MATLAB's (R2018b) least-squares fitting procedure (lsqcurvefit.m) from the Curve Fitting and Optimization Toolboxes to fit the data. Goodness-of-fit was assessed with a *Chi-square* test:

$$\chi^2 = \sum_i \frac{(m_i - t_i)^2}{\sigma_i^2} \qquad (16)$$

where m_i represents the model's output for the *i*th stimulus condition, t_i represents the average threshold data for that condition, and σ_i^2 was the between-subject variance of the threshold estimate.

3.3.5. Code availability

The code used for model simulations in the study are available at:

https://github.com/bipenalo/Adaptive_CS_Mechanisms.git

3.4. Results

3.4.1. Contrast normalization

Fig. 3-3B shows the steady-state response of the *contrast-normalization stage* to a rectangular input signal set at 2.8% and 92% contrast. The output of this stage is contrast

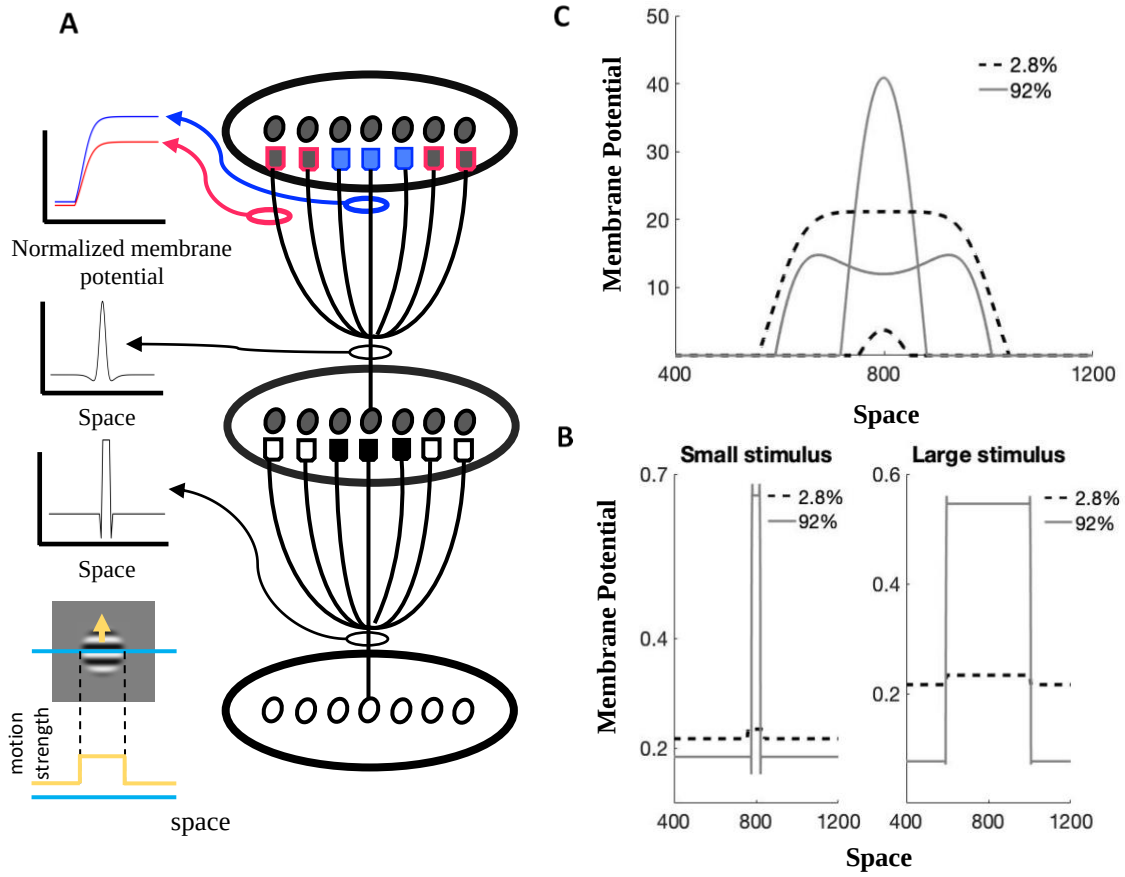

Fig. 3-3. The architecture of the neural model. A Schematic of the model. The bottom, middle and upper ovals represent the input, contrast normalization and adaptive center-surround stages. Black/white and blue/red synaptic symbols represent excitatory and inhibitory connections, respectively. To avoid clutter, only a representative set of connections from a single neuron in each stage of the model is shown. The plots to the left of the schematic highlight some of the properties: The first subplot from the bottom depicts how we represented the input signal to the model, i.e. motion strength varying in size and contrast. The next two subplots depict the center-surround spatial kernels used for the contrast normalization and adaptive center-surround layer, respectively. The top subplot shows the non-linearities for the excitatory center (blue) and inhibitory surround (red) connections. **B** Response of the contrast normalization layer to two stimulus sizes (small and large) and two contrast values (2.8% and 92%). The output is contrast-dependent with higher response to higher contrast. **C** Response of the adaptive center-surround layer to a small and a large stimulus set at low and high contrast. Notice that at low contrast (black dashed line) the membrane potential increases with size (spatial summation). The opposite effect is observed at high contrast. That is, the response of the model is attenuated with increases in stimulus size (spatial suppression).

dependent, i.e., it normalizes its activity within a fixed interval [0,1] and produces higher activities for higher contrast values. The configuration of the spatial kernel *K* (second subplot from the bottom in Fig. 3-3A) was selected to have low-pass characteristics, i.e., we were interested in obtaining a smoother version of the stimulus. An important property of Eqn (8) is that the sensitivity of the network remains, for the most part, constant regardless of changes in background luminance-level. In other words, the network generates an *automatic gain control* (Grossberg, 1988). This is relevant because, as mentioned above, the human visual system is able to operate over a change in illumination by more than a factor of 10^{11} (Stockman & Sharpe, 2006; Stockman et al., 2006). The gain control stage guarantees that the output of the model is mostly invariant to changes in the background input-level. In Fig. 3-4A, we illustrate this property by using luminance inputs. We show the response of the normalization stage to changes in background luminance ranging from mesopic ($\sim 10^{-3} - 10^{1}$) to photopic ($\sim 10^{1} - 10^{8}$) conditions (Stockman & Sharpe, 2006). In this work, we focus primarily on photopic conditions. Except for the lowest luminance value evaluated (0.1 cd m^{-2} which corresponds to mesopic conditions), the normalized response remains proportional to *contrast* and invariant to *background luminance*. As an illustration, we used a two-dimensional image and convolved it with the 2D version of kernel *K* and then we proceeded to compute the steady-state response of Eqn. (8)[4]. Fig. 3-4C shows the original image and its steady-state representation at the output of the contrast-normalization stage. The contrast-dependent (i.e., normalized) response of a

[4] Note that this is the only simulation where we used a two-dimensinal input and a steady-state approximation. For the rest of simulations in this manuscript we used one-dimensional space and numerically solved the differential equations.

one-dimensional cut of the output image is shown in Fig. 3-4D (up). Finally, Fig. 3-4D (down) shows the normalized average pixel output of the image evaluated at different background-luminance values.

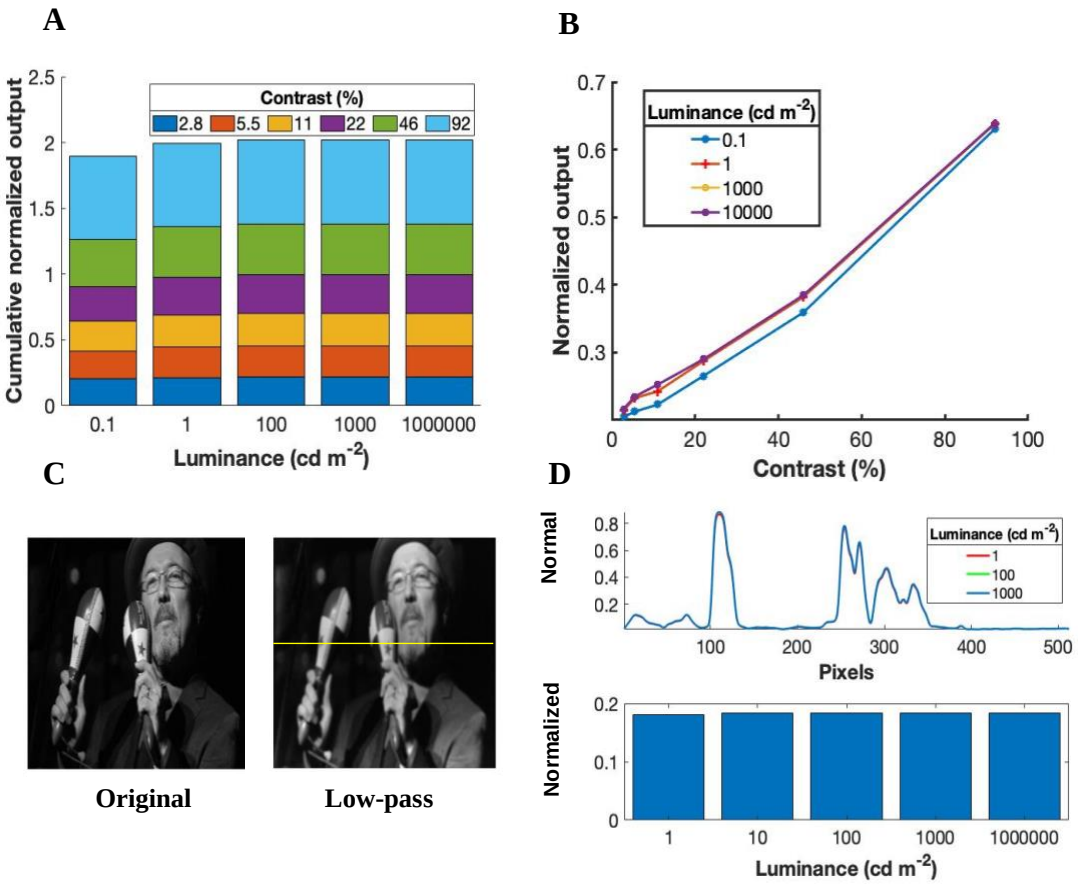

Fig. 3-4. Normalized response of the network as a function of contrast for different luminance levels. **A** The network is mostly invariant to luminance changes varying in orders of magnitude. **B** Normalized output activity as a function of contrast for multiple luminance values. **C** Original image (left) and the contrast-normalized output image after convolution with the 2-D version of kernel K (right). **D (upper)** One dimensional cut of the input image after being convolved with the 2D version of kernel K (see yellow line in subpanel c) and evaluated at multiple luminance level. **(lower)** Average normalized activity of the pixels. Note that the result is mostly the same regardless of background luminance. Photo source: Jeffrey M. Boan, Miami Herald file.

3.4.2. Motion Discrimination as a Function of Contrast and Spatial Size

We simulated data from three sources (Peñaloza et al., 2020; Tadin et al., 2003, 2019). Using a drifting Gabor, Tadin and colleagues (Tadin et al., 2003) demonstrated that

increasing the spatial size of a high-contrast stimulus renders the discrimination of motion-direction more difficult. The opposite effect occurred with a low-contrast stimulus. The authors proposed that this is a perceptual correlate of a center-surround antagonism, whereby high contrast objects are processed by a *spatial suppressive mechanism* and low-contrast sensory inputs are processed by a *spatial summation mechanism*. In other words, the interaction between size and contrast may be explained by an *adaptive center-surround mechanism*. Peñaloza et al. (2020) study is similar to Tadin et al. (2003) with the exception that they used a moving reference-frame for motion.

Fig. 3-5 shows Tadin et al. (2003) and Peñaloza et al. (2020) data as well as the simulation results. In our simulations, the input stimulus is taken to represent motion strength. We used rectangular signals that were varied in width and height, which represented changes in the signal's spatial-size and contrast, respectively. The y-axis, in Fig. 3-5A, represents phase-shift threshold, i.e., the amount of spatial displacement in terms of phase-angle[5] required to discriminate motion direction of a drifting Gabor, and the x-axis represents the stimulus width. There is a good qualitative match between the model (solid lines) and the experimental data (dashed lines). For the low contrast condition, the phase-shift threshold decreases with an increase in stimulus size (the smaller the phase-shift required by observers to reach a given threshold, the better the performance in discriminating the direction of motion). This result signals the presence of a spatial summation mechanism. On the other hand, the performance worsens with stimulus size for

[5] For example, for a Gabor with carrier frequency equal to 0.5 cycle/deg, a spatial shift of 0.1 deg of visual angle corresponds to an equivalent phase shift of 0.05x360 = 18 deg of phase angle.

the high-contrast condition, which is a signature of a spatial suppression mechanism. As seen in Fig. 3-5A our model is capable of predicting this behavior.

Our *linking hypothesis* was that the neuron's membrane potential at steady-state is inversely proportional to the information needed by the observer to reach a performance criterion. For example, if the neuron's potential represents the direction of motion, the higher the activity is, the more likely it is for the observer to report the correct direction of motion. In other words, the higher the membrane potential, the more information the observer has to reach a correct decision that, in turn, translates itself into a lower threshold ($Thr = 1/y$). For low-contrast signals, as the size of the stimulus increases, the model generates an increase in membrane potential in neuron *y* (Fig. 3-3C). The opposite effect occurs at high-contrast values. This is explained by the fact that the inhibitory input to the neuron increases as a function of size *and* contrast. This is evident in Fig. 3-5B, where the slope of the line relating normalized inhibitory input with size is flat at the lowest contrast and it starts to gradually increase with contrast values. In other words, at low-contrast levels, as the stimulus size increases, the neuron's membrane potential is mainly modulated by the excitatory input. Whereas at high-contrast levels, the contribution of the inhibitory input becomes more active as size increases. As a consequence, the neuron's membrane potential decreases with increases in stimulus size, resulting then in a subsequent increase in the performance threshold. In a similar way, Fig. 3-5C shows the ratio of the excitatory term (the term multiplying $(B' - y_i)$ in Eqn. (11) represented as Ex) over the inhibitory term (the term multiplying $(D' + y_i)$ in Eqn. (11) represented as Inh) as a function of size for the lowest and highest stimulus contrasts. The ratio decays as a function of size for the high-contrast condition but saturates for the low contrast condition. This means that the

inhibitory action is more prevalent under stimulus conditions with large and high-contrast motion signals.

These dynamic interactions between excitatory and inhibitory inputs, that in turn modulate the membrane potential of the neuron, result from the multiplicative interactions in the model. This can be better seen from the steady state of a given neuron y_i. As in Eqn (9), this steady-state response can be described by:

$$y_i = \frac{B * Ex - D * Inh}{A + Ex + Inh} \qquad (17)$$

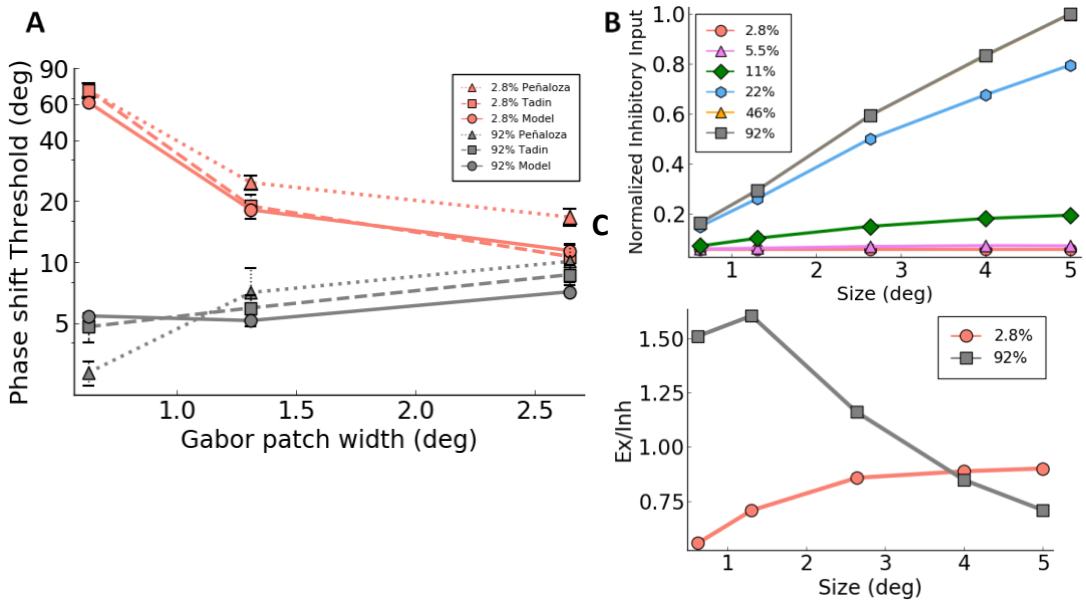

Fig. 3-5. Modeling behavioral data. A Simulation results (solid lines/circle marker) of Tadin et al., (2003) (dashed line/square marker) and Peñaloza et al., (2020) (dotted line/triangle marker) data. The plot represents the phase-shift threshold to discriminate motion direction as a function of size and contrast. Error bars in experimental data represent ± S.E.M. **B** Normalized inhibitory input to the output neuron (y) as a function of size. Different curves represent the different contrast levels. Notice that the slope of the curves gradually increases with contrast. **C** Excitatory (Ex) over inhibitory (Inh) input ratio as a function of size. The lowest and highest contrast values are plotted. At high contrast, the ratio decreases with size which reflects the fact that the excitatory signal saturates whereas the inhibitory signal continues to grow.

Taking into consideration the fact that *B* and *D* are non-negative constants, the inhibitory input to the neuron has both a subtractive (numerator with D > 0) and divisive (denominator) component. Notice that as the inhibitory term becomes more active with increases in size and contrast, the resulting neuronal response starts to be more modulated by inhibition than by excitation. This adaptive transition from excitatory to inhibitory modulation as a function of size and contrast results from multiplicative membrane dynamics and static nonlinearities embedded in the equation. We suggest that these neuronal computations underlie the behavioral transition from spatial summation to spatial suppression as observed in (Glasser & Tadin, 2010; Kwon et al., 2015; Tadin, 2015; Tadin et al., 2003, 2019, 2011).

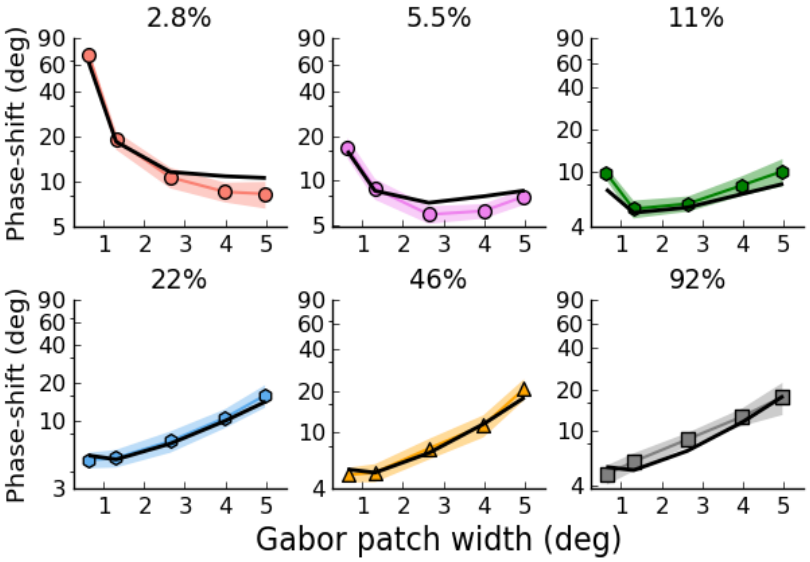

Fig. 3-6. Simulation results of Tadin et al. (2003) data. Filled symbols show Tadin's data. The black solid lines show the model's results. Colored shaded areas represent ± S.E.M. in the experimental data.

Fig. 3-6 shows the breakdown of all contrast values and stimulus sizes evaluated in Tadin et al. (2003; Fig. 3a) along with the model's predictions. In general, the model's

output agrees with the experimental data $(\chi^2(22) = 5.76, p > 0.99)$, signaling the progression from spatial integration to spatial suppression at all contrast stages.

In the literature, there are models (Betts et al., 2012; Tadin & Lappin, 2005) that capture the transition from spatial summation to spatial suppression (see Appendix A.4 for a detailed comparison). Our neural architecture is in agreement with these models in that they are based on center-surround interactions with different contrast response functions for the excitatory center and the inhibitory surround. However, our neural model differs from these models in three fundamental ways. First, these models are not "autonomous" in that the modeler calculates stimulus contrast and enters manually this value to the model equation. Our model extracts contrast automatically (Grossberg, 1973, 1988) according to equations derived from realistic models of neurons. Second, these models lack spatial representations. As we will discuss in the next section, spatial representations in the model make predictions that allow us to explain motion integration/segregation data that are not explained by these models. Third, other models calculate the response strength using either subtractive inhibition (Tadin & Lappin, 2005) or divisive inhibition (Betts et al., 2012). In contrast, our model combines both subtractive and divisive inhibition and, importantly, these properties result from the multiplicative equation whose asymptotic response has an inhibitory term in the numerator *and* the denominator (Eqn. (4)). In the literature, there is no consensus on the exact mechanisms used by neurons to implement surround suppression, with some favoring subtractive inhibition (Sceniak et. al., 2001), divisive inhibition (Cavanaugh et al., 2002), or both (Ayaz & Chance, 2009). Indeed, it may be difficult to distinguish between *in vivo* neurons (Abbott & Chance, 2005). Fourth, whereas previous models characterize the steady-state properties of the system and assume perfect

integration of evidence to reach a decision criteria, the solutions of our model and its associated multiplicative equations characterize the entire time course of the response (e.g., Smith et al., 2015). That is, it includes the dynamics of the response. Finally, our model illustrates the importance of system-level design wherein one can obtain a rich repertoire of behaviors from the nonlinear combination and integration of various canonical building blocks, much like in electronic circuit design. Our model shows how subtle parametric choices, for example the choice of the nonlinearities, can generate flexible and adaptive solutions for extracting information under varying stimulus conditions.

3.4.3. Temporal Integration

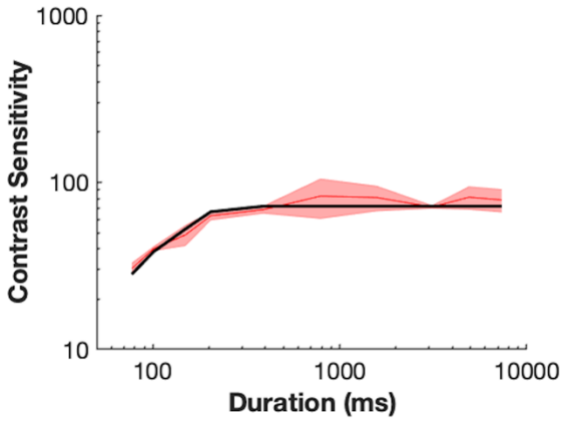

Fig. 3-7. Contrast sensitivity as a function of stimulus duration for direction discrimination of optic flow translation. The salmon line represents the experimental data (Burr & Santoro, 2001) and the black solid line shows the model's response. Colored shaded area represents ± S.E.M. in the experimental data.

To illustrate the dynamics of the model, we simulated data from Burr & Santoro (2001) in which they investigated the temporal summation of optic flow signals. Note that, since our model does not have a motion-detector front-end, the optical flow stimulus was represented in a way similar to the inputs used in the previous simulations, i.e., by rectangular inputs reflecting the strength of optical-flow motion signals. We simulated data

on contrast sensitivity for direction-discrimination of translational optic-flow motion as a function of exposure duration. To do so, we first simulated the shortest stimulus duration with a motion-contrast value close to the experimental results at that exposure duration. After that, we integrated the resulting membrane potential and established a fixed threshold. We took this fixed threshold to represent the neural activity required to achieve the performance level underlying the contrast thresholds showed in the experimental data. We then simulated the various stimulus durations and found the input contrast values generating the integrated membrane-potential values that reach the established threshold before stimulus termination. The model's parameters were identical to the previous simulations. The results of the simulation (solid black line) are shown in Fig. 3-7. Both the data and the model indicate reduced sensitivity for exposure durations shorter than 200 ms. These dynamical properties may explain, on the one hand, the durations of fixations when high-sensitivity is required by task demands, and on the other hand, the reduced sensitivity during brief fixations (less than 200 ms) when the task demands lead to rapid scanning of the scene.

3.4.4. Motion Integration and Segregation as a Function of Contrast

We next asked if the model can explain related but more complex data. Tadin et al. (2019) reported converging correlational and causal evidence showing that the visual system is remarkably effective at motion-based figure-ground segregation under stimulus conditions that also favor spatial suppression.

Fig. 3-8A and 3-8B show an illustration of the stimuli used in Tadin et al. (2019) for motion discrimination and motion segregation tasks, respectively.

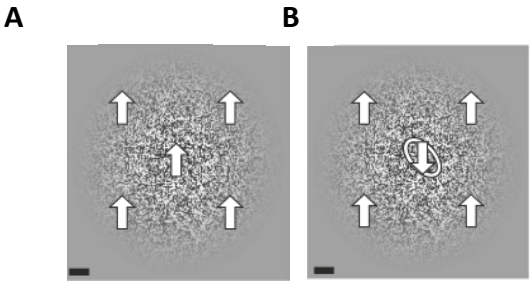

Fig. 3-8. Stimuli and behavioral tasks in Tadin et al. (2019) A motion discrimination task. Observers were asked to identify the motion direction of the moving background. **B** Motion segregation task. Observers were instructed to identify the tilt (left or right) of an oval motion -defined object presented on a moving background. Scale bar is 1°.

In the motion discrimination task (Fig. 3-8A), observers were asked to report the direction of motion (up or down) of a large random-texture pattern. The dependent variable was the stimulus duration (ms) needed to reach a predetermined performance threshold. The stimulus-duration threshold was evaluated at different contrast values. Likewise, in the motion segregation task (Fig. 3-8B), observers were instructed to identify the tilt of an oval motion-defined object presented on a moving background. At all times, the oval motion-defined object moved in the opposite direction to the moving background. Once again, the dependent variable was the stimulus duration (ms) needed to reach a predetermined performance threshold. The stimulus duration threshold was also assessed at varying contrast levels.

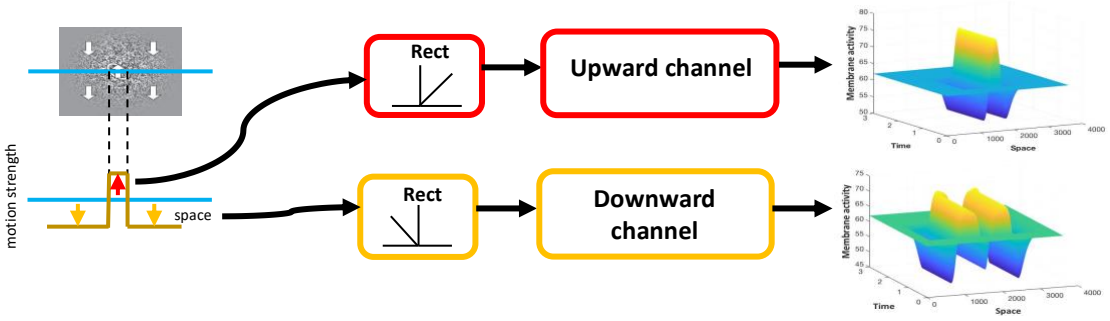

Fig. 3-9. Modified network to account for the motion segregation data in Tadin et al. (2019). Red and orange colors represent opposite motion-direction instances of our feed-forward model, upward and downward motions in this case. The motion strength input was rectified and passed to the corresponding motion channel. The spatiotemporal output of the network is presented in the rightmost panels.

Since segregation in this experiment is based on two different motion signals (up and down), we need one motion system tuned to upwards motion and one tuned to downwards motion. Thus, we simulated this experiment with two instances of our feed-forward model, with one instance for each motion direction. A schematic of the modified network is shown in Fig. 3-9. The colors red and orange represent the upward-motion and downward-motion channels, respectively.

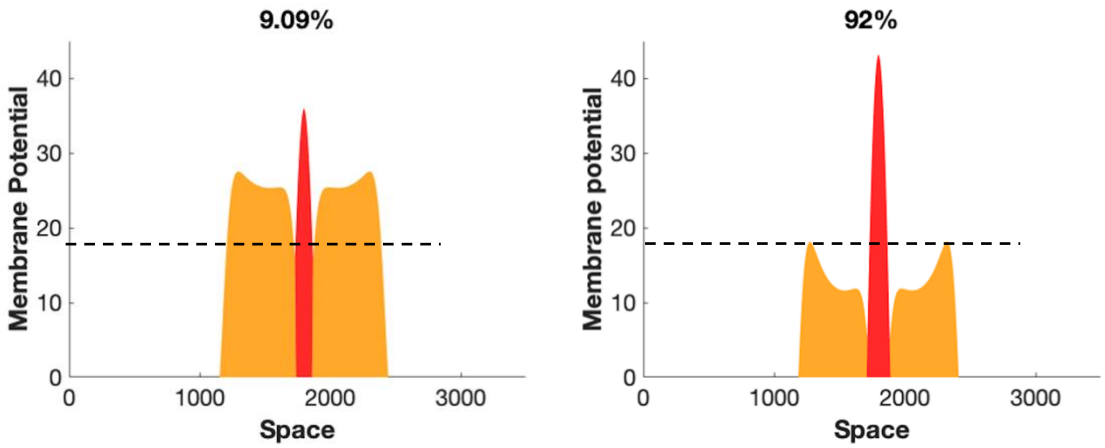

Fig. 3-10. Superimposed outputs of the motion channels. Red and orange colors represent the activities of opposite motion-direction channels, as shown in Fig. 9. The panel on the left shows the membrane potential of both channels for a low-contrast input. The panel on the right shows the membrane potential of both channels for a high-contrast input. As one can see, at high contrast, the activity corresponding to the center region (upward motion) increases whereas the activity corresponding to the surround region (downward motion) decreases. Consider how the center and surround activities may be segregated from each other by taking a reference activity level indicated by the dashed horizontal line. The spatial separation between the motion signals increases with contrast. We took this separation (evolving over time) to reflect the visual's system ability to segregate motion-defined objects from their moving backgrounds. In other words, the larger the spatial separation between the channels the shorter the time required to segregate the motion signals.

The input to the network was again represented as contrast-dependent motion strength. This time however, the signal's direction of motion was represented by the polarity of the input. The upward motion was assigned to the positive polarity and the downward motion to the negative (red and orange arrows in Fig. 3-9). In other words, higher magnitudes of input represented higher motion strengths and the sign of the input represented the direction

of motion. The motion-strength input was subsequently rectified and feedforwarded to the independent motion-direction channels so that each channel received positive-valued inputs corresponding only to their preferred direction. The output signal for the two channels is shown in space-time coordinates in the rightmost part of Fig. 3-9. Fig. 3-10 shows model activities across space for the central upward-motion (red) and the surrounding downward-motion (orange) for two contrast levels (left and right panels). As one can see from the figure, an increase in contrast leads to *(i)* an increase of the central region's activity, *(ii)* a decrease of the surround region's activity, and *(iii)* a more pronounced *spatial segregation* between the center and the surround region activities. After observing the increase spatial segregation at high-contrast in model activities, we asked whether this property can explain Tadin and colleagues data. As a linking hypothesis, we assumed that the spatial separation between the two regions reflects the observers' ability to segregate the two regions. Fig. 3-11 shows the model's fits for Tadin et al. (2019) motion discrimination and motion segregation tasks ($\chi^2(10) = 1.07, p > 0.99$). To simulate the motion discrimination data (Fig. 3-11A salmon line), we measured the simulation time required for the network to reach a fixed activity level at varying contrast intensities. As expected for a large-size motion signal, we observed an increase in duration-threshold with contrast. Our simulation results (black solid line) fit the experimental data within one standard error of the mean.

Fig. 3-11B shows the simulation results for the motion segregation task in Tadin et al. (2019) (salmon line). As discussed before, after superimposing the outputs of the two motion-direction channels, we observed that the spatial separation between the channels increased with contrast. We assumed that this separation reflects the visual system's ability

to segregate motion-defined objects from their moving backgrounds. Thus, we proceeded to measure how this spatial separation between the channels evolved over time, and then measured the time needed to reach a predetermined activity level. As a result, we obtained a decrease in duration threshold as a function of contrast that matched the data very well as shown in Fig. 3-11B (solid line). This effect is explained by the same size- and contrast-dependent interactions introduced by the network. That is, for a large-size motion background the network activity decreases with increments in contrast whereas for small-size motion stimuli (as the central stimulus in the simulation) the response of the network increases or remains mostly constant with contrast. Therefore, at a fixed activity level the spatial separation between the two channels tends to increase with contrast.

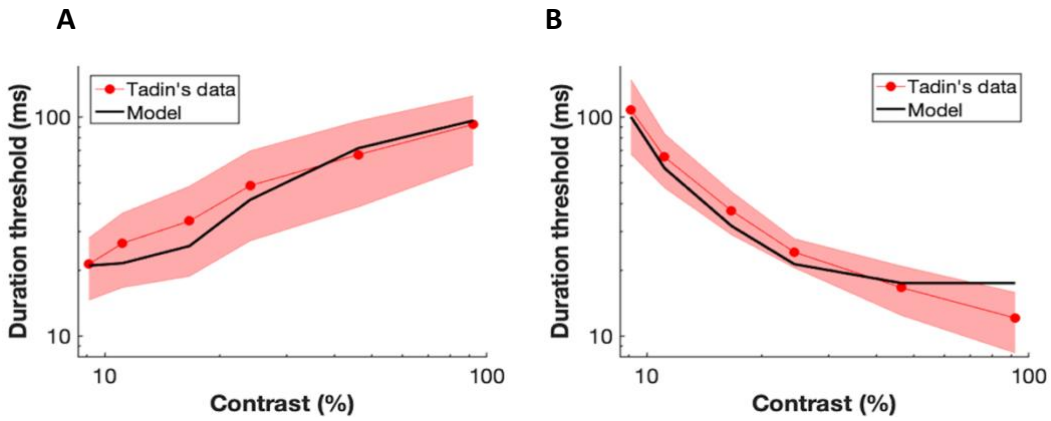

Fig. 3-11. Simulating motion discrimination and motion segregation with the same model. a Data from the motion discrimination task in Tadin et al. (2019) (salmon color) alongside with the simulation result (black solid line). The simulation results stay within one S.E.M. of the experimental data. **b** Experimental results of the motion segregation task (salmon line) and the simulation output (black solid line). The model predicts an increase in performance with contrast in a motion segregation task. Shaded areas in experimental data represent ± S.E.M.

Taken together our results show that our model is capable of capturing the size- and contrast-dependent interactions observed in behavioral data by means of the relationships

between the center and surround receptive fields established through multiplicative neural dynamics.

3.4.5. Simulation of Neural Data

Varying the size of drifting sinusoidal gratings, Sceniak et al. (1999) found that the extent of spatial summation in macaque V1 neurons is contrast dependent. Specifically, they observed that the extent of spatial summation was on average 2.3-fold greater at low stimulus contrast. We tested this effect with our model and the results are presented in Fig. 3-12. In the experimental data, there were neurons showing response suppression while others showed no surround suppression. Keeping the same model's parameters, in Fig. 3-12A we show our simulation results for the surround-suppressed neurons. As seen, the optimal stimulus radius for spatial summation —measured as the maximum peak response (black arrows)— shrank at high stimulus contrast (28%) compared to low stimulus contrast (10%). This is because, for small stimuli, the inhibitory contribution is small (see for example Fig. 3-5C) and the response of the model is mainly modulated by the contrast-controlled excitatory center. Thus, higher stimulus contrasts produce larger responses which, in turn, yields the earlier peak response at 28%.

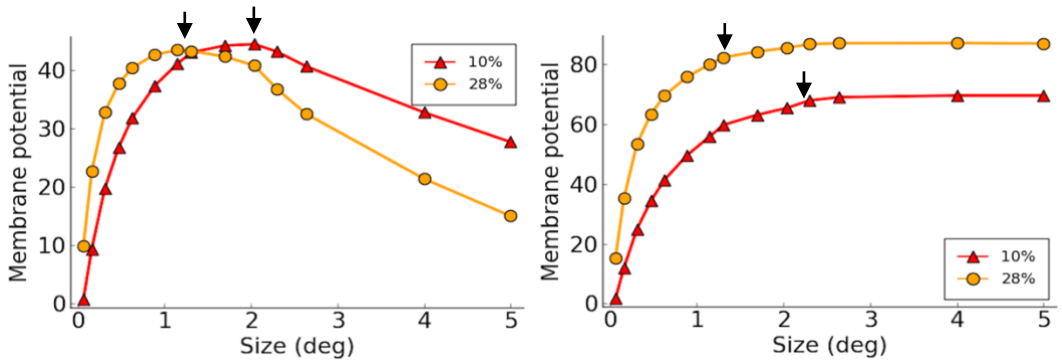

Fig. 3-12. Simulating the effects of contrast on spatial summation of V1 neurons. Using similar contrast values and stimulus sizes as in Sceniak et al. (1999), we simulated the effect of contrast on spatial summation.

To simulate neurons with no surround suppression, we replaced the inhibitory non-linearity with a constant value. Taking the optimal stimulus radius for spatial summation as the radius at which responses first reached 95% of the maximum, in Fig. 3-12B, we

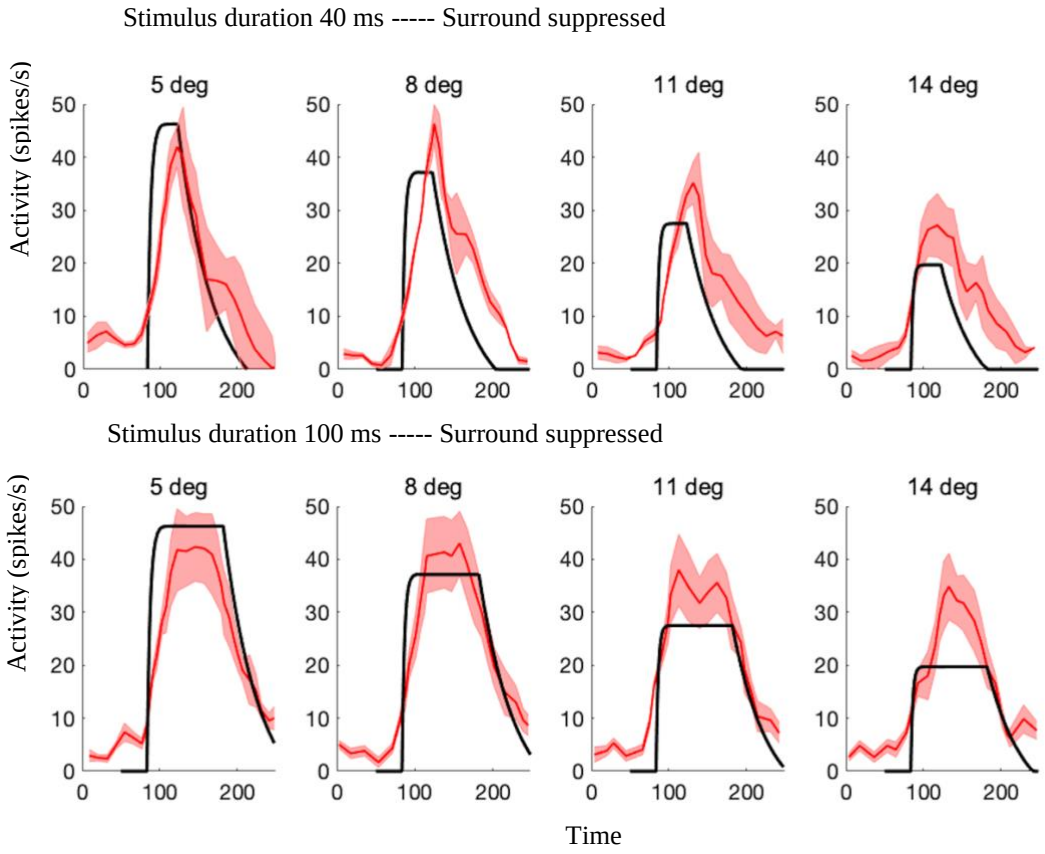

Fig. 3-13. Simulation of the neural activities of surround-suppressed neurons in MT. The upper and lower panel show the simulation results (black lines) for the activities of surround-suppressed MT neurons (salmon color) evaluated at four stimulus sizes when the stimulus presentation time was 40 and 100 ms, respectively. Shaded areas in experimental data represent ± S.E.M.

showed that optimal stimulus radius (black arrows) decreased with contrast too. Similar reasons as in the previous case apply for this condition. Hence, regardless of surround condition, our model predicts that the extent of spatial summation decreases as stimulus contrast increases.

We also model neural data from MT neurons. Particularly, Churan et al. (2008) recorded the activities of surround-suppressed and non-surround-suppressed MT neurons responding to the preferred and anti-preferred directions. They stimulated both types of neurons for a brief (40 ms) and a longer (100 ms) stimulus presentation time. Using the same model's parameters except for a change in the time-scaling parameter (τ in Eqn. (11)) from 1 to 0.00375 (which results in a more prolonged decay of the response), we simulated the data of neurons responding to the preferred directions, evaluating the four stimulus sizes and the two presentation times. Our results showed similar response profiles

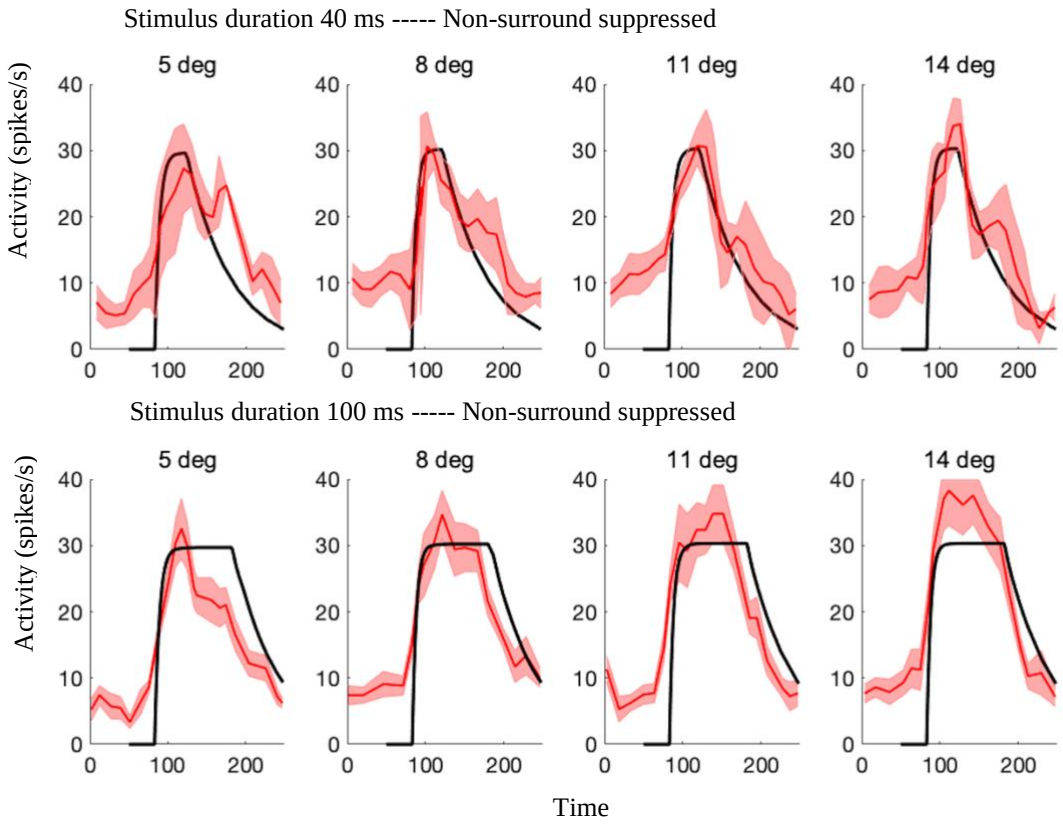

Fig. 3-14. Simulation of the neural activities of non-surround-suppressed neurons in MT. The upper and lower panel show the simulation results (black lines) for the activities of non-surround-suppressed MT neurons (samon lines) evaluated at four stimulus sizes when the stimulus presentation time was 40 and 100 ms, respectively. Shaded areas in experimental data represent ± S.E.M.

as the neural data (Fig. 3-13). For instance, there is the expected drop in neural activity due to the increased inhibition that comes as a product of an increase in stimulus size and the sharp response at the onset of the stimulus accompanied by the subsequent prolonged decay in the response after the stimulus offset.

To simulate the non-surround-suppressed neurons, we modified the model by eliminating the surround inhibition term and changing the passive decay parameter (from 1 to 100) of Eq. 11. After scaling the simulation results (with a constant value of 0.52), we obtained the fits shown in Fig. 3-14, which demonstrates that the modified model can capture the response profile of non-surround-suppressed MT neurons.

3.5. Discussion

In this work, we show how a class of biologically-plausible systems of differential equations known as multiplicative equations can account for the visual system's ability to handle visual inputs of diverse dynamic ranges as those found in ecological vision. The key canonical computational elements in our model are multiplicative interactions, center-surround receptive-fields, and sigmoidal nonlinearities. All these elements have strong neurophysiological support. Building upon these mechanisms, we showed how the brain can deploy adaptive spatial-integration strategies in order to cope with the dynamic range of its inputs. Additionally, we showed that the same mechanisms can also account for both motion discrimination and motion segregation data.

Balancing the demands between segregation and integration processes is a challenging problem that is widespread in biological systems (Braddick, 1993). Visual segregation of moving objects is a good example of this interaction. Our model supports the recently reported evidence for the hypothesis that spatial suppression of background motion signals

is critical for rapid segmentation of moving objects (Tadin et al., 2019). For motion signals that favor spatial suppression (i.e., large high-contrast stimuli) the model predicts short duration-thresholds for segregation and large duration-threshold for discrimination. On the other hand, for motion signals advantaging spatial integration (i.e., large low-contrast stimuli) the network predicts the opposite effect. In other words, our neural model provides an adaptive mechanism between the competing segregation and integration strategies, which are ubiquitous in processing environmental stimuli.

The center-surround anatomy embedded in our architecture is in line with the longstanding hypothesis that highlights the fundamental role of antagonistic center-surround mechanisms in the segmentation of moving objects from their backgrounds (Nakayama & Loomis, 1974; Petkov & Subramanian, 2007). The network is sensitive to small motion signals falling within the center receptive field, and as the stimulus is enlarged to reach the receptive field surround, the neural response is suppressed. This property is prominently found in the motion processing area MT (Schallmo et al., 2018). Therefore, we suggest MT as the possible loci for our proposed adaptive mechanisms of sensory processing. Two lines of imaging studies and one of transcranial magnetic stimulation (TMS) support this hypothesis: First, using an experimental paradigm similar to that of Tadin et al. (2003), Er et al. (2020) showed in an fMRI study that responses in hMT+, but not those in V1, correlated with the perceptual effect. As mentioned before, Peñaloza et al. (2020) conducted a psychophysical study similar to Tadin et al. (2003), with the exception that they used a non-retinotopic rather than a retinotopic reference frame used in Tadin et al. (2003) and Er et al. (2020). Optics of the eyes project neighboring points in the environment to neighboring points on the retina, and these neighborhood relations are

preserved in early visual areas (Sereno et al., 1995). This property is known as retinotopy. However, our perception of the world is not based on a retinotopic reference frame: for example, during steady fixation, a reflector placed on the wheel of a bicycle undergoes a cycloidal motion on retinotopic coordinates. However, we perceive the reflector *in rotation* because the visual system uses a non-retinotopic reference frame following the motion path of the bicycle. Using the Ternus-Pikler display paradigm (Boi et al., 2009), Peñaloza et al. (2020) tested whether Tadin et al.'s (2003) findings generalized to motion in non-retinotopic coordinates. Their findings suggested that similar adaptive center-surround mechanisms also mediate the non-retinotopic processes underlying motion perception. Using 7T fMRI, Thunell et al. (2016) showed that activations in areas V1, V2, and V3 reflected retinotopic properties of motion. On the other hand, activations in the area hMT+ reflected both retinotopic and non-retinotopic properties of motion. Hence, visual area hMT+ appears to be the first visual area encoding non-retinotopic percepts. Finally, using TMS, Tadin et al. (2011) tested the hypothesis that disruption of MT/V5 should weaken spatial suppression and, therefore, improve motion discrimination of large moving patterns. Their findings support this hypothesis, suggesting that area MT/V5 is causally involved in spatial suppression, possibly by debilitating surround suppression strength. Given the similar behavioral findings for adaptive processing of motion in retinotopic (Tadin et al., 2003) and non-retinotopic (Peñaloza et al., 2020) coordinates and in the light of Er et al.'s (2020) and Thunell et al.'s (2016) fMRI findings, visual area hMT+ is a good candidate for the locus of adaptive properties in processing motion.

We would also like to highlight the limitations of our model. For simplicity, we did not include a front-end consisting of motion detectors. Instead, we directly fed the motion

signals to our model through a simple input representation. One purpose of this was to present the adaptive mechanisms as general mechanisms that can be applied to motion and other types of inputs such as luminance and color or to any processing hierarchy that would benefit from these adaptive properties. In fact, we demonstrated contrast normalization by using inputs representing luminance, whereas the rest by using inputs representing motion. However, the model can be more complete by incorporating directionally selective motion-detectors at the front end. Moreover, whether similar properties hold for other stimulus dimensions, such as color, texture, needs to be tested experimentally. Furthermore, extending the model into two spatial dimensions can provide a better representation of visual inputs and allow simulations of more complex stimuli. However, such an extension significantly increases the computation time required to solve the resulting system of differential equations numerically.

Chapter 4. Effects of Spatial Attention on Spatial and Temporal Acuity: A Computational Account

4.1. Introduction

At any moment of our daily life, the amount of visual information that impinges upon our retina is enormous. However, the brain has limited capacity for processing visual information and thus is forced to select information that is most relevant for our cognitive, emotional, and behavioral purposes (e.g., Desimone & Duncan, 1995; Moore & Zirnsak, 2017; Posner & Rothbart, 2007; Reynolds & Chelazzi, 2004). Ample neuropsychological, neuroimaging and behavioral evidence supports the idea that stimuli compete for limited brain resources (e.g., Carrasco, 2011; Desimone & Duncan, 1995; Kinchla, 1992; Reynolds & Chelazzi, 2004). For instance, the biased-competition hypothesis states that neurons whose receptive fields lay at the attended location become more active than surrounding neurons whose responses are suppressed (Desimone & Duncan, 1995). Thus, attention is a selective process that allows the brain to optimize its information-processing capabilities by enhancing the most relevant information while suppressing the less relevant information. In this context, the role of covert spatial attention, i.e., the focusing attention on a specific spatial location in the visual field without eye movements, on perception has been extensively studied (for a review see, Carrasco, 2011). Specifically, in a series of studies, Yeshurun & Carrasco (1998, 1999) and Carrasco et al. (2002) found that spatial attention improves spatial resolution, i.e., performance in both acuity and hyperacuity tasks

improves when a target stimulus appears at an attended location (signaled by a briefly presented peripheral spatial precue). Moreover, covert spatial attention also affects perception in the temporal domain. For example, Yeshurun & Levy, 2003 found the paradoxical result that covert spatial attention *impairs* the detection of brief temporal gaps between successively presented stimuli. In other words, spatial attention deteriorates temporal resolution (Yeshurun, 2004; Yeshurun & Levy, 2003).

If the role of attention is to select and enhance stimuli, why does it impair temporal acuity when the task demands require temporal segregation of stimuli? Different attentional models, e.g., based on decision criteria (e.g., Kinchla et al., 1995) and noise-reduction approaches (e.g., Lu & Dosher, 1998), have been proposed to explain this phenomenon. However, due to experimental design considerations and task demands, these models do not provide a suitable explanation for the effects of spatial attention on spatiotemporal acuity (see the discussion section for a detailed explanation). However, a more plausible explanation is an attentional signal-enhancement mechanism based on interactions between the parvocellular and magnocellular systems (Ogmen, 2005; Yeshurun, 2004; Yeshurun & Levy, 2003; Yeshurun & Sabo, 2012). The parvocellular and magnocellular systems possess different spatiotemporal sensitivities with the parvocellular system being more sensitive to high spatial frequencies and the magnocellular system to high temporal frequencies (Breitmeyer & Ganz, 1977; Keesey, 1972; Kulikowski & Tolhurst, 1973). Thus, an attentional mechanism that *favors* the parvocellular system, and therefore, improves sensitivity to high spatial-frequencies can explain the enhancement in spatial resolution. A cortical competition between the parvocellular and magnocellular systems, in turn, can explain why an enhancement of parvocellular activity by attention can lead to

a decrease in magnocellular activity, which in turn results in reduced temporal acuity (Ogmen, 2005; Yeshurun, 2004; Yeshurun & Levy, 2003; Yeshurun & Sabo, 2012).

Even though the conceptual predictions of this parvo-magno attentional model have been proposed (e.g., Yeshurun & Levy, 2003; Ogmen, 2005), these predictions have not been computationally tested. Therefore, in this work, our goal was to test the prediction that a computational model of retino-cortical dynamics (RECOD model) that incorporates interactions between sustained (parvo) and transient (magno) systems (Öğmen, 1993; Ogmen et al., 2003) can account for the aforementioned effects of attention on spatial and temporal acuity (Ogmen, 2005).

4.2. The general architecture of the model

The general structure of the RECOD model is shown in Fig. 4-1. The primate retina has at least 20 types of retinal ganglion cells (RGCs) which are distinguished by their size, shape, connectivity, neurochemistry or response properties (Grünert & Martin, 2020; Kaplan, 2013). However, the RGCs can be mainly categorized into two major groups: *midget* RGCs and *parasol* RGCs. The midget RGCs show sustained responses to maintained illumination, non-saturating contrast response, and small receptive fields (Lee et al., 2010). Parasol RGCs, on the other hand, present transient response to maintained illumination, saturating contrast-response function, and spread dendritic connections (Lee et al., 2010). Anatomically, the midget RGCs project to the parvocellular layer of the lateral geniculate nucleus (LGN) while the parasol RGCs synapse to the magnocellular layer of the LGN, forming what is known as the *parvocellular* and *magnocellular* streams. In Fig.4-1, the bottom two ellipses represent these two major pathways of RGCs. The top ellipses in Fig. 4-1 represent post-retinal areas that receive selective inputs from the magnocellular

and parvocellular projections. These cortical areas are thought to specialize in complementary visual functions, such as temporal (magnocellular pathway), and spatial (parvocellular pathway) aspects of visual processing.

The RECOD model postulates interactions within and between these pathways. *Intra-channel inhibition* refers to within channel interactions produced by on-center off-surround organization of the receptive fields (represented in Fig. 4-1 by the open and filled synapses, respectively). In addition to feed-forward intra-channel interactions, the model has also feedback (recurrent, re-entrant) interactions at post-retinal levels in the parvocellular pathway, as depicted by the recurrent connections in Fig. 4-1. Furthermore, the model incorporates interactions at cortical levels between parvocellular and magnocellular streams.

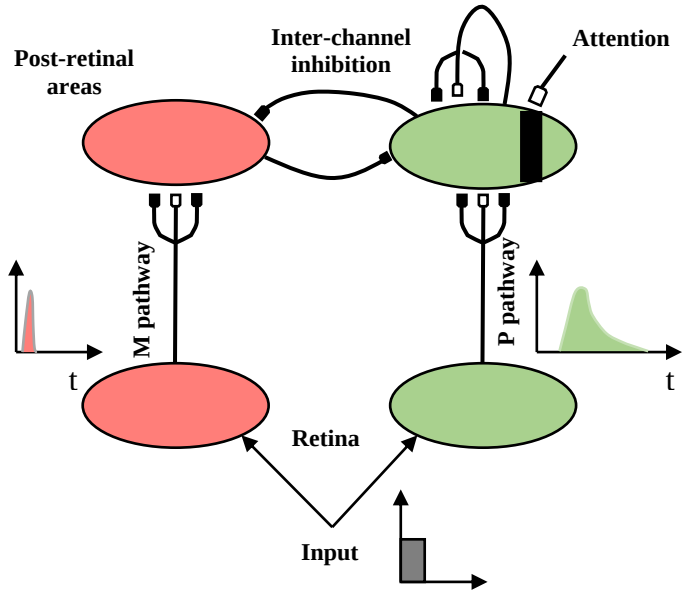

Fig. 4-1. Schematic representation of the RECOD model. The bottom two ellipses represent the retinal magnocellular (parasol, red) and parvocellular (midget, green) retinal ganglion cells. The red and green upper ellipses are the post-retinal magnocellular and parvocellular channels, respectively. The open and filled symbols represent excitatory and inhibitory connections. Spatial attention is included as a facilitatory signal to a specific spatial area in the parvocellular channel (depicted by the black region). The subplots next to each channel present an illustration of the response against time for each pathway.

In particular, it includes reciprocal inhibition (*inter-channel inhibition*) between the post-retinal cells driven by the magnocellular and parvocellular channels (Ogmen et al., 2003; Purushothaman et al., 2003).

One area where the RECOD model has been extensively tested is visual masking (Breitmeyer & Ogmen, 2000, 2006). The reciprocal inhibition between sustained and transient systems plays a key role in explaining type-B masking as well as for explaining the double dissociation between a stimulus' visibility and its effectiveness as a mask (Öğmen et al., 2006). According to RECOD, attention is not an essential component of visual masking (Agaoglu et al., 2016; Breitmeyer & Ogmen, 2000, 2006); rather attention plays a modulatory role and hence is introduced to the model as an "add-on" either as sub-cortical modulation (Breitmeyer et al., 2006; cf. Bachmann, 1984) or as a direct modulation of cortical activities (Ogmen, 2005). Fig. 4-1 shows the latter, i.e., attention signals directly enhancing activities in a specific spatial region (depicted by the black area in Fig. 4-1) of the post-retinal parvocellular neurons. As we will detail in the following sections, direct modulation of parvocellular activities along with the reciprocal inhibition between parvocellular and magnocellular systems make predictions consistent with the positive and negative effects of spatial attention on spatial and temporal acuity, respectively.

4.3. Mathematical description of the model

Here, we present a description of the equations governing the model. The neurophysiological bases of the equations can be found elsewhere (Breitmeyer & Ogmen, 2006; Öğmen, 1993; Ogmen et al., 2003).

4.3.1. Retinal cells with sustained activities

The first step in the model is to achieve temporal adaptation or gain control via (Carpenter & Grossberg, 1981; Ogmen & Gagne, 1990; Ögmen & Gagné, 1990):

$$\frac{1}{\tau}\frac{dz_i}{dt} = \alpha_s(\beta_s - z_i) - \gamma_s(J_s + I_i)z_i \quad , \qquad (1)$$

where z_i represents the concentration of a transducing agent at the *i*th spatial location, β_s is the maximum level of the agent, α_s and γ_s are its replenishment and depletion rates, J_s represents a dark current and I_i is the external input (luminance value) at the *i*th spatial location (Öğmen, 1993). In response to a step input, this equation generates a response that consists of an initial peak activity that is followed by a decay to a plateau. The parameter τ controls the time constant of response dynamics.

The second step consists of cells arranged in a center-surround organization of the form:

$$\frac{1}{\tau}\frac{dw_i}{dt} = -A_s w_i + (B_s - w_i)\sum_{j=i-n_{se}}^{i+n_{se}} G_{j-i}^{se}\left(J_s + I_j\right) z_j -$$

$$(D_s + w_i)\sum_{j=i-n_i}^{i+n_i} G_{j-i}^{si}(J_s + I_j)z_j \qquad (2)$$

The sustained membrane potential w_i is generated by convolving the output signals from the first stage, z_j, with Gaussian kernels representing the excitatory-center and inhibitory-surround receptive fields. The Gaussian kernels are functions of the form $G_k^s = Amp_s e^{\left(\frac{-k^2}{sd_s^2}\right)}$ with parameters Amp_s and sd_s representing the amplitude and spread of the kernel. Notice that Eq. 2 is a shunting equation (Grossberg, 1988), and its parameters [A_s,

B_s, D_s] represent a passive decay and the Nernst potentials for depolarization and hyperpolarization, respectively.

The last step consists of:

$$\frac{dv_i}{dt} = \sigma(-v_i + [w_i - \Gamma_s]^+) \qquad (3)$$

Table 4-1. Choice of parameter values for the sustained (subscript s) and transient (subscript t) instantiations of Eq. 1 and 4.

Parameter	Value
$\boldsymbol{\alpha_s}$	0.4
$\boldsymbol{\beta_s}$	16
$\boldsymbol{\gamma_s}$	0.13
$\boldsymbol{J_s}$	12
$\boldsymbol{\alpha_t}$	8.4
$\boldsymbol{\beta_t}$	40
$\boldsymbol{\gamma_t}$	1.53
$\boldsymbol{J_t}$	12
$\boldsymbol{\tau}$	1

This stage transforms the membrane potential, w_i, from the previous stage into an output signal via a nonlinearity with threshold. The function $[a]^+$, denotes half-wave rectification (also known as linear-above-threshold function, Rectified Linear Unit, ReLU). The parameter Γ_s represents the threshold level of the function. The parameter σ scales the temporal dynamics and determines the persistence of the sustained cell activity. The output of this equation is a parvocellular-sustained retinal response similar to the green sustained response illustrated on the right subplot of Fig 4-1.

4.3.2. Retinal cells with transient activities (magnocellular pathway)

Responses of retinal cells in the magnocellular pathway are generated with the same equations as in the parvocellular stream, namely:

$$\frac{1}{\tau}\frac{du_i}{dt} = \alpha_t(\beta_t - u_i) - \gamma_t(J_t + I_i)u_i \qquad (4)$$

$$\frac{1}{\tau}\frac{dy_i}{dt} = -A_t y_i + (B_t - y_i)\sum_{j=i-n_{te}}^{i+n_{te}} G_{j-i}^{te}\left(J_t + I_j\right)u_j -$$

$$(D_t + y_i)\sum_{j=i-n_i}^{i+n_i} G_{j-i}^{ti}(J_t + I_j)u_j \qquad (5)$$

The parameters' interpretation of Eq. 4 and 5 is the same as in Eq. 1 and 2. The only difference lies in the parameter choice (see Tables 4-1 and 4-2) and in the use of a half-way rectification operation of the form:

$$y_i^t = [y_i - \Gamma_t]^+ \qquad (6)$$

where y_i is the activity of the neuron at location *i* resulting from Eq. 5, and the fixed threshold level Γ_t is set to specifically select the initial overshoot of the response. The result of Eq. 6 is a transient response representing the output of the magnocellular stream as depicted by the red transient response in the left subplot of Fig. 4-1.

Table 4-2. Choice of parameters for the retinal network equations.

Parameter	Values
$\boldsymbol{A_s}$	2
$\boldsymbol{B_s}$	250
$\boldsymbol{D_s}$	10
$\boldsymbol{J_s}$	6
$\boldsymbol{G^{se}}$: $\boldsymbol{Amp}$	1
$\boldsymbol{G^{se}}$: $\boldsymbol{sd}$	28
$\boldsymbol{G^{si}}$: $\boldsymbol{Amp}$	0.00135
$\boldsymbol{G^{si}}$: $\boldsymbol{sd}$	94
$\boldsymbol{n_{se}}$	57
$\boldsymbol{n_{te}}$	80
$\boldsymbol{n_i}$	100
$\boldsymbol{\sigma}$	0.10
$\boldsymbol{\Gamma_s}$	249.62
$\boldsymbol{A_t}$	2
$\boldsymbol{B_t}$	600
$\boldsymbol{D_t}$	10
$\boldsymbol{J_t}$	6

$\boldsymbol{G^{te}}$: $\boldsymbol{Amp}$	1
$\boldsymbol{G^{te}}$: $\boldsymbol{sd}$	14
$\boldsymbol{G^{ti}}$: $\boldsymbol{Amp}$	0.00743
$\boldsymbol{G^{ti}}$: $\boldsymbol{sd}$	112
$\boldsymbol{\Gamma_t}$	587

4.3.3. Post-retinal network

4.3.3.1 Post-retinal sustained/parvocellular cells

The dynamics of post-retinal sustained/parvocellular cells is described by Eq. 7. The first term on the right-hand-side of Eq. 7 is a passive decay term. The second term on the right-hand-side corresponds to feedback excitatory signals, excitatory sustained inputs, and localized afferent sustained inputs modulated by the attentional signal (Att). Finally, the third term is composed of inhibitory feedback signals, afferent inhibitory signals coming from the retinal sustained cells, and transient-on-sustained inhibition from the transient channel (m). The Gaussian kernels G_k^{pi}, Q_k^{mi} determine the spread of the inhibitory afferent and the transient-on-sustained inhibitory signals, respectively. Spatial attention was represented by a rectangular signal of the form: $Att(x) = \Theta\ for\ |x| \leq 40; 0\ otherwise$. The parameter Θ represents the strength of the attentional signal and was used to fit the experimental data ($\Theta = 0.61$). The variable x represents the location in space where attention is focalized. We fixed the spatial spread of the attentional signal to a specific value wide enough to cover the cortical space where the sustained afferent inputs appeared. In this way, the same level of attention is provided regardless of the width of the input stimulus. When spatial attention is absent, the attentional signal, Att, is set to a value of zero. The parameter η represents the relative delay between the parvocellular and magnocellular signals.

$$\frac{dp_i}{dt} = -A_p p_i + (B_p - p_i)\left\{\Phi(p_i) + v_i(t-\eta) + \sum_{j=i-n_e}^{i+n_e} Att_{j-i} v_j(t-\eta)\right\}$$

$$- p_i \left\{ \sum_{j=i-n_{pf}}^{i+n_{pf}} \Phi(p_i) + \sum_{j=i-n_{pi}}^{i+n_{pi}} G_{j-i}^{pi}\, v_j(t-\eta) \right.$$

$$\left. + \sum_{j=i-n_{pi}}^{i+n_{pi}} Q_{j-i}^{mi} m_j \right\} \qquad (7)$$

The excitatory and inhibitory reentrant (feedback) signals are implemented through the nonlinear function $\Phi(x) = 10a\{(a+1)^2 - 1\}$, if $a < 0.05$ and $\Phi(x) = a(a + 0.975)$, otherwise. This function and its parameters have previously been chosen because they achieve sharpening of boundary signals for dynamic inputs (Breitmeyer et al., 2006; Öğmen, 1993; Ogmen et al., 2003).

4.3.3.2 Post-retinal transient/magnocellular cells

The dynamics of post-retinal transient/magnocellular cells is described by Eq. 8. Similar to the post-retinal sustained cells, the first term on the right-hand-side of Eq. 8 is a passive decay. The second term corresponds to excitatory inputs from the afferent magnocellular pathway. The third term between curly brackets in the equation is the inhibitory signal which consists of afferent inhibitory inputs from the magnocellular pathway and, importantly, of inter-channel inhibitory signals via the *sustained-on-transient interaction*. H_k^{mi} is a Gaussian kernel determining the spread of inhibition and parameters ε and ϵ are non-negative constants.

$$\frac{dm_i}{dt} = -A_m m_i + (B_m - m_i)\varepsilon \cdot y_i^t(t) - m_i \left\{ \sum_{j=i-n_t}^{i+n_t} H_{j-i}^{mi} y_j^t(t) + \epsilon \cdot p_i \right\} \qquad (8)$$

Table 4-3. Choice of parameters for the post-retinal network equations

Parameter	Values
$\boldsymbol{A_p}$	1
$\boldsymbol{B_p}$	1
$\boldsymbol{G^{pi}}$: $\boldsymbol{Amp}$	0.00135
$\boldsymbol{G^{pi}}$: $\boldsymbol{sd}$	94
$\boldsymbol{Q^{mi}}$: $\boldsymbol{Amp}$	1
$\boldsymbol{Q^{mi}}$: $\boldsymbol{sd}$	94
$\boldsymbol{n_e}$	20
$\boldsymbol{n_{pi}}$	100
$\boldsymbol{n_{pf}}$	57
$\boldsymbol{n_t}$	80
$\boldsymbol{\eta}$	2
$\boldsymbol{A_m}$	10
$\boldsymbol{B_m}$	1
$\boldsymbol{H^{mi}}$: $\boldsymbol{Amp}$	0.00743
$\boldsymbol{H^{mi}}$: $\boldsymbol{sd}$	112
$\boldsymbol{\varepsilon}$	0.6
$\boldsymbol{\epsilon}$	9

4.3.4. Simulation methods

The system of ordinary differential equations was numerically solved using the CVODE package (version 5.6.1, Dec 2020) from the SUNDIALS library (Hindmarsh et al., 2005). We used the backward-differentiation formula method, which is a classic method to solve stiff differential equations. However, solving large systems of ODEs can be time-consuming. Thus, we limited the model to only one spatial dimension with 580 neurons representing a spatially one-dimensional retinotopic map. For the spatial resolution simulations, the input to the model was represented as a one-dimensional cut of the 2-D stimulus used in Yeshurun et al. (1999) (see yellow 1-D cut in Fig. 4-2A). Observers were asked to report the location of the gap in this Landolt-square stimulus. A depiction of the retinal ON-channel and OFF-channel responses to the yellow 1-D cut is shown by the red and green pulse responses in Fig. 4-2A. For computational simplicity, in

this study, we took the pulse response to the gap size in the OFF-channel to represent the input stimulus of the model (green area in Fig. 4-2A). Our linking assumption was that the higher the cortical parvocellular-response was to this gap area, the better was the observer's ability to detect and report the location of the gap. Therefore, larger gap sizes produce more robust post-retinal parvocellular activities, which, in turn, result in better spatial resolution performance. To fit the neural responses to the data, we took the sum of the activities of the cortical parvocellular neurons responding to the gap area as representing the percent correct performance. Similarly, we represented reaction time (RT) as the model time required for the integrated response of a post-retinal sustained neuron to reach a pre-defined fixed activity level. Therefore, just as in the experimental data, as the neuronal response increases with gap size, the time to reach the pre-defined threshold decreases accordingly.

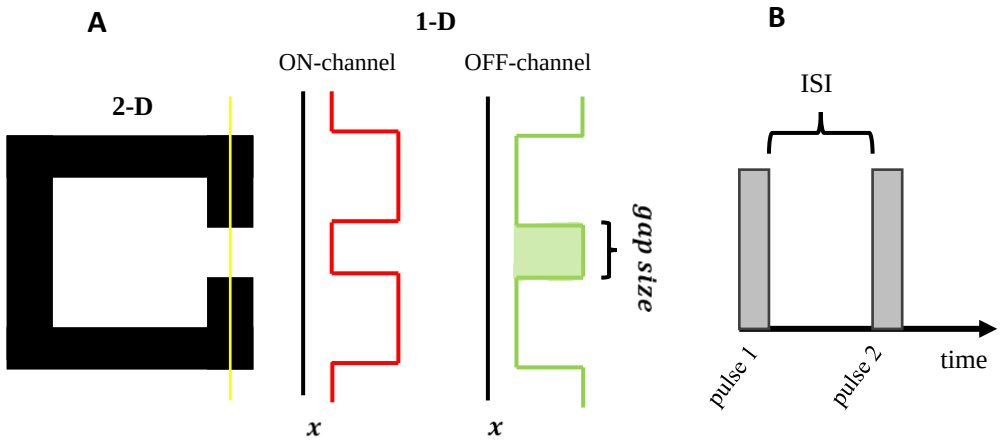

Fig. 4-2. Schematic of the input stimulus used in the simulations. A. The leftmost figure is the 2-D representation of the stimulus used to measure spatial resolution in Yeshurun et. al (1999). For computational simplicity we limit the simulations to one spatial dimension. The vertical yellow line represents a cut in one spatial dimension. The red pulsed signal in the middle of panel A expresses the response of the retinal ON-channel to the 1-D yellow cut. The rightmost sketch in the panel depicts the response of the retinal OFF-channel. We took the response of the OFF-channel to the gap size (shaded green area in the figure) as the input stimulus for the spatial resolution simulations. **B.** Representation of the input stimulus used for the simulation of the temporal resolution task. The time between the offset of the first pulse and the onset of the second pulse (ISI) was varied in this simulation.

Mathematically, the gap-size target stimulus was represented as a rectangular signal of the form: $I = 20\ for\ |x| \leq\ gap; 10\ otherwise$, where the variable x represents spatial location. The width of the input signal, I, was changed through the gap variable, which represents the different gap-size values (see Fig. 4-2A).

For the temporal resolution task, observers were presented with two brief pulses separated by an interstimulus interval (ISI). The corresponding input to the network is shown in Fig. 4-2B. Our model assumes that temporal resolution relies mainly on the magnocellular activity. Thus, to relate the simulation results to the experimental data in the temporal resolution task, we expressed performance as $pulse1_{magno\_peak} * pulse2_{magno\_peak} * t_s$, where $pulse1_{magno\_peak}$ and $pulse2_{magno\_peak}$ represent the amplitude of the post-retinal magnocellular responses to the first and second pulses, respectively. The value t_s represents the temporal separation between these two responses. In other words, our linking assumption was that observers used the response pulses and the temporal separation between them to decide whether they perceived one or two pulses. Stronger pulses indicated more reliably the occurrence of two, rather than one pulse, and the more these pulses were separated in time, the higher was the likelihood of having two as opposed to one input pulse. Hence, the performance was taken as the product of the post-retinal magnocellular peak responses to the two pulses multiplied by the temporal separation between them (t_s). Accordingly, the temporal-resolution task-performance increases with temporal separation between the input signals and the amplitude of the magnocellular responses.

The model has a total of 48 parameters. However, to fit the simulated results to the data, we fixed 45 parameters and took only three parameters as free viz., the standard deviations of the center and surround Gaussians of the sustained retinal neurons ($\boldsymbol{G^{se}}$ and $\boldsymbol{G^{si}}$ in Table 4-2), and the attentional strength parameter Θ. The remaining parameter values for the RECOD model were taken from previous works (Breitmeyer et al., 2006; Purushothaman et al., 2002).

4.4. Simulation of behavioral data

4.4.1. The effects of spatial attention in spatial resolution tasks

The effects of spatial attention on spatial resolution have been reported in several studies (e.g., Bonder et al., 2018; Carrasco et al., 2002; Yeshurun & Carrasco, 1998, 1999). For instance, using peripheral pre-cuing, Yeshurun & Carrasco (1999) explored the effect of covert transient attention (i.e., attending to the stimulus without eye movements) on performance in spatial acuity and hyperacuity tasks. To do so, they designed a psychophysical experiment that proceeded in the following way (see Fig. 4-3 for an illustration). First, observers were asked to fixate on a fixation dot for 1000 ms. After that, either a spatial cue (a horizontal line lasting 54 ms) appeared in a peripheral location just above the location of the future target, or a neutral cue (a green dot with a duration of 54 ms) appeared in the center of the display. The cue was followed by an inter-stimulus interval (ISI, the time between the offset of the cue and the onset of the target stimulus) of 67 ms. Then, a stimulus designed to measure spatial resolution —viz., a square stimulus with a gap inserted in one of its sides or a "Landolt-square"— appeared briefly (80 ms) in one of 16 possible locations. Finally, a mask of 200 ms followed the target stimulus. In this

experiment, the observer's task was to indicate, as rapidly and accurately as possible, whether the gap was on the left or right side of the Landolt-square.

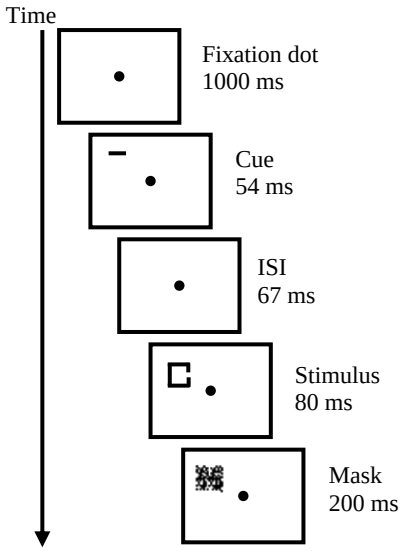

Fig. 4-3. Depiction of the experiment and sequence of presentation in Yeshurun et al., 1999. See main text for a description.

The results of this experiment showed an improvement in performance (% correct) for the condition where the target location was indicated by a spatial cue (*cued condition*), in comparison to the condition where no spatial information was provided (*neutral condition*). Likewise, a reduction in reaction times for the cued condition in comparison to the neutral condition was observed. Yeshurun and Carrasco (1999) attributed these effects to a signal enhancement mechanism in which attention increases the strength of the sensory representation in the attended location, resulting in improved spatial resolution (Lu & Dosher, 1998; Posner, 1980; Yeshurun & Carrasco, 1998, 1999).

4.4.1.1 Schematic explanation of model predictions

We start with a simple, intuitive explanation of how our model can account for the data in Yeshurun & Carrasco (1999) using a schematic representation (Fig. 4-4). During the cued condition, the spatial cue, represented by a rectangular signal, is presented first and generates through Eqs 7 and 8, both a fast transient response in the post-retinal magnocellular stream (red signal in Fig. 4-4) and a slower sustained response in the post-retinal parvocellular stream (green signal in Fig. 4-4). Then, a *spatially localized* signal (dark grey signal in Fig. 4-4) is used to modulate the excitatory input to the post-retinal sustained response generated by the target stimulus (Att in Eq. 7). This event, in turn, yields a higher neural activity in the parvocellular stream (red outlined green response at

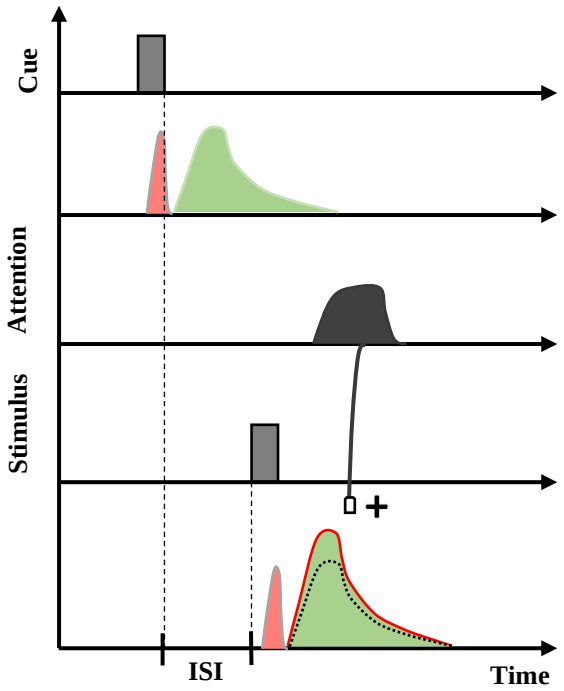

Fig. 4-4. Schematic explanation of model predictions for the effect of spatial attention on spatial resolution. The cue and target stimuli generate transient (salmon) and sustained (green) responses in the post-retinal magnocellular and parvocellular streams. In addition, the excitatory modulatory effect of spatia attention on the parvocellular stream produces a response increment (solid red-line response), translating itself into improved spatial resolution.

the bottom of Fig. 4-4), which, according to our model, is processing the spatial aspects of the stimulus (e.g., the gap resolution).

On the other hand, when the spatial cue is absent, i.e., during the neutral condition, the attentional mechanism remains silent, and no excitatory signal is sent to the post-retinal sustained response of the stimulus (represented by the dotted black signal in Fig. 4-4). Consequently, no enhancement in the sustained stream is produced, which means that spatial resolution performance is not as good as when there is a spatial cue. Thus, just as in the experimental results, we expect to have a better performance in terms of accuracy and reaction time for the cued condition compared to the neutral condition. In our simulation, we tested this hypothesis

4.4.1.2 Simulation results

We simulated the main result in Yeshurun et al. (1999). They measured performance in a gap resolution task. As expected, participants' performance increased as a function of gap size (Fig. 4-5A). Critical, however, was the finding that the condition where the target location was spatially cued showed a significant performance improvement compared to the condition where no target-location information was provided. The same effect of spatial attention was observed in terms of reaction times (Fig. 4-5B), i.e., the time required to report the correct location of the gap was shorter for the cued condition than for the neutral condition.

In our simulations, we used rectangular signals as inputs to the system (see simulation methods). The width of the signal was taken to represent changes in gap size. After passing these signals to the system, the dynamics of the equations governing the magnocellular and parvocellular retinal-pathways generate transient and sustained responses, respectively.

These responses, in turn, are feedforwarded to the post-retinal network. The result is the post-retinal transient/magno and sustained/parvo responses shown in Fig. 4-5. In this work,

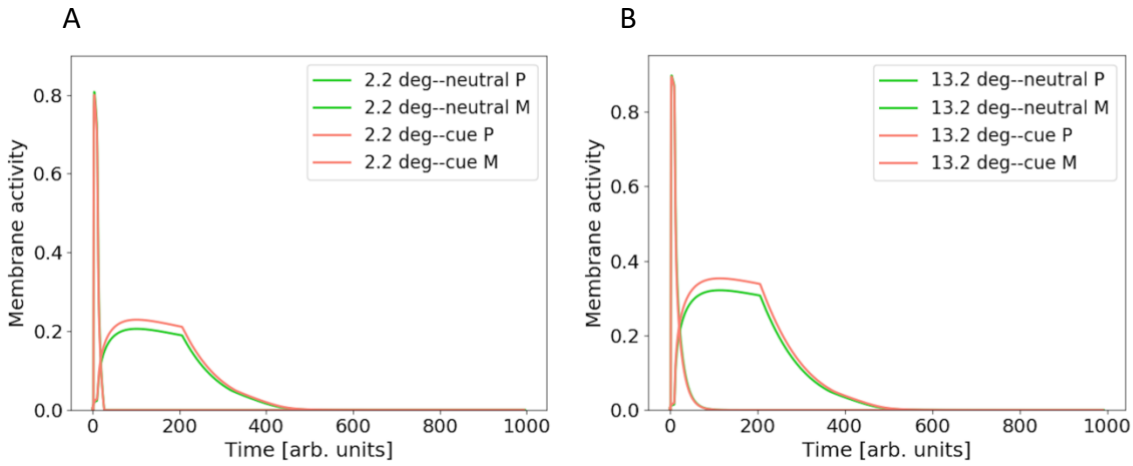

Fig. 4-5. Simulated response profile of the post-retinal magnocellular (M) and parvocellular (P) neurons in the spatial resolution task. A. Transient and sustained post-retinal responses to a 2.2 deg target stimulus. Salmon and green colors represent the cue and neutral condition, respectively. Note the increase in membrane activity with the attentional cue. **B.** Transient and sustained post-retinal responses to a 13.2 deg target stimulus.

we assume that the parvocellular stream primarily processes perceptual tasks requiring high spatial acuity (e.g., discrimination of small gaps in an object). Therefore, enhancement in the activity of the parvocellular pathway by spatial attention improves the sensitivity to high spatial frequencies, resulting in better spatial resolution performance. Our linking hypothesis relating the simulation values to the behavioral data was to assume that the strength of the neural activity of the post-retinal parvocellular neurons responding to the gap size reflected the ability of the observer to arrive at a correct decision. In other words, the larger the membrane activity of these neurons, the better the performance at discriminating the correct location of the gaps. For instance, Fig. 4-5 shows the response profile of a neuron located at the center of a small gap (Fig. 4-5A) and a large gap (Fig. 4-5B) for the post-retinal magnocellular and post-retinal parvocellular neurons, respectively.

As observed, the parvo-sustained response increases as a function of gap size (compare the response to a 2.2 deg and 13.2 deg gap size in Fig. 4-5A and 4-5B, respectively) and cueing condition (compare the salmon and green lines in both figures). Thus, improved performance with gap size and attention is expected.

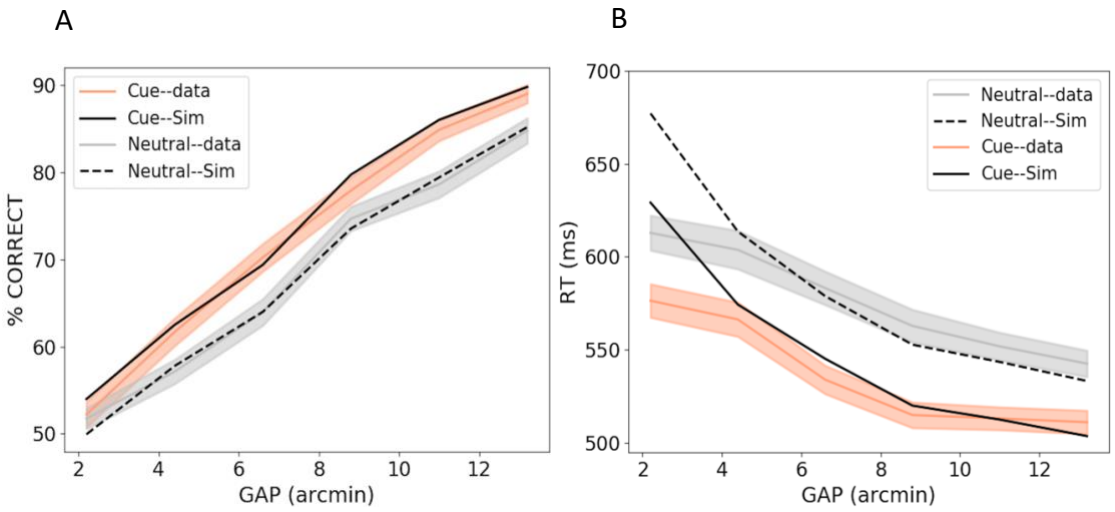

Fig. 4-6. Simulating the results in Yeshurun et al. (1999). A. Percent correct as a function of gap size. The grey line shows the experimental data for the neutral condition. The salmon line shows the experimental data for the cued condition. The black dashed line and the solid black line depict the simulation result for the neutral and cued conditions. **B.** Reaction time as a function of gap size. Colored shaded areas represent ± S.E.M. in the experimental data.

The final simulation results along with behavioral data are presented in Fig. 4-6. The solid black line in both figures represents the simulation results for the cued condition, whereas the dashed black lines indicate the simulation results for the neutral condition. As shown in Fig. 4-6A, our model shows an excellent quantitative fit to the data ($R^2 = 0.98$). The model can capture (1) the general improvement in performance as a function of gap size and (2) the improvement in performance in the cued condition due to the effect of spatial attention. Further, except for the smallest gap size, our model can also predict the qualitative shape of the results relating reaction-time duration to gap size detection (Fig. 4-6B). To simulate reaction time data, we measured the model time required for the post-

retinal parvocellular neurons to reach a fixed activity level at varying gap widths (see simulation methods). Our model predicts a decrease in reaction time with gap size and a more robust reduction in reaction time for the cue condition compared to the neutral condition. For the smallest gap size, our model disagrees with the experimental data. It is not clear why experimental reaction-times tend to be flat at very small gap sizes. It may be due to a change in decision criterion-level in case the observers find these small gap trials more difficult than relatively larger gaps. With the exception of this discrepancy, our model offers an excellent quantitative fit to the data.

4.4.2. The effects of spatial attention on temporal resolution tasks

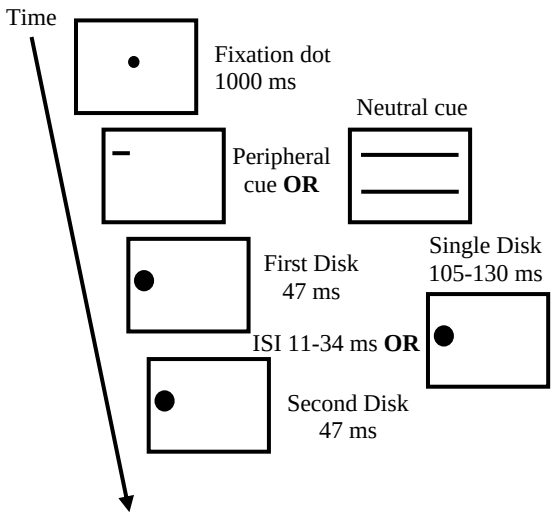

Fig. 4-7. Depiction of the experiment and sequence of presentation in Yeshurun et al., 2003. See main text for a description.

Just as in the spatial domain, the effects of spatial attention on temporal resolution have also been studied (Carrasco et al., 2002; Yeshurun, 2004; Yeshurun & Levy, 2003; Yeshurun & Marom, 2008). Yeshurun & Levy (2003) investigated the effects of transient

spatial attention on temporal resolution. To assess temporal resolution, they designed a psychophysical experiment to measure the two-flash fusion threshold. The two-flash fusion threshold is a classic method for quantifying temporal resolution. This paradigm measures the minimum time interval required for observers to detect two successive flashes of light as actually two separate flashes rather than as a single flash (Fig. 4-7). To evaluate the effect of spatial attention on temporal resolution, Yeshurun & Levy combined the two-flash fusion threshold paradigm with peripheral pre-cuing, which is a common way to manipulate transient spatial attention (Yeshurun & Levy, 2003). The experiment was conducted as follows (Fig. 4-7): at the beginning of each trial, observers fixated on a fixation dot for 1000 ms. Right after that, either a peripheral (a short horizontal line just above the location of the target) or a neutral cue (two long lines across the display) was presented, followed by the first stimulus flash (3°disk presented for 47 ms). In half of the trials, an ISI was introduced between the two flashes. The ISI served as the independent variable in the experiment. Instead of two flashes, a longer-lasting single flash was presented on the other half of the trials. In each trial, observers reported whether they saw a single flash or two separate flashes.

4.4.2.1 Schematic explanation of model predictions

As stated above, spatial attention degrades the performance of the visual system on temporal resolution tasks. To gain an intuition of how our model can explain this experimental finding, we first present a schematic explanation of the model predictions in Fig. 4-8. The grey rectangular signals in Fig. 4-8 represent briefly flashed stimuli, each of which generates a transient and sustained post-retinal response (Eqs. 8 and 7, respectively). In the presence of a peripheral cue, an attentional signal (dark grey signal) modulates the

post-retinal sustained response produced by the first flashed stimulus (S1). The effect is an enhancement of the activity in the post-retinal parvocellular stream (red-outlined sustained response in Fig. 4-8).

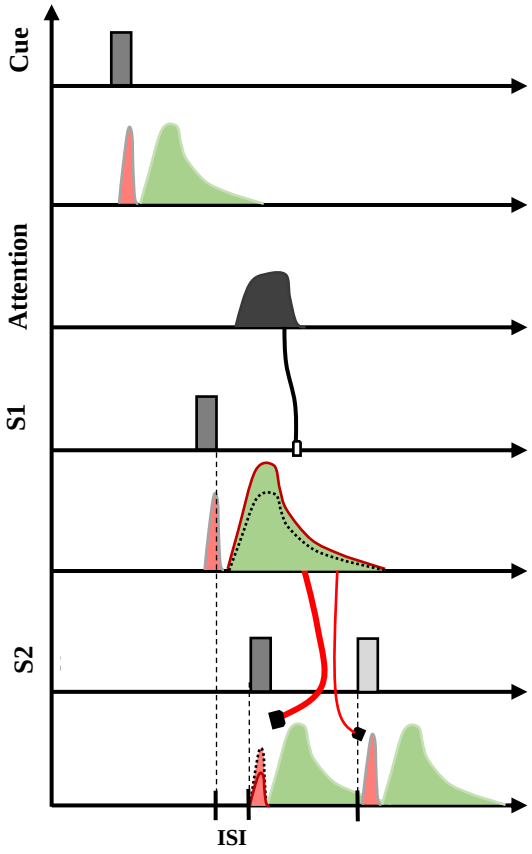

Fig. 4-8. Schematic explanation of model predictions for the effect of spatial attention on temporal resolution. The sustained response of the first flash (S1) inhibits via inter-channel interactions the magnocellular response of the second flash (S2). This effect is significantly stronger at shorter interstimulus intervals (thick inhibitory red line). As the interstimulus interval between the flashes increases (light grey S2), inhibition decreases (thin inhibitory red line). When attention is added as a modulatory signal to the sustained response of S1, the inhibitory effect to the transient response of S2 increases (red-outline transient response), this outcome disappears when attention is absent.

Through the inter-channel inhibition between the parvocellular and magnocellular streams, the newly enhanced sustained signal of the first flash inhibits the transient response of the magnocellular signal produced by the second flash (S2). This result is

represented by the attenuated transient response at the bottom of Fig. 4-8 (outlined in red). Our model postulates that temporal resolution relies mainly on the magnocellular activity. Therefore, by inhibiting this activity, the ability of the observer to discriminate between the two flashes gets hampered.

In contrast, during the neutral cue condition, no spatial-attentional signal is generated. As a result, the sustained response of S1 is not enhanced (dotted sustained response in Fig. 4-8), and the resulting inhibition to the transient response of S2 is not as potent as in the peripheral cue condition (dotted transient response in Fig. 4-8). The black dotted signals represent the sustained and transient responses for the neutral condition in Fig. 4-8. Based on our modeling assumption, this unattenuated transient response translates into better performance in the temporal resolution task when compared to the cue condition.

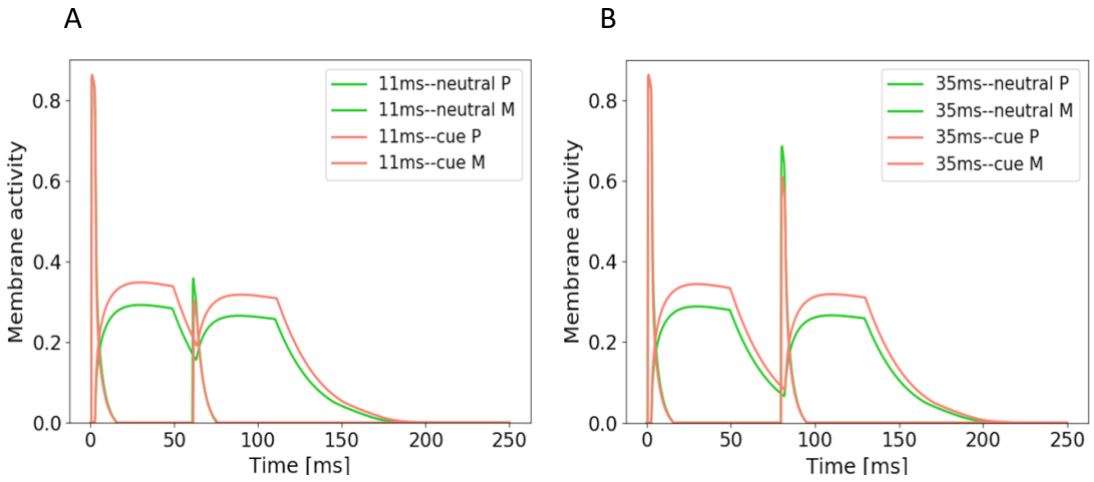

Fig. 4-9. Simulated response profile of the post-retinal magnocellular (M) and parvocellular (P) neurons in the temporal resolution task. A. Transient and sustained post-retinal responses of two flashes separated by a ISI of 11 ms. Salmon and green colors represent the cue and neutral condition, respectively. **B.** Transient and sustained post-retinal responses of the two flashed stimuli separated by a 35 ms ISI. Note that as a function of ISI, the amplitude of the magno response of the second flash is increased. Also, regardless of ISI, the response of the cue condition (salmon line) is attenuated compared to the neutral condition (green line).

4.4.2.2 Simulation results

Our model predicts that, regardless of attentional effect, the observer's performance must exhibit a monotonic increase as a function of ISI. Per our linking assumption, two factors explain this outcome: First, as ISI increases, the temporal interval between the two cortical magnocellular responses increases and second, the amplitude of the second magnocellular response increases. This latter result is because the final membrane activity of the post-retinal magnocellular stream is a function of the inter-channel inhibition coming from the post-retinal parvocellular stream. Hence, as the temporal separation between the flashes increases, the inhibitory effect of the sustained response of the first flash on the transient response of the second flash is less effective due to the expected decay of the sustained response with time. This effect is illustrated in Fig. 4-9. Panel A shows the transient and sustained post-retinal responses for the stimulus at an ISI of 11 ms. Panel B presents the responses at 35 ms ISI. Notice that the second flash's transient response is less suppressed at the large ISI. This result comes from the poor inhibition coming from the sustained response of the first flash. Secondly, the model also predicts the deterioration of performance with spatial attention. As described in the previous section, the attentional enhancement to the response of the post-retinal sustained channel of the first flash results in more potent inhibition of the post-retinal transient response of the second flash. This effect, in turn, yields a poor performance on the detection of the second flash. This effect is absent when the spatially localized attentional cue signal is removed. In Fig. 4-9, this effect is shown by the less prominent magnocellular membrane activity for the cue condition (salmon lines) compared to the neutral condition (green lines).

Finally, Fig. 4-10A shows the main result of Yeshurun & Levy (2003) and our model's simulation (black lines). As a function of ISI, the observers' performance (measured as $d' = z(hit) - z(false\ alarm)$) monotonically increased for both the *neutral* and the *cue* condition. Nevertheless, contrary to spatial resolution, when attention was directed to the target location, the observer's ability to detect the temporal gap degraded (salmon line in Fig. 4-10). Moreover, in Fig. 4-10B we show that the model predicts the negative

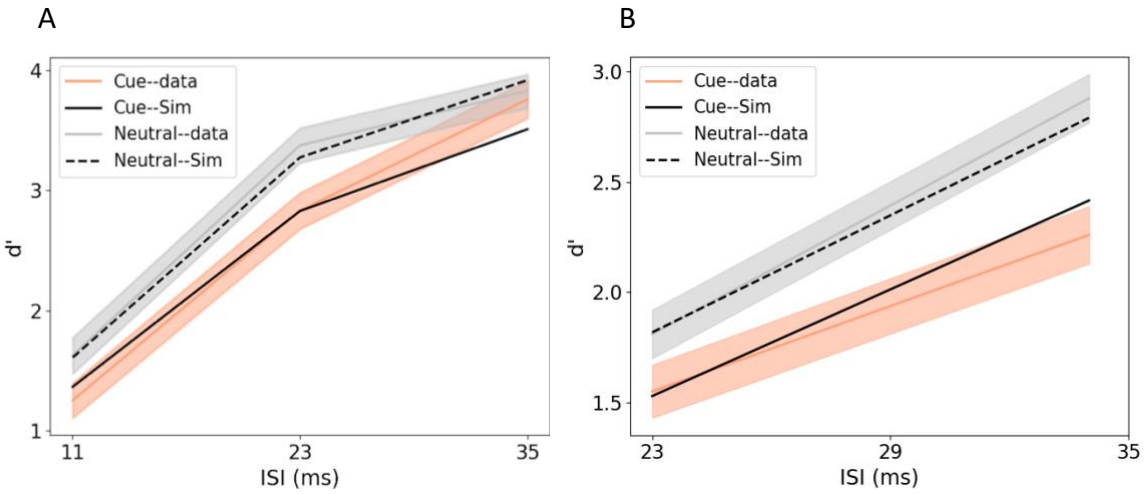

Fig. 4-10. Simulating the results in Yeshurun et al. (2003). A. Performance as a function of ISI. The grey line shows the experimental data for the neutral condition. The salmon line shows the experimental data for the cued condition. The black dashed line and the solid black line depict the simulation result for the neutral and cued conditions. **B.** Same as A, but the stimulus size was changed from 3° to 0.3°. Colored shaded areas represent $\pm$ S.E.M. in the experimental data.

effect of spatial attention on temporal resolution even when the stimulus size was reduced from 3° to 0.3° (see discussion).

We fit the simulation results to the data by taking the performance to be a function of the peak amplitudes of the magnocellular responses and the temporal separation between them (see simulation methods). Thus, the model's performance is expected to increase as a function of ISI and magnocellular responses. Furthermore, due to the role of spatial attention and reciprocal inhibition between the parvo-magno channels, the performance for

the cue condition must be poorer than the performance for the neutral condition. Our simulation results demonstrate that our model captures this effect. The solid and dashed black lines in Fig. 4-10 show the simulated results for the cue and neutral condition, respectively $(R^2 = 0.95)$.

4.5. Discussion

Several mechanisms have been proposed to explain the effects of attention on perceptual discriminability. One such mechanism suggests that the attentional precueing effects are due to observers adopting a more liberal decision criterion (Kinchla et al., 1995; Palmer, 1994; Shiu & Pashler, 1995). Other explanations propose external noise-reduction approaches, which come in two formulations (Carrasco, 2011): On the one hand, attention alters the properties of perceptual filters, letting pass the signals of interest at specific locations and diminishing external noise at other locations (Dosher & Lu, 2000b, 2000a; Lu & Dosher, 1998) and/or, on the other hand, attention enables observers to suppress distractors outside the focus of attention. Yet, another attentional mechanism is a signal enhancement process, whereby attention facilitates the processing of the relevant information at the selected location, thereby improving the quality of the stimulus representation (Carrasco et al,, 2000; Carrasco et al., 2002; Dosher & Lu, 2000b, 2000a; Lu & Dosher, 1998; Yeshurun & Carrasco, 1999). Our model architecture and the selected behavioral data we simulated support the signal enhancement hypothesis. One reason for this is that the spatial precue in Yeshurun et al. (1999) indicated the target location, but it was not informative about the correct response. Therefore, this shows that the improved performance with attention is a product of sensory and not of decision processes. A second reason is that both studies (Yeshurun & Carrasco, 1999; Yeshurun & Levy, 2003) used

suprathreshold and isolated targets. Thus, whatever observed attentional benefit could not be explained in terms of noise-reduction approaches since the target stimuli could not be confused with the background (the spatial uncertainty introduced by threshold-level targets is precluded when suprathreshold stimuli is used) (Carrasco et al., 2000; Carrasco et al., 2002; Nachmias, 2002; Palmer, 1994; Pelli, 1985) or with added external noise (since there were no distractors). Therefore, the most likely attentional mechanism is a signal enhancement process that facilitates processing the stimulus of interest at the attended location.

The attentional signal-enhancement of relevant information can be achieved in different ways. One proposed mechanism suggests that attention may reduce the noise of the internal representation of the stimulus by decreasing the variance in the perceived quality of the stimulus (Prinzmetal et al., 1998; Prinzmetal et al., 1997). However, this approach presupposes that attention must always help performance and, therefore, it could not account for the deterioration of temporal resolution with attention observed in Yeshurun et al. (2003). Another possible mechanism that could, in fact, explain the pre-cuing effects of attention on spatial and temporal resolution is a mechanism that reduces the size of receptive fields (Desimone & Duncan, 1995). As neurons with small receptive fields are more sensitive to small stimuli, an attentional mechanism that reduces the receptive field size of neurons explains the enhanced performance in spatial resolution. At the same time, smaller receptive fields yield a reduction in spatial summation and this, in turn, generates prolonged temporal integrations, which explains the deterioration in temporal resolution (Hart, 1987; Yeshurun, 2004; Yeshurun & Levy, 2003). A related attentional mechanism proposed to explain several attentional effects suggests that

attention attracts the receptive fields of neurons towards the attended location (Baruch & Yeshurun, 2014). According to this view, by attracting the receptive fields of lower layer afferent neurons toward the locus of location, the receptive field of higher layer neurons shrinks. This attentional shrinkage in the receptive field results in improved spatial resolution. Furthermore, the attentional attraction of receptive fields increases the number of responding neurons; combining this with the inherent variability in individual neurons' responses results in a temporal population response that rises faster and decays later than the case without the attentional attraction of receptive fields. As a result, this outcome is taken to explain the decrement in temporal resolution. The previous two mechanisms rely on reducing spatial summation (by receptive filed shrinkage) to account for the decrease in temporal resolution. However, the effect of spatial summation for small stimuli is negligible (e.g., Brown, 1965; Schieting & Spillmann, 1987; Yeshurun & Levy, 2003). In fact, using a stimulus of diameter 0.3° (thought to be unaffected by spatial summation), Yeshurun & Levy (2003) showed that the detrimental effect of attention on temporal resolution was observed even with this small stimulus. This result demonstrates that the attentional mechanism underlying the temporal-resolution effect does not solely depend on reducing the stimulus summation area. Finally, as pointed out elsewhere (Yeshurun, 2004; Yeshurun & Levy, 2003), a probable but less parsimonious proposal is two independent attentional mechanisms that, through reduction of receptive field size and lengthening of temporal integration, could explain the enhanced spatial resolution and the reduced temporal resolution, respectively.

Attention is a complex process. Depending on task demands, experimental designs, and other factors, it is very likely that not just one but multiple mechanisms provide attentional

facilitation. Nevertheless, a signal-enhancing mechanism that facilitates the parvocellular over the magnocellular pathway and that incorporates reciprocal inhibition between the two pathways seem to account better for the effects of spatial attention on spatial and temporal resolution and to be more biologically viable. For instance, by facilitating the parvocellular cells, spatial attention enhances spatial resolution as these neurons are more sensitive to higher spatial frequencies (e.g., Kaplan, 2003, 2013). At the same time, the facilitation of parvo-like sustained responses and the subsequent inhibition of magnocellular neurons also explains the decrement of temporal resolution. Evidence of this sustained-on-transient inhibition is supported by the finding that isoluminant stimuli and red backgrounds greatly reduce the effects of transient spatial attention on temporal resolution (Yeshurun, 2004). Similarly, evidence of transient-on-sustained inhibition is supported by similar findings showing that red backgrounds attenuate the magnitude of metacontrast masking (Breitmeyer & Breier, 1994; Breitmeyer & Williams, 1990), a phenomenon thought to be mediated by reciprocal inhibitory interactions between sustained and transient channels (Banta & Breitmeyer, 1985; Breitmeyer & Ganz, 1976). These results are expected because the magnocellular system is relatively colorblind (e.g., Merigan & Maunsell, 1993; Schiller & Logothetis, 1990) and inhibited by diffused red light (e.g., Livingstone & Hubel, 1984; Schiller & Malpeli, 1978); therefore, the parvo system becomes the primary mediator in processing these stimuli. Consequently, in the context of Yeshurun's (2004) results, as the red background suppresses the magno-system, any parvo-on-magno inhibition yields a negligible effect on performance, thus reducing the adverse effect of attention on temporal resolution. Yet in another study, using a pulsed-pedestal and a steady-pedestal paradigm, Yeshurun et al. (2012) found a differential effect

of transient spatial attention on performance with these two paradigms. The pulsed-pedestal paradigm consists of simultaneously flashed pedestal elements— a background luminance field —plus a test stimulus. The steady-pedestal paradigm consists of a test stimulus plus continuously presented luminance pedestals. These two paradigms are designed to stimulate either the parvocellular (by saturation of the magno system) or the magnocellular system (by saturation of the parvo system), respectively (e.g., Leonova et al., 2003; Pokorny, 2011; Pokorny & Smith, 1997). The results showed that attention had a differential effect on performance in the pulsed-pedestal paradigm compared to no effect on the steady-pedestal paradigm, suggesting that transient spatial attention favors parvocellular over magnocellular processing. Finally, Fig. 5B shows that our model, which favors the parvo-magno inhibitory hypothesis, accounts for the detrimental effect of spatial attention on temporal resolution in a stimulus condition where the mere reduction of spatial summation area seems insufficient as an explanation.

To summarize. In this work, we tested the predictions of a neural model that accounts for the effects of spatial attention on spatial and temporal acuity via enhancement of the parvocellular pathway and reciprocal inhibition between the parvocellular and magnocellular channels. To our knowledge, this is the first study to provide computational evidence for this attentional mechanism.

Chapter 5. Conclusions and future work

In this dissertation, we studied mechanisms by which the visual system adapts to ecological vision given its intrinsic limitations in terms of dynamic range and processing capacity. Via behavioral experiments and computational modeling, we studied mechanisms such as an adaptive spatial-integration strategy and attention as functional solutions to overcome these limitations. We concentrated on the visual system due to its essential role in our daily life and survival. In the first study, using visual psychophysics, we demonstrated that the visual system uses an adaptive center-surround mechanism to compute motion in non-retinotopic coordinates. This result is significant because motion perception is carried out mostly non-retinotopically (i.e., computing motion using a reference frame other than the retina). Thus, when the environment requires the visual system to augment the signal-to-noise ratio (e.g., at low-contrast, foggy conditions), spatial summation takes place to avoid the input information being swamp by noise. Nevertheless, when the signals in the environment are strong (e.g., at high-contrast, bright conditions), the visual system may require the suppression of the input signals to avoid saturation of its processing neurons. In summary, this finding sheds light on the adaptive spatial-integration strategy by which the brain extracts motion information under different conditions in ecological vision.

The neural mechanism that implements this adaptive center-surround mechanism is still unknown. Therefore, in the second study, we proposed a biologically plausible

mechanistic model to explain how the brain realizes the adaptive spatial-integration strategy that helps it adapt to the constant changes in the environment. The critical characteristic of mechanistic models is that they explain a phenomenon in terms of its (assumed) components parts, their interactions, and their organization. In other words, mechanistic models explain how phenomena emerge from their constituent parts. Consequently, in our model, we showed that the combination of three canonical computational elements—viz., multiplicative interactions, center-surround receptive-field organization, and sigmoidal nonlinearities—not only captures the transition from a spatial summation mechanism (for weak signals) to a spatial suppression mechanism (for strong signals) but also the interactions between motion discrimination and motion segregation processes. This result is relevant not only from the computational theoretical point of view, but also from the clinical and engineering perspective. For instance, deficits in spatial suppression, i.e., the ability to perceptually suppress large background-like motions, have been found to occur in several special populations, including the elderly and individuals with schizophrenia or depression (for a review, see Tadin, (2015)). Our biologically plausible, quantitative, and mechanistic model for adaptive center-surround mechanisms can potentially be used to study the neural mechanisms underlying this processing deficit more realistically. Specifically, by fitting our model to data from these special populations and systematically finding the parameters of interest to the system's behavior, the perceptual deficit could be traced to specific brain areas or mechanisms. For instance, with the current architecture, our model predicts that an increase/decrease of the saturation parameter (Amp, see sensitivity analysis in appendix B.3) for the center/surround non-linearities generates a reduction in spatial suppression, which explains the deficits in spatial

suppression shown by the elderly population. Malfunctioning in the excitatory or inhibitory mechanisms of neurons could be at the root of the behavioral deficit. This information could then be used to generate new treatments or contribute to the development of new drugs. Future simulations and refinements to the model need to be done to test these predictions.

Moreover, our model is a dynamical system, i.e., it describes how the system develops over time. We believe, therefore, that our results are of great interest to artificial intelligence and machine learning researchers interested in producing systems that learn through real-time interactions with the environment. Moreover, this approach is fundamental in the context of the quest to develop low-energy and computationally-efficient intelligent computing platforms that have motivated the interest to explore the brain in recent years (Bengio et al., 2015; Kulkarni & Rajendran, 2018; NAE, 2009; Pfeiffer & Pfeil, 2018). For example, as our model incorporates the dynamics of the system, an AI system that interacts with the environment in real time will benefit from the neural strategies used by the brain (e.g. the adaptive center-surround mechanism) to *efficiently* adapt to the signals in the world.

Another important future direction for our model is to expand it to two spatial dimensions. In this way, more complex input stimuli (e.g., images) can be used, and a richer representation of the input stimuli is obtained. The main obstacle to attain this goal is computing time since the number of differential equations dramatically increases in two spatial dimensions. However, with the advent of quantum computing, increases in computing power can easily overcome these limitations.

Finally, in the last study, we modeled the effects of spatial attention on temporal and spatial acuity. In particular, we tested the predictions of a biologically plausible neural model (the RECOD model) that explains these attentional effects via interactions between channels with different spatiotemporal sensitivities—viz., the magnocellular and parvocellular channels. More specifically, our model supports the hypothesis that spatial attention by enhancing the parvocellular pathway produces improved performance in spatial resolution tasks, whereas the performance decrement in temporal resolution tasks is explained via parvo-magno inhibitory interactions. By helping the brain select and optimize the most relevant information for our survival, spatial attention serves a critical functional role in our daily interactions with the environment. For instance, spatial attention is exploited by the visual system to compensate for its processing deficiencies in the periphery, i.e., spatial attention enhances spatial resolution and decreases temporal resolution in favor of temporal integration in the periphery, two characteristics necessary for better decision-making. Further experiments need to be conducted to study the effects of peripheral spatial attention on the adaptive center-surround mechanism. It is interesting to observe whether or not there is further impairment of motion discrimination of large high-contrast stimuli in the presence of spatial attention.

To summarize, in order to successfully interact with our everchanging and dynamic environment, the primate brain uses a synergy of multiple structural and functional solutions. This dissertation studied spatial integration strategies, spatial attention, and the interaction between structural visual pathways as solutions used by the primate brain to adapt to varying ecological conditions given its intrinsic limitations. Understanding the neural mechanisms by which the primate brain adapts to the environment is fundamental

in our collective effort to reverse-engineering the human brain. In the coming decades, findings in this field will have revolutionary impacts on basic science, engineering applications, clinical treatments, and new drug discoveries. This dissertation serves as a small contribution to the advancement of our knowledge in this direction.

Publications

Chapter 2:

Conference publication

- Peñaloza, B., Herzog, M. H., & Öğmen, H. (2019). Adaptive center-surround mechanisms in non-retinotopic processes. Presented at the Vision Sciences Society (VSS) 2019 Annual Conference, *Journal of Vision*; 19(10):295b. doi: https://doi.org/10.1167/19.10.295b

Peer-reviewed journal publication

- Peñaloza, B., Herzog, M. H., & Öğmen, H. (2020). Non-retinotopic adaptive center-surround modulation in motion processing. *Vision Research*, *174*, 10–21. https://doi.org/10.1016/j.visres.2020.05.007

Chapter 3:

Conference publication

- Peñaloza, B., Herzog, M. H., & Öğmen, H. (2021). Adaptive Trade-off between Sensitivity and Spatial Resolution and its Implications for Motion Discrimination and Segregation. Presented at the Vision Sciences Society (VSS) 2021 Annual Conference, to appear in abstract form in the *Journal of Vision*.

Peer-reviewed journal publication

- Peñaloza, B., Herzog, M. H., & Öğmen, H. (2021). Adaptive mechanisms of visual motion discrimination, integration, and segregation. *Vision Research*, *188*, 96–114. https://doi.org/10.1016/J.VISRES.2021.07.002

Chapter 4:

Peer-reviewed journal publication

- Peñaloza, B., & Öğmen, H. (*to be submitted*). Effects of Spatial Attention on Spatial and Temporal Acuity: A Computational Account.

References

Abbott, L. F., & Chance, F. S. (2005). Drivers and modulators from push-pull and balanced synaptic input. In *Progress in Brain Research* (Vol. 149, pp. 147–155). Elsevier. https://doi.org/10.1016/S0079-6123(05)49011-1

Agaoglu, M. N., Clarke, A. M., & Herzog, M. H. (2017). Motion-based nearest vector metric for reference frame selection in the perception of motion, *16*(2016), 1–16. https://doi.org/10.1167/16.7.14.doi

Agaoglu, S., Breitmeyer, B., & Ogmen, H. (2016). Metacontrast masking and attention do not interact. *Attention, Perception, and Psychophysics*, *78*(5), 1363–1380. https://doi.org/10.3758/s13414-016-1090-y

Allman, J., Miezin, F., & McGuinness, E. (1985). Stimulus Specific Responses from Beyond the Classical Receptive Field: Neurophysiological Mechanisms for Local-Global Comparisons in Visual Neurons. *Annual Review of Neuroscience*, *8*(1), 407–430. https://doi.org/10.1146/annurev.ne.08.030185.002203

Anderson, S J, & Burr, D. C. (1991). Spatial summation properties of directionally selective mechanisms in human vision. *Journal of the Optical Society of America. A, Optics and Image Science*, *8*(8), 1330–1339. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1919836

Anderson, Stephen J., & Burr, D. C. (1985). Spatial and temporal selectivity of the human motion detection system. *Vision Research*, *25*(8), 1147–1154. https://doi.org/10.1016/0042-6989(85)90104-X

Ayaz, A., & Chance, F. S. (2009). Gain modulation of neuronal responses by subtractive and divisive mechanisms of inhibition. *Journal of Neurophysiology*, *101*(2), 958–

968. https://doi.org/10.1152/jn.90547.2008

Bachmann, T. (1984). The process of perceptual retouch: Nonspecific afferent activation dynamics in explaining visual masking. *Perception & Psychophysics*, *35*(1), 69–84. https://doi.org/10.3758/BF03205926

Banta, A. R., & Breitmeyer, B. G. (1985). Stationary patterns suppress the perception of stroboscopic motion. *Vision Research*. Vision Res. https://doi.org/10.1016/0042-6989(85)90228-7

Baruch, O., & Yeshurun, Y. (2014). Attentional attraction of receptive fields can explain spatial and temporal effects of attention. *Visual Cognition*, *22*(5), 704–736. https://doi.org/10.1080/13506285.2014.911235

Bengio, Y., Lee, D.-H., Bornschein, J., Mesnard, T., & Lin, Z. (2015). Towards Biologically Plausible Deep Learning. Retrieved from http://arxiv.org/abs/1502.04156

Betts, L. R., Sekuler, A. B., & Bennett, P. J. (2012). Spatial characteristics of motion-sensitive mechanisms change with age and stimulus spatial frequency. *Vision Research*, *53*(1), 1–14. https://doi.org/10.1016/j.visres.2011.11.004

Boi, M., Ogmen, H., & Herzog, M. H. (2011). Motion and tilt aftereffects occur largely in retinal, not in object, coordinates in the Ternus-Pikler display. *Journal of Vision*, *11*(3), 7–7. https://doi.org/10.1167/11.3.7

Boi, M., Ogmen, H., Krummenacher, J., Otto, T. U., & Herzog, M. H. (2009). A (fascinating) litmus test for human retino- vs.non-retinotopic processing. *Journal of Vision*, *9*(13), 5–5. https://doi.org/10.1167/9.13.5

Boi, Marco, Oğmen, H., Krummenacher, J., Otto, T. U., & Herzog, M. H. (2009). A

(fascinating) litmus test for human retino- vs. non-retinotopic processing. *Journal of Vision*, *9*(13), 5.1-511. https://doi.org/10.1167/9.13.5

Bonder, T., Gopher, D., & Yeshurun, Y. (2018). The joint effects of spatial cueing and transcranial direct current stimulation on visual acuity. *Frontiers in Psychology*, *9*(FEB), 159. https://doi.org/10.3389/fpsyg.2018.00159

Braddick, O. (1993). Segmentation versus integration in visual motion processing. *Trends in Neurosciences*. https://doi.org/10.1016/0166-2236(93)90179-P

Breitmeyer, B. G., & Breier, J. I. (1994). Effects of background color on reaction time to stimuli varying in size and contrast: Inferences about human M channels. *Vision Research*, *34*(8), 1039–1045. https://doi.org/10.1016/0042-6989(94)90008-6

Breitmeyer, B. G., & Ganz, L. (1976). Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychological Review*, *83*(1), 1–36. https://doi.org/10.1037/0033-295X.83.1.1

Breitmeyer, B. G., & Ganz, L. (1977). Temporal studies with flashed gratings: Inferences about human transient and sustained channels. *Vision Research*, *17*(7), 861–865. https://doi.org/10.1016/0042-6989(77)90130-4

Breitmeyer, B. G., Kafaligönül, H., Öğmen, H., Mardon, L., Todd, S., & Ziegler, R. (2006). Meta- and paracontrast reveal differences between contour- and brightness-processing mechanisms. *Vision Research*, *46*(17), 2645–2658. https://doi.org/10.1016/j.visres.2005.10.020

Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception and Psychophysics*, *62*(8),

1572–1595. https://doi.org/10.3758/BF03212157

Breitmeyer, B. G., & Williams, M. C. (1990). Effects of isoluminant-background color on metacontrast and stroboscopic motion: Interactions between sustained (P) and transient (M) channels. *Vision Research*, *30*(7), 1069–1075. https://doi.org/10.1016/0042-6989(90)90115-2

Breitmeyer, B., & Ogmen, H. (2006). *Visual Masking: Time Slices Through Conscious and Unconscious Vision*. *Visual Masking: Time Slices Through Conscious and Unconscious Vision*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198530671.001.0001

Brown, J. L. (1965). Flicker and intermittent stimulation. In C. H. Graham (Ed.), *Vision and visual perception* (pp. 251–320). New York: John Wiley & Sons, Inc.

Burkitt, A. N. (2006). A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input. *Biological Cybernetics*, *95*(1), 1—19. https://doi.org/10.1007/s00422-006-0068-6

Burr, D. C., & Santoro, L. (2001). Temporal integration of optic flow, measured by contrast and coherence thresholds. *Vision Research*, *41*(15), 1891–1899. https://doi.org/10.1016/S0042-6989(01)00072-4

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn3136

Carpenter, G. A., & Grossberg, S. (1981). ADAPTATION AND TRANSMITTER GATING IN VERTEBRATE PHOTORECEPTORS. *J. Theoretical Neurobiology*, *42*, 1–42. https://doi.org/10.1007/978-4-431-54880-5

Carrasco, Marisa. (2011, July 1). Visual attention: The past 25 years. *Vision Research*.

Pergamon. https://doi.org/10.1016/j.visres.2011.04.012

Carrasco, Marisa, Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF: Support for signal enhancement. In *Vision Research* (Vol. 40, pp. 1203–1215). Vision Res. https://doi.org/10.1016/S0042-6989(00)00024-9

Carrasco, Marissa, Williams, P. E., & Yeshurun, Y. (2002). Covert attention increases spatial resolution with or without masks: Support for signal enhancement. *Journal of Vision*, *2*(6), 467–479. https://doi.org/10.1167/2.6.4

Cavanaugh, J. R., Bair, W., & Anthony Movshon, J. (2002a). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *Journal of Neurophysiology*, *88*(5), 2530–2546. https://doi.org/10.1152/jn.00692.2001

Cavanaugh, J. R., Bair, W., & Anthony Movshon, J. (2002b). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *Journal of Neurophysiology*, *88*(5), 2530–2546. https://doi.org/10.1152/jn.00692.2001

Churan, J., Khawaja, F. A., Tsui, J. M. G., & Pack, C. C. (2008). Brief motion stimuli preferentially activate surround-suppressed neurons in macaque visual area MT. *Current Biology*. https://doi.org/10.1016/j.cub.2008.10.003

Clarke, A. M., & Herzog, M. H. (2013). Does spatio-temporal filtering account for nonretinotopic motion perception ? Comment on Pooresmaeili , Cicchini ,. *Journal of Vision*, *13*(2013), 1–14. https://doi.org/10.1167/13.10.19.doi

Clarke, A. M., Öğmen, H., & Herzog, M. H. (2015). A computational model for reference-frame synthesis with applications to motion perception. *Vision Research*, *126*, 242–253. https://doi.org/10.1016/j.visres.2015.08.018

Desimone, R., & Duncan, J. (1995, November 28). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*. Annual Reviews Inc. https://doi.org/10.1146/annurev.ne.18.030195.001205

Dosher, B. A., & Lu, Z. L. (2000a). Mechanisms of perceptual attention in precuing of location. In *Vision Research* (Vol. 40, pp. 1269–1292). Vision Res. https://doi.org/10.1016/S0042-6989(00)00019-5

Dosher, B. A., & Lu, Z. L. (2000b). Noise exclusion in spatial attention. *Psychological Science*, *11*(2), 139–146. https://doi.org/10.1111/1467-9280.00229

Eifuku, S., & Wurtz, R. H. (1998). Response to motion in extrastriate area MSTI: Center-surround interactions. *Journal of Neurophysiology*, *80*(1), 282–296. https://doi.org/10.1152/jn.1998.80.1.282

Ellias, S. A., & Grossberg, S. (1975). Pattern formation, contrast control, and oscillations in the short term memory of shunting on-center off-surround networks. *Biological Cybernetics*. https://doi.org/10.1007/BF00327046

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex*, *7*(2), 181–192. https://doi.org/10.1093/cercor/7.2.181

Feng, J. (2001). Is the integrate-and-fire model good enough? - A review. *Neural Networks*. https://doi.org/10.1016/S0893-6080(01)00074-0

Gallace, A., & Spence, C. (2009). The Cognitive and Neural Correlates of Tactile Memory. *Psychological Bulletin*, *135*(3), 380–406. https://doi.org/10.1037/a0015325

Gautama, T., & Van Hulle, M. M. (2001). Function of center-surround antagonism for

motion in visual area MT/V5: A modeling study. *Vision Research*, *41*(28), 3917–3930. https://doi.org/10.1016/S0042-6989(01)00246-2

Gegenfurtner, K. R., Kiper, D. C., Beusmans, J. M. H., Carandini, M., Movshon, J. A., & Zaidi, Q. (1994). Chromatic properties of neurons in macaque MT. *Visual Neuroscience*, *11*(3), 455–466. https://doi.org/10.1017/S095252380000239X

Glasser, D. M., & Tadin, D. (2010). Low-level mechanisms do not explain paradoxical motion percepts. *Journal of Vision*, *10*(4), 20.1-9. https://doi.org/10.1167/10.7.837

Graham, D. J. (2011). Visual perception: Lightness in a high-dynamic-range world. *Current Biology*. https://doi.org/10.1016/j.cub.2011.10.003

Grossberg, S. (1973). Contour Enhancement, Short Term Memory, and Constancies in Reverberating Neural Networks. *Studies in Applied Mathematics*, *52*(3), 213–257. https://doi.org/10.1002/sapm1973523213

Grossberg, S. (1988). Nonlinear neural networks: principles, mechanisms, and architectures. *Neural Networks*, *1*, 17–61. https://doi.org/10.1016/0893-6080(88)90021-4

Grossberg, S., Léveillé, J., & Versace, M. (2011). How do object reference frames and motion vector decomposition emerge in laminar cortical circuits? *Attention, Perception, and Psychophysics*, *73*(4), 1147–1170. https://doi.org/10.3758/s13414-011-0095-9

Grünert, U., & Martin, P. R. (2020, September 1). Cell types and cell circuits in human and non-human primate retina. *Progress in Retinal and Eye Research*. Elsevier Ltd. https://doi.org/10.1016/j.preteyeres.2020.100844

Hart, W. M. (1987). The temporal responsiveness of vision. In MR A., HW M. (Ed.),

Adler's physiology of the eye: clinical application. (8 ed., pp. 429–57). St. Louis, MO: Mosby.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, *9*(2), 181–197. https://doi.org/10.1017/S0952523800009640

Hutmacher, F. (2019). Why Is There So Much More Research on Vision Than on Any Other Sensory Modality? *Frontiers in Psychology*, *10*, 2246. https://doi.org/10.3389/fpsyg.2019.02246

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*(2), 201–211. https://doi.org/10.3758/BF03212378

Johansson, G. (1974). Vector analysis in visual perception of rolling motion - A quantitative approach. *Psychologische Forschung*, *36*(4), 311–319. https://doi.org/10.1007/BF00424568

Kalloniatis, M., & Luu, C. (2005). *Principles of Vision*. *Webvision: The Organization of the Retina and Visual System.* Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21413377

Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (1999). Dynamics of spatial summation in primary visual cortex of alert monkeys. *Proceedings of the National Academy of Sciences*, *96*(21), 12073–12078. https://doi.org/10.1073/pnas.96.21.12073

Kaplan, E. (2003). The M, P, and K Pathways of the Primate Visual System. *The Visual Neurosciences*, (January 2003), 481–493. Retrieved from https://www.researchgate.net/profile/Ehud_Kaplan/publication/256685795_The_M_P_and_K_pathways_in_the_Primate_Visual_System/links/54b5063c0cf28ebe92e4a

ede/The-M-P-and-K-pathways-in-the-Primate-Visual-System.pdf

Kaplan, E. (2013). The M, P, and K Pathways of the Primate Visual System revisited. *The Visual Neurosciences*, (January 2013), 481–493.

Keesey, U. T. (1972). Flicker and pattern detection: a comparison of thresholds. *Journal of the Optical Society of America*, *62*(3), 446–448. https://doi.org/10.1364/JOSA.62.000446

Kinchla, R. A. (1992). Attention. *Annual Review of Psychology*, *43*(1), 711–742. https://doi.org/10.1146/annurev.ps.43.020192.003431

Kinchla, Ronald A., Chen, Z., & Evert, D. (1995). Precue effects in visual search: Data or resource limited? *Perception & Psychophysics*, *57*(4), 441–450. https://doi.org/10.3758/BF03213070

Kulikowski, J. J., & Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *The Journal of Physiology*, *232*(1), 149–162. https://doi.org/10.1113/jphysiol.1973.sp010261

Kulkarni, S. R., & Rajendran, B. (2018). Spiking neural networks for handwritten digit recognition—Supervised learning and network optimization. *Neural Networks*, *103*, 118–127. https://doi.org/10.1016/j.neunet.2018.03.019

Kwon, O. S., Tadin, D., & Knill, D. C. (2015). Unifying account of visual motion and position perception. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(26), 8142–8147. https://doi.org/10.1073/pnas.1500361112

Lauffs, M. M., Choung, O. H., Öğmen, H., & Herzog, M. H. (2018). Unconscious retinotopic motion processing affects non-retinotopic motion perception. *Consciousness and Cognition*, *62*, 135–147.

https://doi.org/10.1016/j.concog.2018.03.007

Lauffs, M. M., Choung, O. H., Öğmen, H., Herzog, M. H., & Kerzel, D. (2019). Reference-frames in vision: Contributions of attentional tracking to nonretinotopic perception in the Ternus-Pikler display. *Journal of Vision*, *19*(12), 1–15. https://doi.org/10.1167/19.12.7

Lauffs, M. M., Öğmen, H., & Herzog, M. H. (2017). Unpredictability does not hamper nonretinotopic motion perception. *Journal of Vision*, *17*(9), 1–10. https://doi.org/10.1167/17.9.6

Lee, B. B., Martin, P. R., & Grünert, U. (2010). Retinal connectivity and primate vision. *Progress in Retinal and Eye Research*, *29*(6), 622–639. https://doi.org/https://doi.org/10.1016/j.preteyeres.2010.08.004

Leonova, A., Pokorny, J., & Smith, V. C. (2003). Spatial frequency processing in inferred PC- and MC-pathways. *Vision Research*, *43*(20), 2133–2139. https://doi.org/10.1016/S0042-6989(03)00333-X

Liu, L. D., Haefner, R. M., & Pack, C. C. (2016). A neural basis for the spatial suppression of visual motion perception. *ELife*, *5*. https://doi.org/10.7554/elife.16167

Livingstone, M. S., & Hubel, D. H. (1984). Anatomy and physiology of a color system in the primate visual cortex. *Journal of Neuroscience*, *4*(1), 309–356. https://doi.org/10.1523/jneurosci.04-01-00309.1984

Lu, Z. L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research*, *38*(9), 1183–1198. https://doi.org/10.1016/S0042-6989(97)00273-3

Merigan, W. H., & Maunsell, J. H. R. (1993). How parallel are the primate visual pathways? *Annual Review of Neuroscience*. Annu Rev Neurosci. https://doi.org/10.1146/annurev.ne.16.030193.002101

Moore, T., & Zirnsak, M. (2017). Neural Mechanisms of Selective Visual Attention. *Annual Review of Psychology*. https://doi.org/10.1146/annurev-psych-122414-033400

Nachmias, J. (2002). Contrast discrimination with and without spatial uncertainty. *Vision Research*, *42*(1), 41–48. https://doi.org/10.1016/S0042-6989(01)00254-1

NAE. (2009). National academy of engineering - Reverse-engineer the brain.

Nakayama, K., & Loomis, J. M. (1974). Optical velocity patterns, velocity sensitive neurons, and space perception: a hypothesis. *Perception*, *3*(1), 63–80. https://doi.org/10.1068/p030063

Öğmen, H. (1993). A neural theory of retino-cortical dynamics. *Neural Networks*, *6*(2), 245–273. https://doi.org/10.1016/0893-6080(93)90020-W

Ogmen, H., & Gagne, S. (1990). Neural network architectures for motion perception and elementary motion detection in the fly visual system. *Neural Networks*, *3*(5), 487–505.

Öĝmen, H., & Gagné, S. (1990). Neural models for sustained and ON-OFF units of insect lamina. *Biological Cybernetics*, *63*(1), 51–60. https://doi.org/10.1007/BF00202453

Öğmen, Haluk. (2007). A theory of moving form perception: Synergy between masking, perceptual grouping, and motion computation in retinotopic and non-retinotopic representations. *Advances in Cognitive Psychology*, *3*(1–2), 67–84. https://doi.org/10.2478/v10053-008-0015-2

Ogmen, Haluk, Breitmeyer, B. G., & Melvin, R. (2003). The what and where in visual masking. *Vision Research*, *43*(12), 1337–1350. https://doi.org/10.1016/S0042-6989(03)00138-X

Öğmen, H. (2005). "Spatio-temporal Dynamics of Visual Perception across Neural Maps and Pathways", in: E. Bayro-Corrochano (Ed) Handbook of Geometric Computing: Applications in Pattern Recognition, Computer Vision, Neural Computing, and Robotics, Chapter 1, pp. 3-27, Springer Verlag (2005).

Öğmen, Haluk, Breitmeyer, B. G., Todd, S., & Mardon, L. (2006). Target recovery in metacontrast: The effect of contrast. *Vision Research*, *46*(28), 4726–4734. https://doi.org/10.1016/j.visres.2006.09.006

Öğmen, Haluk, & Herzog, M. H. (2010). The geometry of visual perception: Retinotopic and nonretinotopic representations in the human visual system. *Proceedings of the IEEE*, *98*(3), 479–492. https://doi.org/10.1109/JPROC.2009.2039028

Öğmen, Haluk, Otto, T. U., & Herzog, M. H. (2006). Perceptual grouping induces non-retinotopic feature attribution in human vision. *Vision Research*, *46*(19), 3234–3242. https://doi.org/10.1016/j.visres.2006.04.007

Otto, T. U., Ogmen, H., & Herzog, M. H. (2006). The flight path of the phoenix V The visible trace of invisible elements in human vision. *Journal of Vision*, *6*, 1079–1086.

Pack, C. C., Hunter, J. N., & Born, R. T. (2005). Contrast dependence of suppressive influences in cortical area MT of alert macaque. *Journal of Neurophysiology*, *93*(3), 1809–1815. https://doi.org/10.1152/jn.00629.2004

Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research*, *34*(13).

https://doi.org/10.1016/0042-6989(94)90128-7

Pantle, A., & Picciano, L. (1976). A multistable movement display: Evidence for two separate motion systems in human vision. *Science*, *193*(4252), 500–502. https://doi.org/10.1126/science.941023

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*, *2*(9), 1508. https://doi.org/10.1364/josaa.2.001508

Peñaloza, B., Herzog, M. H., & Öğmen, H. (2020). Non-retinotopic adaptive center-surround modulation in motion processing. *Vision Research*, *174*, 10–21. https://doi.org/10.1016/j.visres.2020.05.007

Petkov, N., & Subramanian, E. (2007). Motion detection, noise reduction, texture suppression, and contour enhancement by spatiotemporal Gabor filters with surround inhibition. *Biological Cybernetics*, *97*(5–6), 423–439. https://doi.org/10.1007/s00422-007-0182-0

Pfeiffer, M., & Pfeil, T. (2018). Deep Learning With Spiking Neurons: Opportunities and Challenges. *Frontiers in Neuroscience*, *12*(October). https://doi.org/10.3389/fnins.2018.00774

Pikler, J. (1917). Sinnesphysiologische Untersuchungen. *Leipzig, Germany: Barth.*

Pokorny, J. (2011). Review: Steady and pulsed pedestals, the how and why of post-receptoral pathway separation. *Journal of Vision*, *11*(5), 7–7. https://doi.org/10.1167/11.5.7

Pokorny, Joel, & Smith, V. C. (1997). Psychophysical signatures associated with magnocellular and parvocellular pathway contrast gain. *Journal of the Optical*

Society of America A, *14*(9), 2477. https://doi.org/10.1364/josaa.14.002477

Pollen, D. A., Nagler, M., Daugman, J., Kronauer, R., & Cavanagh, P. (1984). Use of gabor elementary functions to probe receptive field substructure of posterior inferotemporal neurons in the owl monkey. *Vision Research*, *24*(3), 233–241. https://doi.org/10.1016/0042-6989(84)90125-1

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Posner, M. I., & Rothbart, M. K. (2007, December 6). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology*. Annual Reviews. https://doi.org/10.1146/annurev.psych.58.110405.085516

Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998). Phenomenology of Attention: 1. Color, Location, Orientation, and Spatial Frequency. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(1), 261–282. https://doi.org/10.1037/0096-1523.24.1.261

Prinzmetal, W., Nwachuku, I., Bodanski, L., Blumenfeld, L., & Shimizu, N. (1997). The Phenomenology of Attention. *Consciousness and Cognition*, *6*(2–3), 372–412. https://doi.org/10.1006/ccog.1997.0313

Publio, R., Ceballos, C. C., & Roque, A. C. (2012). Dynamic Range of Vertebrate Retina Ganglion Cells: Importance of Active Dendrites and Coupling by Electrical Synapses. *PLoS ONE*, *7*(10), 48517. https://doi.org/10.1371/journal.pone.0048517

Purushothaman, G., Lacassagne, D., Bedell, H. E., & Öğmen, H. (2002). Effect of exposure duration, contrast and base blur on coding and discrimination of edges. *Spatial Vision*, *15*(3), 341–376. https://doi.org/10.1163/15685680260174074

Purushothaman, G., Öğmen, H., & Bedell, H. E. (2003). Suprathreshold Intrinsic Dynamics of the Human Visual System. *Neural Computation*, *15*(12), 2883–2908. https://doi.org/10.1162/089976603322518786

Rayner, K. (2009). *Eye movements and attention in reading, scene perception, and visual search*. *Quarterly Journal of Experimental Psychology* (Vol. 62). https://doi.org/10.1080/17470210902816461

Reynolds, J. H., & Chelazzi, L. (2004, June 24). Attentional modulation of visual processing. *Annual Review of Neuroscience*. Annual Reviews. https://doi.org/10.1146/annurev.neuro.26.041002.131039

Sceniak, M. P., Hawken, M. J., & Shapley, R. (2001). Visual spatial characterization of macaque V1 neurons. *Journal of Neurophysiology*, *85*(5), 1873–1887. https://doi.org/10.1152/jn.2001.85.5.1873

Sceniak, M. P., Ringach, D. L., Hawken, M. J., & Shapley, R. (1999). Contrast's effect on spatial summation by macaque V1 neurons. *Nature Neuroscience*, *2*(8), 733–739. https://doi.org/10.1038/11197

Schallmo, M.-P., Kale, A. M., Millin, R., Flevaris, A. V, Brkanac, Z., Edden, R. A., … Murray, S. O. (2018). Suppression and facilitation of human neural responses. *ELife*, *7*. https://doi.org/10.7554/elife.30334

Schieting, S., & Spillmann, L. (1987). Flicker adaptation in the peripheral retina. *Vision Research*, *27*(2), 277–284. https://doi.org/10.1016/0042-6989(87)90190-8

Schiller, P. H., & Malpeli, J. G. (1978). Functional specificity of lateral geniculate nucleus laminae of the rhesus monkey. *Journal of Neurophysiology*, *41*(3), 788–797. https://doi.org/10.1152/jn.1978.41.3.788

Schiller, Peter H., & Logothetis, N. K. (1990). The color-opponent and broad-band channels of the primate visual system. *Trends in Neurosciences, 13*(10), 392–398. https://doi.org/10.1016/0166-2236(90)90117-S

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience, 4*(8), 819–825. https://doi.org/10.1038/90526

Shiu, L. P., & Pashler, H. (1995). Spatial attention and vernier acuity. *Vision Research, 35*(3), 337–343. https://doi.org/10.1016/0042-6989(94)00148-F

Smith, P. L., Sewell, D. K., & Lilburn, S. D. (2015). From shunting inhibition to dynamic normalization: Attentional selection and decision-making in brief visual displays. *Vision Research, 116*, 219–240. https://doi.org/10.1016/j.visres.2014.11.001

Sperling, G., & Sondhi, M. M. (1968). Model for visual luminance discrimination and flicker detection. *Journal of the Optical Society of America, 58*(8), 1133–1145. https://doi.org/10.1364/JOSA.58.001133

Stockman, A., Langendörfer, M., Smithson, H. E., & Sharpe, L. T. (2006). Human cone light adaptation: From behavioral measurements to molecular mechanisms. *Journal of Vision, 6*(11), 1194–1213. https://doi.org/10.1167/6.11.5

Stockman, A., & Sharpe, L. T. (2006, May). Into the twilight zone: The complexities of mesopic vision and luminous efficiency. *Ophthalmic and Physiological Optics*. https://doi.org/10.1111/j.1475-1313.2006.00325.x

Tadin, D. (2015). Suppressive mechanisms in visual motion processing: From perception to intelligence. *Vision Research, 115*, 58–70. https://doi.org/10.1016/j.visres.2015.08.005

Tadin, D., Kim, J., Doop, M. L., Gibson, C., Lappin, J. S., Blake, R., & Park, S. (2006).

Weakened center-surround interactions in visual motion processing in schizophrenia. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *26*(44), 11403–11412. https://doi.org/10.1523/JNEUROSCI.2592-06.2006

Tadin, D., & Lappin, J. S. (2005). Optimal size for perceiving motion decreases with contrast. *Vision Research*, *45*(16), 2059–2064. https://doi.org/10.1016/j.visres.2005.01.029

Tadin, D., & Lappin, J. S. (2012). *Linking Psychophysics and Physiology of Center-Surround Interactions in Visual Motion Processing*. https://doi.org/10.1093/acprof

Tadin, D., Lappin, J. S., Blake, R., & Grossman, E. D. (2002). What constitutes an efficient reference frame for vision? *Nature Neuroscience*, *5*(10), 1010–1015. https://doi.org/10.1038/nn914

Tadin, D., Lappin, J. S., Gilroy, L. A., & Blake, R. (2003). Perceptual consequences of centre–surround antagonism in visual motion processing. *Nature*, *424*(6946), 312–315. https://doi.org/10.1038/nature01812

Tadin, D., Park, W. J., Dieter, K. C., Melnick, M. D., Lappin, J. S., & Blake, R. (2019). Spatial suppression promotes rapid figure-ground segmentation of moving objects. *Nature Communications*, *10*(1), 2732. https://doi.org/10.1038/s41467-019-10653-8

Tadin, D., Silvanto, J., Pascual-Leone, A., & Battelli, L. (2011). Improved motion perception and impaired spatial suppression following disruption of cortical area MT/V5. *Journal of Neuroscience*, *31*(4), 1279–1283. https://doi.org/10.1523/JNEUROSCI.4121-10.2011

Takeuchi, T. (1998). Effect of contrast on the perception of moving multiple Gabor

patterns. *Vision Research*, *38*(20), 3069–3082. https://doi.org/10.1016/S0042-6989(98)00019-4

Ternus, J. (1926). Experimentelle Untersuchungen über phänomenale Identität. *Psychological Research*, *7*, 81–136.

Thunell, E., van der Zwaag, W., Ögmen, H., Plomp, G., & Herzog, M. H. (2016). Retinotopic encoding of the Ternus-Pikler display reflected in the early visual areas. *Journal of Vision*, *16*(3), 26. https://doi.org/10.1167/16.3.26

Tootell, R. B. ., Hadjikhani, N. K., Mendola, J. D., Marrett, S., & Dale, A. M. (1998). From retinotopy to recognition: fMRI in human visual cortex. *Trends in Cognitive Sciences*, *2*(5), 174–183. https://doi.org/10.1016/S1364-6613(98)01171-1

Tootell, R. B. H., Mendola, J. D., Hadjikhani, N. K., Ledden, P. J., Liu, A. K., Reppas, J. B., … Dale, A. M. (1997). Functional analysis of V3A and related areas in human visual cortex. *Journal of Neuroscience*, *17*(18), 7060–7078. https://doi.org/10.1523/jneurosci.17-18-07060.1997

Tsui, J. M. G., Hunter, J. N., Born, R. T., & Pack, C. C. (2010). The role of V1 surround suppression in MT motion integration. *Journal of Neurophysiology*, *103*(6), 3123–3138. https://doi.org/10.1152/jn.00654.2009

Watson, A. B. (2017). QUEST+: A general multidimensional Bayesian adaptive psychometric method. *Journal of Vision*, *17*(3), 10. https://doi.org/10.1167/17.3.10

Watson, A. B., & Turano, K. (1995). The optimal motion stimulus. *Vision Research*, *35*(3), 325–336. https://doi.org/10.1016/0042-6989(94)00182-L

Wilson, H. R., & Kim, J. (1998). Dynamics of a divisive gain control in human vision. *Vision Research*, *38*(18), 2735–2741. https://doi.org/10.1016/S0042-

6989(97)00328-3

Yeshurun, Y. (2004). Isoluminant stimuli and red background attenuate the effects of transient spatial attention on temporal resolution. *Vision Research, 44*(12), 1375–1387. https://doi.org/10.1016/j.visres.2003.12.016

Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature, 396*(6706), 72–75. https://doi.org/10.1038/23936

Yeshurun, Y., & Carrasco, M. (1999). Spatial attention improves performance in spatial resolution tasks. *Vision Research, 39*(2), 293–306. https://doi.org/10.1016/S0042-6989(98)00114-X

Yeshurun, Y., & Levy, L. (2003). Transient spatial attention degrades temporal resolution. *Psychological Science*, *14*(3), 225–231. https://doi.org/10.1111/1467-9280.02436

Yeshurun, Y., & Sabo, G. (2012). Differential effects of transient attention on inferred parvocellular and magnocellular processing. *Vision Research, 74*, 21–29. https://doi.org/10.1016/j.visres.2012.06.006

Appendices

Appendix A

Our stimulus consisted of Gabor patches embedded in squares elements. The purpose of the square elements was to strengthen the formation of group motion and thereby strengthen the underlying non-retinotopic reference-frame. To show that the results of Experiments 1 and 4 are independent of the presence of these square elements, we ran on one subject (BP) a control condition where we removed them from the stimulus. In Fig. A-1, we show the results of the control condition for Experiment 1. The control data (dashed lines) are similar to the data from the same subject in Experiment 1 (solid lines). For the high contrast stimulus there is a small increase in threshold for the control condition; however, the general interaction between size and contrast remains.

Fig. A-2 shows the results of the same control condition (i.e., stimuli without the squares) for Experiment 4 obtained from subject BP and compared to the data of subject BP in Experiment 4. The only major difference is that the minimum contrast required for the no-squares control condition was higher (8% vs 5.5%).

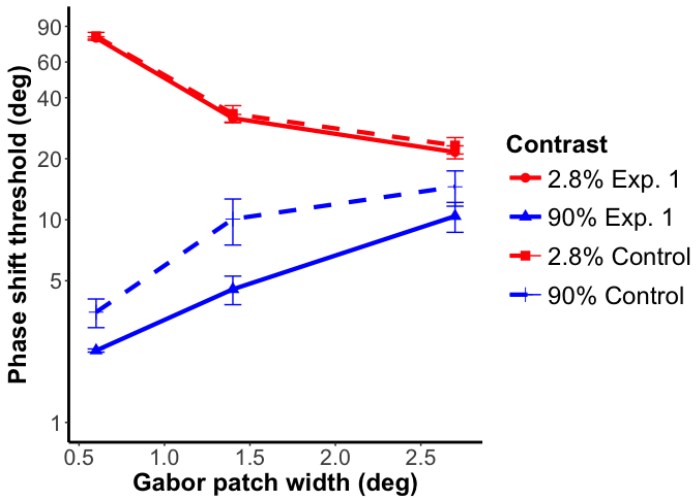

Fig. A-1. Results of the no-square control condition for Experiment 1. Solid lines represent the results from Experiment 1 for subject BP whereas dashed lines show the results from the control for the same subject. Error bars are ± S.E.M.

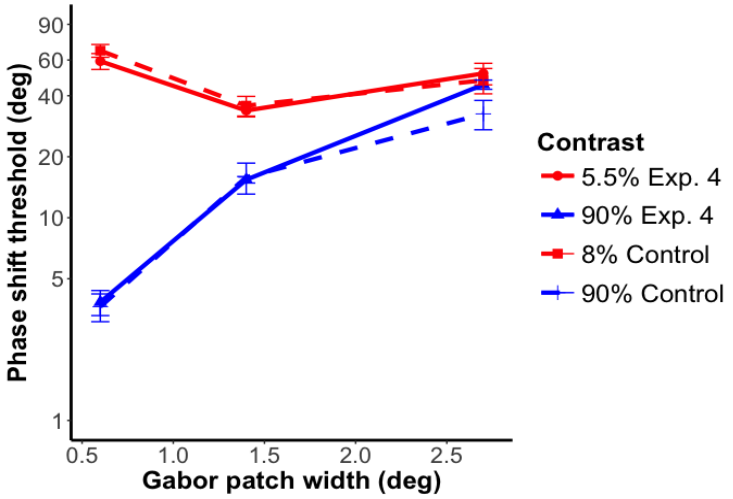

Fig. A-2. Results of the no-square control condition for Experiment 4. Solid lines represent the results from Experiment 4 whereas dashed lines show the results from the control. The data come from one subject. Error bars are ± S.E.M.

The *retinotopic reference-frame and non-retinotopic drift information condition* and the *non-retinotopic reference-frame and retinotopic drift-information condition* showed chance performance due to the mismatch between the reference-frame system and the motion information. Fig. A-3 illustrates the average percent correct across observers (N=6) as a function of phase-shift value for a low-contrast stimulus (5.5%).

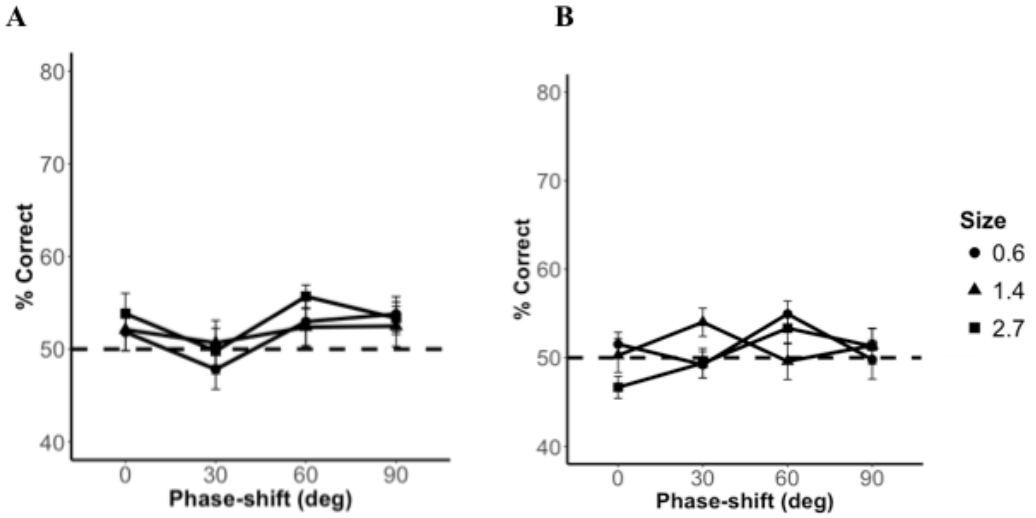

Fig. A-3. (A) Average percent correct across observers (N=6) is shown as a function of phase-shift δ value for the low-contrast stimulus in the retinotopic reference-frame and non-retinotopic drift information condition. **(B)** Average percent correct across observers (N=6) is shown as a function of phase-shift δ value for the low-contrast stimulus in the non-retinotopic reference-frame and retinotopic drift-information condition. Error bars are ± S.E.M.

Appendix B

B.1 Integration, signal-to-noise ratio, and resolution

We will illustrate the relationships between integration, signal-to-noise ratio, and resolution with a simple example. Assume *x* is the signal and *n* is the noise. Assume that noise is additive and zero mean, leading to the measured signal expressed as $\tilde{x} = x + n$. Integration is a low-pass filter and produces averaging when divided by the integration interval (Gille et al., 1981). If we integrate the signal over space and divide by the length of the integration interval, we obtain $\frac{1}{|space|}\int_{space}\tilde{x} = \frac{1}{|space|}\int_{space}(x+n) \approx \bar{x} + \bar{n}$, where $\bar{x}$ and $\bar{n}$ denote the mean values of these variables. Since $\bar{n} = 0$, we obtain $\frac{1}{|space|}\int_{space}\tilde{x} \approx \bar{x}$. In other words, integration cancels noise and provides the mean of the signal. Since the signal is averaged, i.e., low-pass filtered, spatial or temporal resolution (depending on which dimension the integration is carried on) is reduced.

B.2 Multiplicative vs Additive Models of Membrane Potential

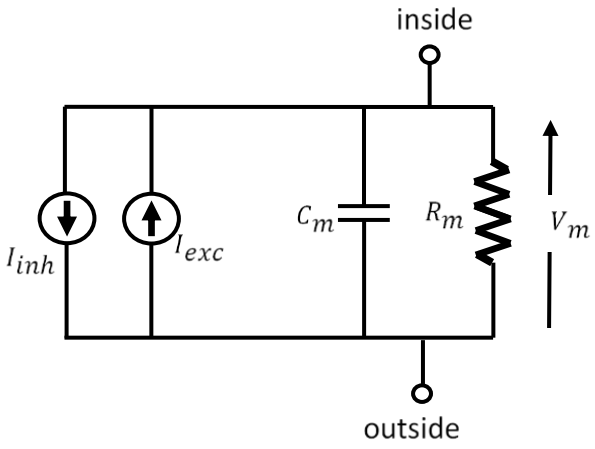

Fig. B-1. Equivalent electrical circuit of the additive model of membrane potential. Contrary to the multiplicative models, the additive models of membrane potential do not achieve normalization.

As mentioned in Section 3.2.1 and as it can been seen from Eqn. (2), in the multiplicative model of membrane potential the inputs are multiplied by the membrane potential on the right-hand side of the differential equation. This property leads to the fact that the inputs do appear in the denominator of the steady-state equation thereby allowing nonlinear processing, including contrast normalization. A simpler model for membrane potential neglects the variable-nature of conductances across the membrane and models inputs as current-sources into a passive RC-circuit as shown in Fig. B-1. From this equivalent electrical circuit, we can derive the following membrane equation:

$$C_m \frac{dV_m}{dt} = -\frac{1}{R_m} V_m + I_{exc} - I_{inh} \qquad (B1)$$

Generalizing this membrane voltage to a neuron with activity x_i

$$\frac{dx_i}{dt} = -Ax_i + \quad I^*_{exc} \; - \; I^*_{inh} \qquad (B2)$$

where $A = \frac{1}{C_m R_m}$, $I^*_{exc} = \frac{1}{C_m} I_{exc}$ and $I^*_{inh} = \frac{1}{C_m} I_{inh}$.

This is known as the additive model of membrane potential or equivalently the leaky-integrator model for the membrane potential. It is a constant-coefficient linear first-order differential equation with the steady-state:

$$x_i = \frac{1}{A}(I^*_{exc} \; - \; I^*_{inh}) \qquad (B3)$$

Hence the output is a linear function of its inputs and this equation itself cannot achieve normalization.

B.3 Sensitivity Analysis

Most parameters of the model are not critical to the predictions. They are basically scaling parameters, meaning that a change in these parameters does not compromise the behavior and dynamics of the model other than changing the operational space of the other parameters and/or the scaling of the results. There are, however, a set of parameters/mechanisms that are critical for the behavior of the model (see Table 3-1). Specifically, the spread of the Gaussian center-surround receptive field and the center-surround non-linearities of the second stage of the model are critical to the final predictions. For these parameters of interest we conducted a sensitivity analysis.

Methods: The sensitivity analysis was performed by varying the original value of the selected parameter in a percentage range while keeping the rest of the parameters fixed. This was done systematically to each of the selected parameters. The final results for each analyzed parameter were plotted against the original fit in a single plot. The summary of the model parameters is shown in Table 3-1.

Results:

σ_c*: Spread of the center-Gaussian receptive field.* The original value of the spread of the center-Gaussian receptive field was changed in the range of $\pm 15\%$. The results are shown in Fig. B-2.

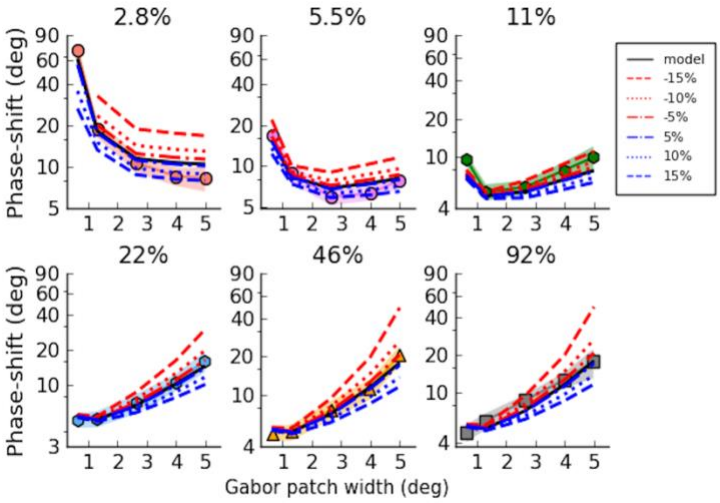

Fig. B-2. Sensitivity analysis for parameter σ_c. Blue and red curves denote the percentage increments and decrements of the parameter from its original value.

As the σ_c parameter is increased (blue lines), the threshold predictions dropped. This is a result of the larger pool of excitatory signals that come as a result of the increase in the spread of the center Gaussian. This is especially evident at the lowest contrast where the thresholds for spatial summation are lower at +15% compared to the rest of values. Also, at higher contrast values a decrease in the value of σ_c is reflected in an increase of the thresholds at large stimulus sizes (due to a larger contribution of the inhibitory surround signals).

σ_s: Spread of the surround-Gaussian receptive field. The original value of the spread of the surround-Gaussian receptive field was also changed in the range of $\pm 15\%$. The results are shown in Fig. B-3. The effect of an increase/decrease of the spread of the surround Gaussian is especially manifested in the increase/decrease of the spatial summation thresholds at the lowest contrast level. At mid-high contrast values, changes in the spread of the surround Gaussian has no major effect on the results.

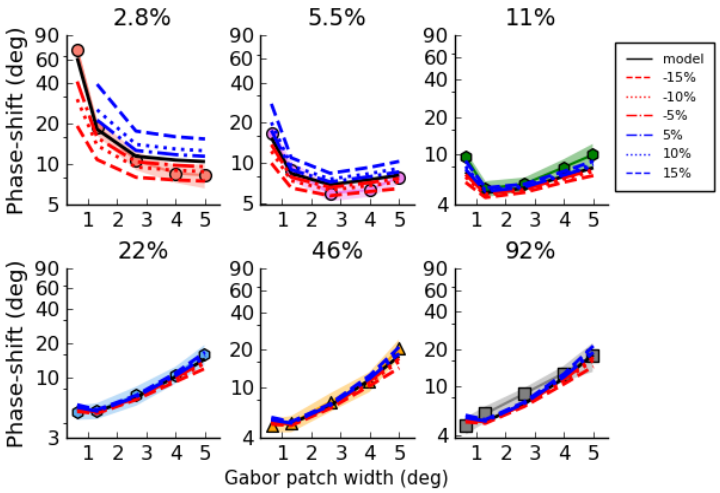

Fig. B-3. Sensitivity analysis for parameter σ_s. Blue and red curves denote the percentage increments and decrements of the parameter from its original value.

a_c: *Midpoint of the sigmoidal center non-linearity.* This is a critical parameter for the model's predictions. The a_c parameter controls the rightward/leftward shifts of the center non-linearity. At low contrast values, in particular, small rightward/leftward ($\pm$3%) shifts of the non-linearity change the sensitivity of the model. This is because at low contrasts values, the contrast normalized activity of the network operates in the transitional region of the non-linearity where small shifts to the midpoint parameter have significant effects on the output of the non-linearity (see the 2.8% vertical grey line of the contrast-response function in Fig. B-4, left panel).

For instance, rightward shifts (+) of the center nonlinearity result in a reduction of the output activity of the non-linearity which ends up in a decrease of the excitatory activity with a resulting increase in the output's thresholds (blue curves in Fig. B-4). The opposite occurs for a leftward shift (red curves in Fig. B-4). At mid-to-high contrast, however, changes to $\boldsymbol{a_c}$ do not affect the output of the model as shown in the lower panels of Fig B-4.

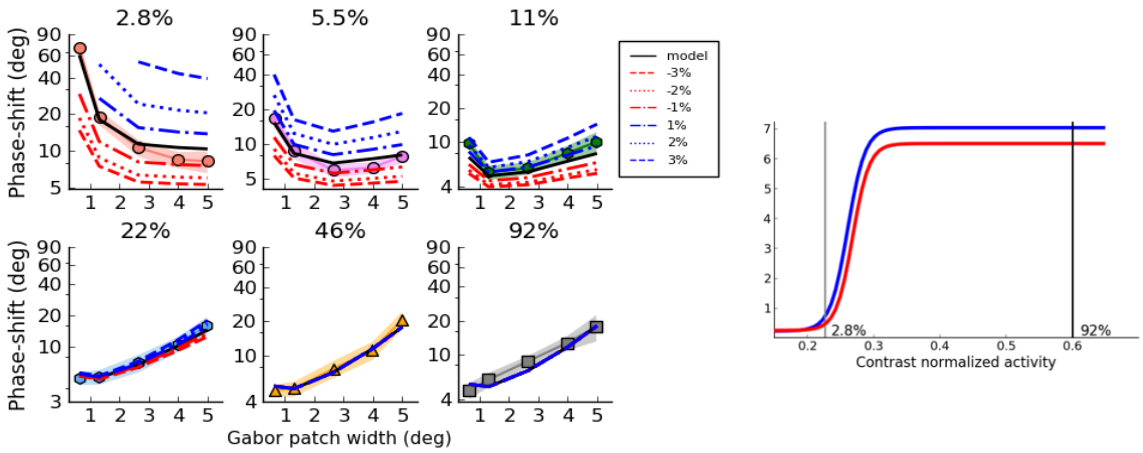

Fig. B-4. Sensitivity analysis for parameter a_c. Right) Results from the sensitivity analysis. Blue and red curves denote the percentage increments and decrements of the parameter from its original value. Left) Center (blue) and surround (left) non-linearities of the model. The vertical grey and black lines show the approximated normalized activity for an input pattern set at 2.8% and 92% contrast, respectively.

$\boldsymbol{a_s}$***:*** *Midpoint of the sigmoidal surround non-linearity.* Similarly to the $\boldsymbol{a_c}$ parameter, the $\boldsymbol{a_s}$ is a critical parameter for the model's predictions. This time, however, rightward shifts of the sigmoidal non-linearity produce a decrease in inhibition and therefore a drop in the thresholds as shown by the blue lines in Fig. B-5.

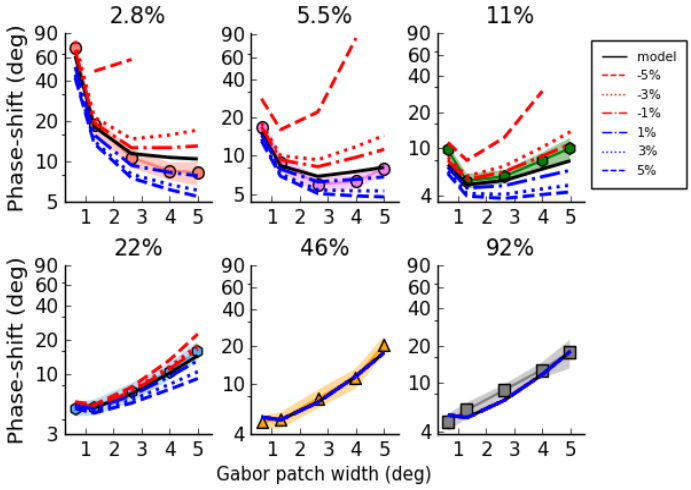

Fig. B-5. Sensitivity analysis for parameter a_s. Blue and red curves denote the percentage increments and decrements of the parameter from its original value.

b_c: Slope of the center non-linearity. The slope parameter of the non-linearities is another critical parameter in the model. Changes in b_c ($\pm 15\%$) produced strong variations in the outputs of the model. The explanation is similar to the midpoint parameters. Basically, changes in the slope of the non-linearity affects the results at low contrast, but not at mid-to-high contrast values. Increases in the slope of the center non-linearity yield a decrease in the output of the nonlinear function. This, in turn, produces lower excitatory activity and higher threshold outputs (blue curves in Fig. B-6).

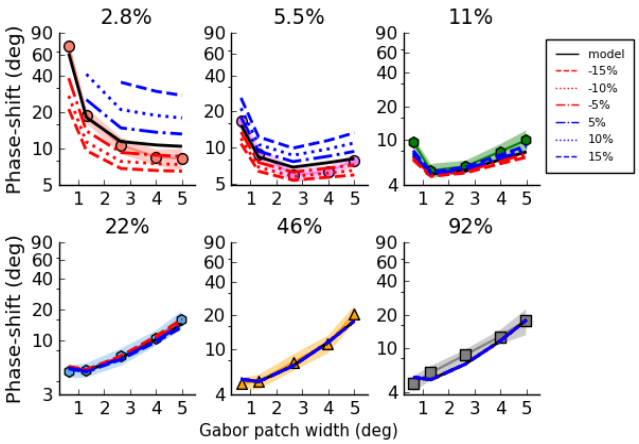

Fig. B-6. Sensitivity analysis for parameter b_c. Blue and red curves denote the percentage increments and decrements of the parameter from its original value.

b_s: Slope of the surround non-linearity. A similar argument to the case of the slope of the center non-linearity can be raised when changes ($\pm 15\%$) in the slope parameter of the surround non-linearity are introduced. This time, however, increases in the slope of the surround non-linearity yield a decrease in the output of the nonlinear function. This, in turn, produces lower inhibitory activity and, as a consequence, lower threshold outputs (blue curves in Fig. B-7).

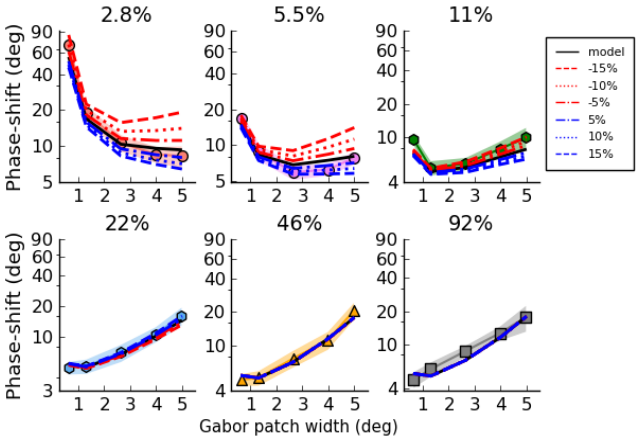

Fig. B-7. Sensitivity analysis for parameter b_s. Blue and red curves denote the percentage increments and decrements of the parameter from its original value.

Amp_c: Saturation of the center non-linearity. Changes in the Amp_c are not critical for the performance of the model. They mainly produced changes in the scale of the outputs. An increment in the value of Amp_c generates a drop in the thresholds (blue curves in Fig. B-8) which is a product of the increase in the output of the center non-linearity. The opposite occurs with decreases in Amp_c (red curves in Fig. B-8). To see the effects on performance of other non-critical parameters we refer the reader to the supplementary materials.

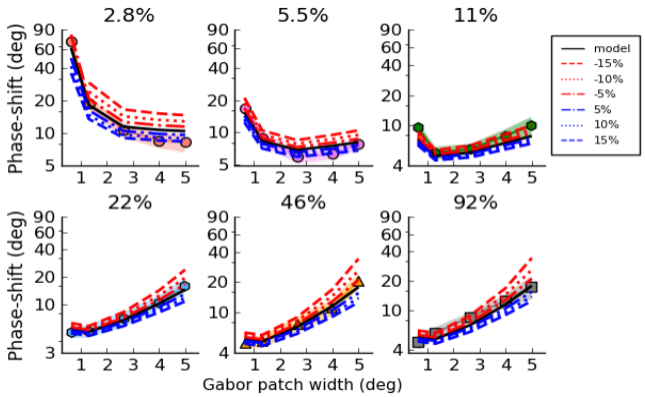

Fig. B-8. Sensitivity analysis for parameter Amp_c. Blue and red curves denote the percentage increments and decrements of the parameter from its original value.

Amp_s: Saturation of the surround non-linearity. Similar to Amp_c, changes in Amp_s mainly yield variations in the scale of the outputs. An increment in the value of Amp_s generates an increase in the thresholds (blue curves in Fig. B-9) which is a product of the increase in the output of the surround non-linearity (which generates higher inhibition in the model). The opposite occurs with decreases in Amp_s (red curves in Fig. B-9).

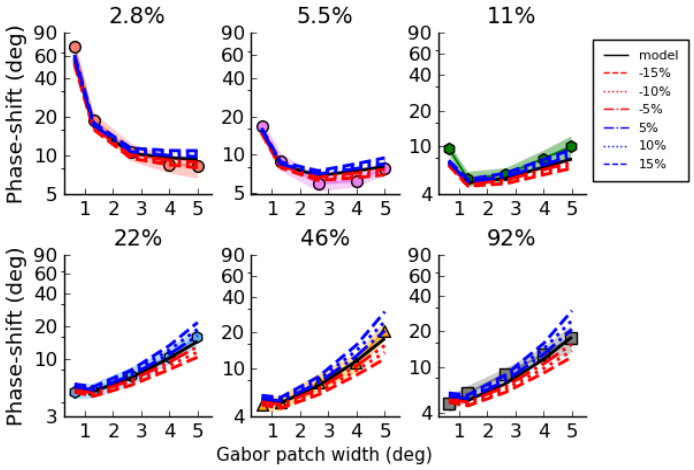

Fig. B-9. Sensitivity analysis for parameter Amp_s. Blue and red curves denote the percentage increments and decrements of the parameter from the original value.

B': Nernst potential for depolarization. As an illustration on the effect of varying a non-critical parameter, we changed the Nernst potential for depolarization (B') in a range of $\pm 15\%$. As a result, we observe that changes in this parameter mainly generate changes in the scale of the results, but do not affect the predictions of the model.

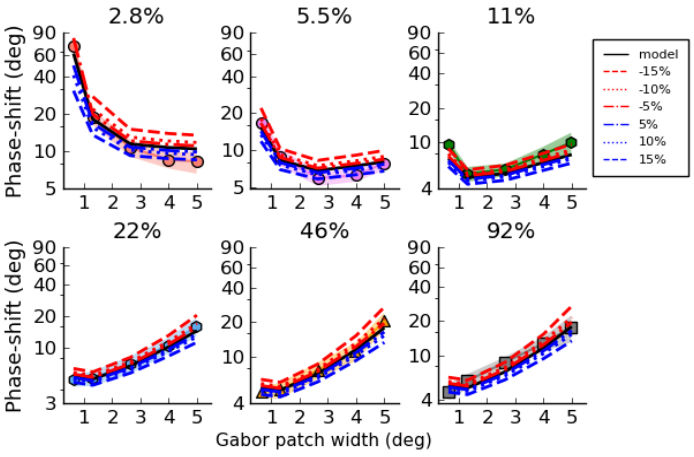

Fig. B-10. Sensitivity analysis for parameter B'. Blue and red curves denote the percentage increments and decrements of the parameter from the original value.

τ: Time-scaling parameter. This parameter is a constant that scales the derivative and hence controls the response-time. We changed this parameter to values ranging from 0.00375 to 4 (the original value was 1) and the simulation duration was equivalent to 100 ms. As shown in Fig. B-11, this variable does not affect the behavioral results of the model. However, it does affect how fast neural responses change and, as mentioned in Section 3.4.5, a value of 0.00375 gave better fits to neural data.

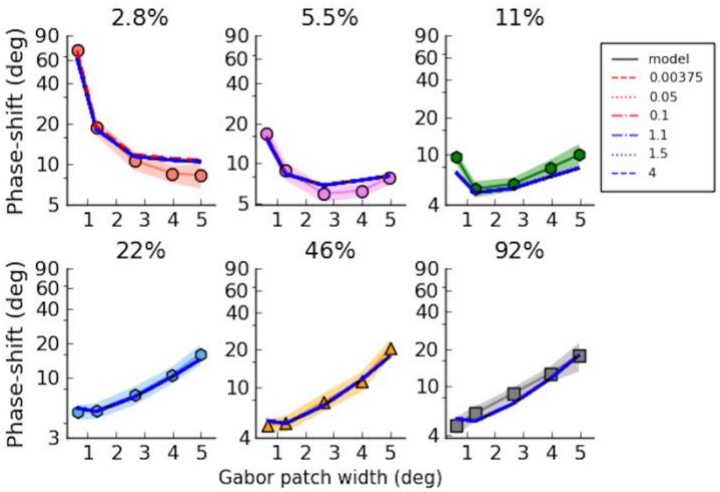

Fig. B-11. Sensitivity analysis for parameter τ. Blue and red curves denote the parameter's increments and decrements from the original value, respectively. Changes in this parameter did not affect the predictions of the model.

B.4 Adding Gaussian noise:

As a final analysis, we proceeded to add Gaussian noise to the membrane potential of the contrast normalization layer following the equation:

$$\frac{1}{\tau}\frac{dx_i}{dt} = -Ax_i + (1 - x_i)\sum_{j=i-n}^{i+n} K_{j-i}^{center} \cdot I_j - x_i \sum_{j=i-n}^{i+n} K_{j-i}^{surround} \cdot I_j + \lambda \cdot n(t) \qquad (B4)$$

where λ represents the amplitude of the noise and $n(t)$ is a random variable following a standard normal distribution (zero mean and a standard deviation of one). The rest of the parameters where identical to the previous simulations. The noise amplitude was varied according to a percentage of the membrane potential value when no noise was present. For instance, the plot in Fig. B-12A shows the noisy membrane potential of a neuron in the contrast normalization layer when noise was set to 5% and 30%, respectively. Fig. B-12B shows the output of the model evaluated at two contrast values and three stimulus sizes when noise was only introduced to the contrast normalization layer. As seen, the network is largely insensitive to noise, especially at high contrast (right plot) where the non-linearities of the center and surround receptive fields in the second layer operate in their saturation region. Thus, at high contrast, noisy changes in the membrane potential of the first layer do not impact the results. At low contrast values (left plot), however, noisy variations in the contrast normalization layer produced small deviations in the output mostly due to the nonlinearities which, at low-contrasts, operate in their transitional regions where small changes in the input can have a significant impact on the output of the nonlinearity. Lastly, simulations were run when Gaussian noise was added to the second layer only. A similar equation as in Eqn. B4 was followed, but we used the membrane potential equation for the adaptive center-surround layer instead. Fig. B-12C (left) shows the output of a neuron in the second layer evaluated at low and high noise levels. In Fig. B-12C (right), at low-contrast stimulus condition, we evaluated the response of the model to low and high noise levels. Aside from small deviations in the outputs, the results showed that the model is mainly unaffected by Gaussian noise. Similar results were obtained at high-contrast values.

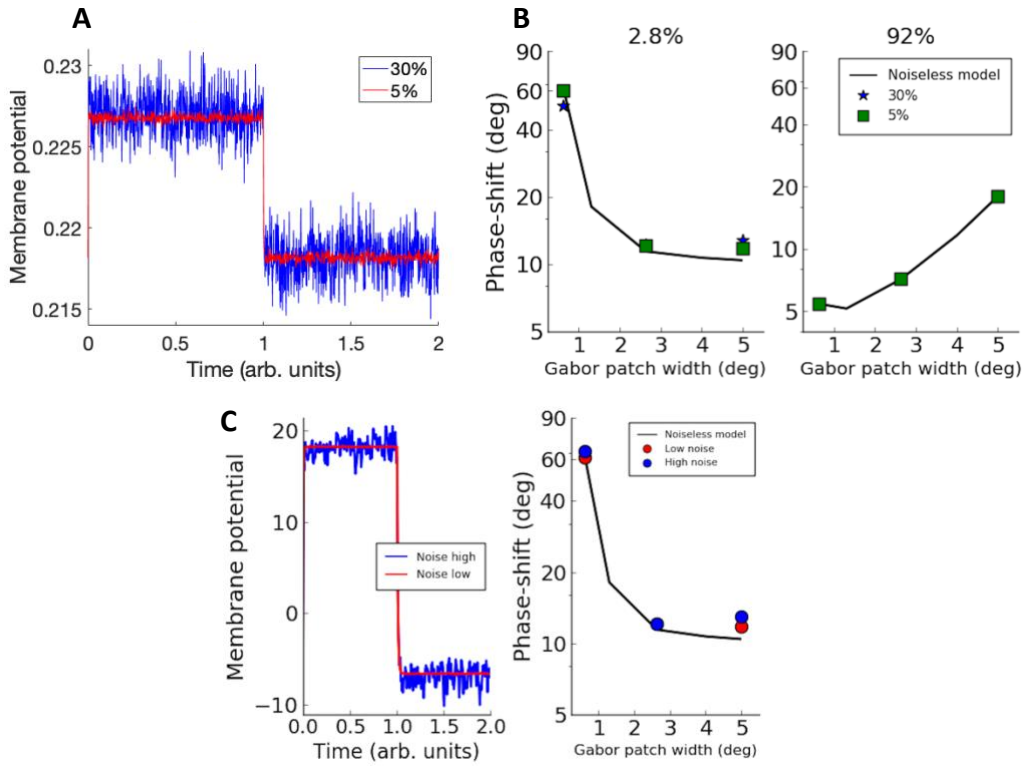

Fig. B-12. Simulations with Gaussian noise. A Example of the response of a neuron in the contrast normalization layer to two noise levels. **B** Output of the model to noise variation in the first layer evaluated at low contrast (left) and high contrast (right). **C** Noisy membrane potential of a neuron in the second layer (left) and output of the model to noise changes in the second layer evaluated at low contrast (right).

B.5 Model comparison

Here we compare our model to the size model and the gain model described in Tadin & Lappin (2005) and Betts et al. (2012), respectively. Analytically, our architecture differs from these models in a few important ways. First, in these models contrast is captured by hand, i.e. the stimulus contrast is manually introduced to a Naka-Rushton function of the form:

$$K(c) = A\frac{c^n}{c^n + c50^n} \qquad (B5)$$

In our model, however, the stimulus contrast is computed by the contrast normalization stage (Eqn. 8) which at steady state ($\frac{x_i}{dt} = 0$) and with the selected parameters, gives the steady-state response:

$$x_i = \frac{\sum_{j=i-n}^{i+n} K_{j-i}^{center} . I_j}{0.5 + \sum_{j=i-n}^{i+n} K_{j-i}^{center} . I_j + \sum_{j=i-n}^{i+n} K_{j-i}^{surround} . I_j} \qquad (B6)$$

This equation generates an automatic gain control and bounds the response of the system to values between 0 and 1.

Second, these models are limited in terms of spatial representation because they compute the neural response at a single point in space. For example, the response of the system to a stimulus of size *w* is estimated as the integral of the Gaussian that overlaps with the stimulus, as provided by the *erf function*. Our network, on the other hand, captures the response of the neurons in space by calculating the 1D spatial convolution of the input signals with the receptive field kernels following the convolution operation:

$$K * I = \sum_{j=i-n}^{i+n} K_{j-i} \cdot I_j \qquad (B7)$$

Third, the responses of these models are exclusively given in terms of either subtractive inhibition or divisive inhibition. For example, for the size model (Tadin & Lappin, 2005), the response of the model is equal to:

$$R = R_0 + E - I \qquad (B8)$$

where R_0 is a baseline response, and the excitatory response (E) and the inhibitory response (I) vary in terms of size and contrast following the equations:

$$E(w, c) = K_e(c) \cdot \mathrm{erf}\left(\frac{w}{\alpha(c)}\right)$$

$$(B9)$$

$$I(w,c) = K_i(c) \cdot \text{erf}\left(\frac{w}{\beta}\right)$$

The function $\alpha(c) = \frac{S}{1+m \cdot e^{-\frac{k}{c}}}$ is a decreasing logistic function used to estimate the rate at which α changes with stimulus contrast. For the gain model (Betts et al., 2012), on the other hand, the response of the system is provided in terms of divisive inhibition as in:

$$R(w,c) = \frac{K_e(c) \cdot E(w)}{1 + K_i(c) \cdot I(w)} \qquad (B10)$$

Here, $K(c)$ is the contrast response coming from the Naka-Rushton function, and $E(w) = \text{erf}\,(\frac{w}{\alpha})$ and $I(w) = \text{erf}\,(\frac{w}{\beta})$ are the integral of the portion of the Gaussian activated by the stimulus of size *w*. In contrast to these models, the steady-state response of our architecture has both subtractive and divisive inhibition terms that result from the dynamics of the equation. For instance, solving Eqn. (11) for $\frac{y_i}{dt} = 0$, we obtain:

$$y_i = \frac{B \sum_{j=i-n}^{i+n} G_{j-i}^{center} \cdot f(x_j) - D \sum_{j=i-n}^{i+n} G_{j-i}^{surround} \cdot g(x_j)}{A + \sum_{j=i-n}^{i+n} G_{j-i}^{center} \cdot f(x_j) + \sum_{j=i-n}^{i+n} G_{j-i}^{surround} \cdot g(x_j)} \qquad (B11)$$

Note that the inhibitory term from the surround neurons is present in the numerator as subtractive inhibition and in the denominator as divisive inhibition. Lastly, while the output of these models represents the steady-state response of the system, our model encompasses the whole dynamic of the response which makes our model more realistic than the existing ones.

We also performed goodness of fit analysis. The models were fit to the log-transformed group average phase-shift thresholds given in Fig. 3-6. Fits were evaluated using a χ^2 goodness of fit test with $N - p$ degrees of freedom, where N equals the number of data

points and p is the number of free parameters. Additionally, we also calculated an adjusted R^2 to assess the fits. All models generated good fits to the data with $\left[\chi^2(22) = 5.24, p > 0.99;\ R^2_{adj} = 0.94\right]$ for our current model, $\left[\chi^2(21) = 4.32, p > 0.99;\ R^2_{adj} = 0.97\right]$ for the gain model, and $\left[\chi^2(20) = 3.84, p > 0.99;\ R^2_{adj} = 0.95\right]$ for the size model, respectively. For visualization, Fig. B-13 shows the fits of each model to the data. Hence, in terms of motion discrimination data, the models perform similarly. However, as shown in the manuscript, our model is also capable of capturing motion integration/segregation data. To the best of our knowledge, our model is the only computational model applied to integration/segregation data of Fig. 3-11.

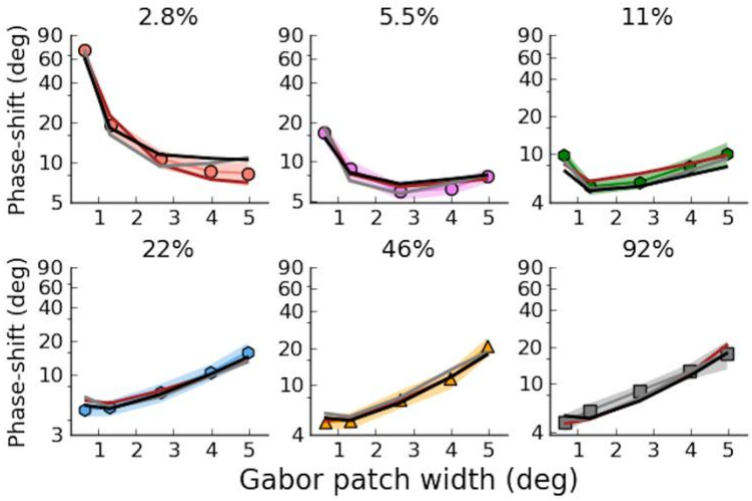

Fig. B-13. Model comparison. The solid black lines show our simulation results. The grey solid lines represent the fits of the gain model and the brown solid lines represent the fit of the size model.

B.6 Model's response to a Gaussian pattern

To demonstrate that the model brings about similar results when the input pattern is different than a rectangular signal, we ran simulations using a Gaussian envelope as input.

We evaluated the model at three input sizes and at two contrast conditions. The results for the rectangular input pattern and the Gaussian envelop are shown in Fig. B-14. As observed, regardless of the shape of the input pattern the model generates similar results.

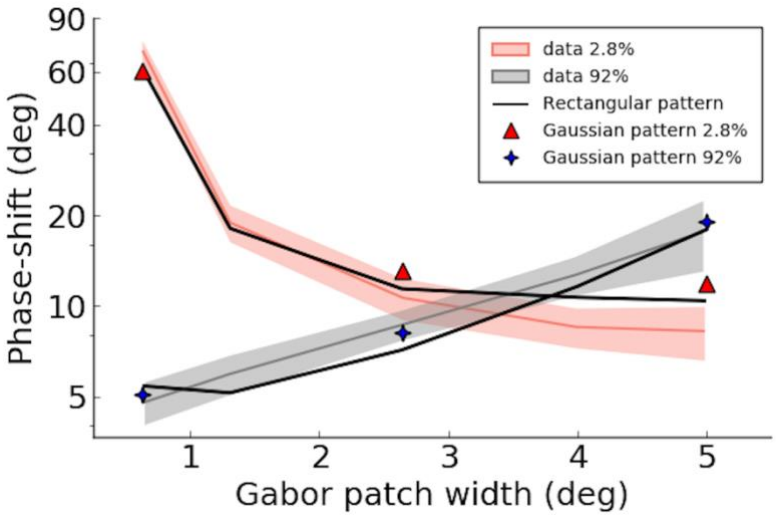

Fig. B-14. Model responses to Gaussian pattern. The solid black lines show the simulation results to a rectangular pattern. The red triangles and the blue stars marks show the simulation results to a Gaussian input pattern.