The Dissertation Committee for Kathryn L Bonnen
certifies that this is the approved version of the following dissertation:

3D motion: Encoding and perception

Committee:

Alexander C Huk, Supervisor

Lawrence K Cormack , Co-supervisor

Carlos Carvalho

Ila Fiete

Wilson Geisler

Mary Hayhoe

3D motion: Encoding and perception

by

Kathryn L Bonnen

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2018

For Sarah & Asadel

Acknowledgments

I would like to acknowledge the many people who have supported me and my scientific endeavors throughout my PhD. There are more people in that number than I can name here. Without them this dissertation would not have been possible. Financial support for my studies was provided by the National Science Foundation Graduate Research Fellowship Program and the Harrington Fellowship program.

I am especially grateful for the guidance and mentorship of my advisors, Alex and Larry. It was a privilege to work with both of them. Alex's guidance smoothed my transition from engineer to scientist. And in the final years of my PhD, I have appreciated his general tolerance and (dare I say) support of my wandering interests. I am immensely grateful that Larry agreed to help a young second-year graduate student with her statistics course project. That collaboration evolved into my dissertation and gave me an excuse to sit on the very comfortable couch in Larry's office.

Thanks to my committee (Mary Hayhoe, Bill Geisler, Ila Fiete and Carlos Carvalho) who have served and continue to serve as incredible mentors, teachers and scientific collaborators. Thanks to Adam Kohn for his collaboration, and to Jonathan Pillow and Alison Preston for their mentorship during the early part of my graduate career.

I have thoroughly enjoyed working alongside the members of the Huk lab (past and present) and graduate school would have been incomplete without the

many other fabulous people across the Center for Perceptual Systems and the Institute for Neuroscience. I'd especially like to thank: Johannes Burge for his thoughtful guidance; Jake Yates and Leor Katz for helping me to take myself a bit less seriously; Jon Matthis for introducing me to a way of collecting data that involves significantly more hiking; and Sarah Nordquist, Lauren Kreeger, and Liz Arnold for the many lunches and happy hours that kept us all mostly sane. I am also grateful to the host of highschool/undergraduate/post-baccalaureate researchers who I have had the pleasure of mentoring (Devon, Skylar, Jasmine, Adebisi, Siri, Austin, Jeanette). I hope you have learned as much from me, as I have from you.

Many thanks to the communities in Austin who have given me respite from my doctoral work, especially the South Austin Community Garden, the Episcopal Student Center, First Bytes and the Training Room.

I also wish to thank my undergraduate mentors at Michigan State University, who set me on this path, especially: Dr. Anil Jain and Dr. Brendan Klare whose guidance and support led me to pursue a graduate degree; and Dr. Robert Bell, my first research mentor, who encouraged me to pursue my interests wherever they led.

To my family and close friends, thank you for being there. Thanks to Anna for being my other half during graduate school, and to Conrad for his encouragement and support. Finally thanks to my little brother, who continues to put up with me, and to my parents who have been there through it all.

Kathryn L Bonnen

The University of Texas at Austin
May 2018

3D motion: Encoding and perception

Publication No. _____

Kathryn L Bonnen, Ph.D.
The University of Texas at Austin, 2018

Supervisors: Alexander C Huk &
Lawrence K Cormack

The visual system supports perception and inferences about events in a dynamic, three-dimensional (3D) world. While remarkable progress has been made in the study of visual information processing, the existing paradigms for examining visual perception and its relation to neural activity often fail to generalize to perception in the real world which has complex dynamics and 3D spatial structure. This thesis focuses on the case of 3D motion, developing dynamic tasks for studying visual perception and constructing a neural coding framework to relate neural activity to perception in a 3D environment.

First, I introduce target-tracking as a psychophysical method and develop an analysis framework based on state space models and the Kalman filter. I demon-

strate that target-tracking in conjunction with a Kalman filter analysis framework produce estimates of visual sensitivity that are comparable to those obtained with a traditional forced-choice task and a signal detection theory analysis. Next, I use the target-tracking paradigm in a series of experiments examining 3D motion perception, specifically comparing the perception of frontoparallel motion with the perception of motion-through-depth. I find that continuous tracking of motion-through-depth is selectively impaired due to the relatively small retinal projections resulting from motion-through-depth and the slower processing of binocular disparities.

The thesis then turns the neural representation of 3D motion and how that underlies perception. First I introduce a theoretical framework that extends the standard neural coding approach, incorporating the environment-to-retina transformation. Neural coding typically treats the visuals stimulus as a direct proxy for the pattern of stimulation that falls on the retina. Incorporating the environment-to-retina transformation results in a neural representation fundamentally shaped by the projective geometry of the world onto the retina. This model explains substantial anomalies in existing neurophysiological recordings in primate visual cortical neurons during presentations of 3D motion and in psychophysical studies of human perception. In a series of psychophysical experiments, I systematically examine the predictions of the model for human perception by observing how perceptual performance changes as a function of viewing distance and eccentricity. Performance in these experiments suggests a reliance on a neural representation similar to the one described by the model.

Taken together, the experimental and theoretical findings reported here advance the understanding of the neural representation and perception of the dynamic 3D world, and adds to the behavioral tools available to vision scientists.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Scientists and philosophers have long sought to understand the relationship between the external world and visual experience, seeking to understand perception as a set of statistical inferences about that external world given some internal representation (Fechner, 1860; von Helmholtz, 1867). Remarkable progress has been made in describing the perception of 2D motion (frontoparallel motion) and the neural activity that underlies that perception, particularly in the context of discrete perceptual decisions (Maunsell & Newsome, 1987; Born & Bradley, 2005). However in the real world, motion is rarely restricted to a frontoparallel plane and motion perception is a continuous process often leveraged as part of a sensorimotor control loop.

This thesis examines 3D motion perception from multiple perspectives, from its role in sensorimotor control loops to the neural mechanisms that underlie its perception. It begins with an examination of the perception of 3D motion using a novel psychophysical paradigm (Chapters 2-3) and then discusses the value of such naturalistic behavioral paradigms in the context of moving beyond the notion of an IID (independently and identically distributed) trial (Chapter 4). Chapters 5-6 examine the neural mechanisms underlying 3D motion perception. I introduce a neural coding model for the representation of 3D motion in primate cortical neurons (Chapter 5), which is then linked to human perception via a series of 3D motion estimation experiments (Chapter 6). Appendix A evaluates speed discrimination across monocular fields, examining a motion cue that is potentially critical to 3D motion perception during self-motion. The general discussion (Chapter 7) integrates this body of work with a focus on potential future experiments.

The purpose of this first chapter is to provide background relevant to the content of this thesis. This introduction consists of 6 sections. The first three (1.1-1.3) focus on psychophysics and provide short introductions to signal detection theory and ideal observer models, bodies of theoretical work that facilitate the analysis and interpretation of psychophysics. Sections 1.4 and 1.5 then give a short overview of the the neurophysiological work relevant to motion perception and an introduction to the basic neural coding models which this thesis builds upon. The final section (1.6) provides an overview of the existing perceptual and neurophysiological work on 3D motion.

1.1 Psychophysics

Fechner published Elemente der Psychophysik in 1860, founding the field of psychophysics (Fechner, 1860). Driven by a dual-aspect monistic view of the mind-body problem, he sought to discover the laws that governed the relationship between the physical external world and internal psychical world. His great contribution was recognizing that measurement would be key to the discovery of these laws and that a "just noticeable difference" in the physical stimulus could serve as an indirect unit of measurement (Wozniak, 1999). Part 1 of the Elemente der Psychophysik, introduced the classical methodologies for measuring the relationship between a physical stimulus and an observer's perception of that physical stimulus: method of adjustment, method of limits, and method of constant stimuli. He applied these methods to measure a variety of perceptual phenomena (e.g. lifted weights, visual brightnesses, tactile and visual distances). These methods became central to the study of perception.

The methodologies proposed by Fechner measured the difference threshold (i.e. the just noticeable difference or *jnd*) which is the minimal change to the physical stimulus that can be detected/discriminated by an observer (Weber, 1834). The thresholds reported in modern psychophysical experiments are the stimulus differences corresponding to a particular level of performance (e.g. 75% correct). The three classical methodologies were: 1) *The method of limits* – On each trial the value of the stimulus is varied in small ascending or descending steps. For each step the observer reports whether the stimulus is smaller than, equal to, or larger than the standard stimulus. 2) *The method of constant stimuli* – Each trial consists of the presentation of a single stimulus value. Over the course of the experiment observers on a set of predetermined stimulus values that range from those that definitely can't be distinguished from the standard to those that are easily distinguished. Here threshold is calculated using the collected data to estimate a psychometric function. 3) *The method of adjustment* – the observer adjusts the value of the stimulus and sets its value equal to the standard stimulus (Fechner, 1860; Treutwein, 1995).

One of the criticisms of these psychophysical approaches is that they are relatively inefficient and require a huge number of measurements (Treutwein 1995; see also Introduction of Chapter 2). This is particularly true for the method of constant stimuli which became the most prominent method for two important reasons:

1) it did not suffer from the measurement biases observed in the method of limits and the method of adjustment and 2) the eventual development of signal detection theory provided a mathematical framework for properly measuring sensitivity and bias in psychophysical responses. The huge number of measurements required by this form of psychophysics motivated the development of adaptive staircase methods (Dixon & Mood, 1948) and Bayesian and maximum-likelihood procedures (e.g. Quest, Watson & Pelli 1983; ML-TEST, Harvey 1986; and ZEST, King-Smith et al. 1993) for more efficient estimation of thresholds.

1.2 Signal Detection Theory

The invention of Signal Detection Theory (SDT) was a key advancement for the field of psychophysics. Originally developed for applications in radar technology (Peterson et al., 1954), it was adapted for the analysis and interpretation of psychophysics. The main assumption in SDT for psychophysics is that the perceptual decision takes place in the presence of uncertainty. Take a simple detection task in which observers are asked to report whether the target stimulus is present or absent. The stimulus (target present or target absent) results in a noisy internal response. From that internal representation of the stimulus, the observer reports "yes" they saw the stimulus or "no" they did not. Taken together, the presentation of the stimulus and the observer response results in one of four possible outcomes: hit (target present, observer reports yes), miss (target present, observer reports no), correct rejection (target absent, subject reports no), or false alarm (target absent, subject reports yes). Figure 1.1 depicts two hypothetical internal response curves for this simple detection task. The choice of the decision criterion determines the relative proportions of hits/misses and correct rejections/false alarm. By measuring the hits and the false alarms, one can calculate the observer's decision criterion and estimate the noise present during the decision process, separating observer's response biases from measures of observer sensitivity.

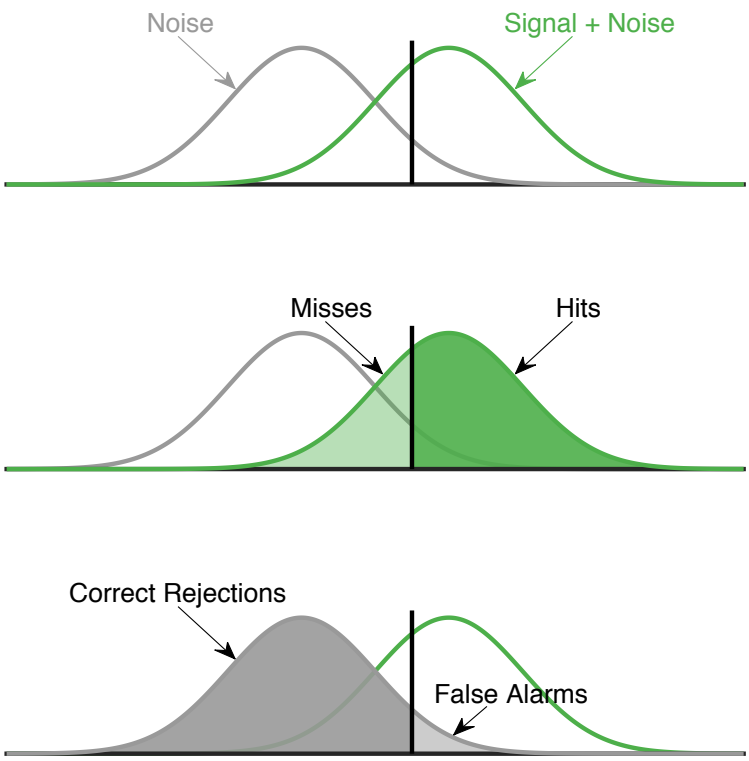

Figure 1.1: Signal Detection Theory. The two distributions for a classic detection experiments are depicted in the top panel: the 'Signal + Noise' (i.e., signal present) and the 'Noise' (i.e., signal absent). The criterion (vertical black line) determines the relative proportion of Misses/Hits and Correct Rejections/False Alarms (middle and lower panel).

An extensive body of work has focused on connecting signal detection theory to the study of sensory systems and the neural response underlying perception, and extending it to better model perceptual decision and their underlying neural activity (Ratcliff, 1978; Smith & Ratcliff, 2004; J. Palmer et al., 2005; Ratcliff & McKoon, 2008). This is discussed in much greater detail in Chapter 4.

1.3 Ideal Observer Analysis

Ideal observer analyses have played a critical role in expanding the understanding of vision and perception. In vision science the ideal observer is typically designed to perform a specific task or set of tasks. It has access to all available stimulus information and performs optimally given the information and any additional constraints. These hypothetical devices are valuable for a number of reasons (laid out in greater detail by Geisler 2011), including: 1) Identifying the task-relevant stimulus properties. 2) Explaining how to use those task-relevant stimulus properties to perform the specified task. 3) Providing a point of comparison (or baseline performance expectation). 4) Generating hypotheses and models of real performance on tasks.

The earliest use of an ideal observer in vision science focused on the problem of visual detection of luminance increments or decrements and how visual detection is limited by photon noise. This early work compared the performance of real observers with ideal observers limited only by photon noise (Rose, 1948; Barlow, 1957; de Vries, 1943; Cohn & Lasley, 1974). Ideal observer theory has since been applied to a wide variety of tasks including acuity-related tasks, contrast sensitivity, contrast discrimination, color discrimination, letter identification, contour grouping, cue integration, motion, attention, etc (see Geisler 2011 for a detailed review).

Although the notion of the ideal observer is really quite general, most ideal observers in vision science are built upon the framework of traditional psychophysics and signal detection theory, with some notable exceptions focusing on sensorimotor control primarily in the context of reaching (Baddeley et al., 2003; Todorov, 2005; Körding & Wolpert, 2006; Burge et al., 2008a). Chapter 2 introduces an Bayesian ideal observer for a novel tracking paradigm, demonstrating that visual sensitivity can be measured via performance on sensorimotor tasks.

1.4 Neurophysiology of 2D motion

The visual pathways governing the neural processing of frontoparallel motion (i.e. 2D motion) in primates has been studied in great detail (see Maunsell & Newsome 1987 or Born & Bradley 2005 for reviews). Evidence of the neural processing of motion is observed first in the primary visual cortex (V1), where a subset of both simple and complex cells are direction-selective. However, in the middle temporal

area (MT) more than 90% of neurons exhibit direction selectivity (Zeki, 1974). This is part of the reason that MT is often described as central to motion processing in the primate brain.

MT receives its primary input from neurons in layer 4 of V1; although only a subset of neurons in V1 are direction-selective, the neurons that project from V1 to MT are predominantly direction-selective (Movshon & Newsome, 1996). MT does also receive input from cortical areas V2 and V3, as well as subcortical pathways via the superior colliculus (Born & Bradley, 2005). Neurons in MT are retinotopically organized preserving the spatial organization of the visual field (Maunsell & Van Essen, 1983b). The receptive fields of MT neurons increase in size as a function of eccentricity. Each neuron's receptive field diameter is roughly equivalent to the eccentricity of the receptive field center (Mikami et al., 1986). Furthermore, MT neurons are selective to motion direction and speed (Zeki, 1974; Maunsell & Van Essen, 1983b; Albright, 1984). Such selectivity is well-described by von Mises tuning functions in the case of motion direction, and log-Gaussian tuning in the case of motion speed (Nover et al., 2005).

The neurophysiological work I have summarized here, along with the work described in the subsequent two sections (1.5-1.6) of this introduction forms the foundation for the neural coding model described in Chapter 5, which seeks to extend existing neural coding models for 2D motion to 3D motion.

1.5 Neural Coding

The neural coding approach works to find a probabilistic relationship between a stimulus and pattern of neural activity (see Pouget et al. 2003 for review). Ultimately this allows one to perform encoding (i.e., characterizing the neural activity resulting from a particular stimulus) and to perform decoding (i.e. estimating the stimulus from a pattern of neural activity). These approaches are possible because populations of neurons respond to single features of the world (i.e., variables) and their responses vary systematically as a function of changes to that single feature (e.g., orientation in primary visual cortex, Hubel & Wiesel 1959; 2D motion direction in middle temporal area, Maunsell & Van Essen 1983b; wind direction in the cricket cercal system, J. P. Miller et al. 1991; human faces in inferotemporal cortex, Perrett et al. 1985, etc.).

A classic example of the neural coding approach was applied to the cricket cercal system (J. P. Miller et al., 1991; Theunissen & Miller, 1991). Hair cells on the cricket cerci are sensitive to wind direction, meaning the cells respond differentially as a function of the direction of air current. The mean response of a cercal cell as a function of wind direction (i.e. the tuning curve) is well-approximated by a rectified cosine:

$$f_i(s) = r_i^{max} \left(cos\left(s - s_i\right)\right)^+, \quad where\ (x)^+ = \begin{cases} x, & x > 0 \\ 0, & otherwise \end{cases} \tag{1.5.1}$$

where s is the wind direction presented to the system, s_i is the preferred direction for hair cell i and r_i^{max} is the maximum firing rate of that cell. A full encoding model for a cercal cell's response to wind direction includes the equation above and a noise model (typically Poisson). A Poisson-independent maximum likelihood estimator for wind direction in cercal cells relies on the neurons' tuning curves and takes into account the Poisson noise model, assuming that the noise is independent across cells. The likelihood is given by:

$$P(\mathbf{r}|s) = \prod_i = 1^n e^{f_i(s)\Delta t} * (f_i(s)\Delta t)^{r_i \Delta t} * \frac{1}{(r_i \Delta t)!} \tag{1.5.2}$$

where $\mathbf{r}$ is the vector of cell responses. The maximum likelihood estimator finds the value of s that maximizes $P(\mathbf{r}|s)$. The cercal system relies on the response of just four cells evenly spaced along the wind direction axis. In the primate visual system, such maximum likelihood decoders are applied to systems with many more neurons spread more finely across the stimulus axis (Paradiso, 1988). Important extensions to the general neural coding approach include the development of Bayesian estimators, joint coding of stimulus variables, and treatments of the issue of dependent noise correlations (Pouget et al., 2003; Averbeck et al., 2006).

Chapter 5 employs the canonical neural coding approach to 3D motion direction in MT neurons, extending the neural coding model by explicitly modeling the environment-to-retina transformation. This relatively simple approach continues to be powerful, making striking predictions for human perceptual behavior in

3D motion direction estimation tasks and predicting perceptual phenomena such as the Pulfrich effect (see Chapter 6.

1.6 3D motion

Three-dimensional (3D) motion serves as a unifying theme of this dissertation; I examine both the perception of 3D motion (see Chapters 3 and 6) and the neural mechanisms underlying 3D motion (see Chapter 5). Cormack et al. (2017) provides an extensive review of the subject prior to the contributions of this dissertation. Here, I provide a brief introduction to the subject from the perspectives of perception and neurophysiology.

Though 3D motion processing is critical to animal behavior, studies of the visual processing of motion and the visual processing of depth have largely remained separate. This separation was reinforced by early perceptual discoveries like Julesz's random dot stereograms, stimuli that established that depth could be perceived with binocular disparity signals alone (i.e. in the absence of monocular signals; Julesz 1971). In these stimuli, the images presented to the left and right eye are composed of white noise and thus have no perceptible structure when viewed monocularly. Horizontal offsets between the left and eye images are introduced for some parts of the left and right eye image. When viewed binocularly these offsets result in identifiable depth structure in the percept. Subsequent work introduced the dynamic random element stereogram, which had disparities that changed over time and drew a new set of random elements on each frame. The resulting stimuli were entirely disparity-based, but observers could perceive 3D motion. Taken together this work led to the prevailing attitude that motion and depth were separate *modules* in visual processing (Cumming & Parker, 1994).

The issue of neuronal selectivity for motion-through-depth in the visual cortex was considered in the early studies of neuronal selectivity in the middle temporal area (MT; Zeki 1974; Maunsell & Van Essen 1983a; Albright et al. 1984). These studies showed that cortical neurons contained populations with tuning for static disparities, and there appeared to be a handful of 'opposed-movement' cells (i.e. cells that were tuned to motion in the opposite direction in the two eyes). However, no study could distinguish a population of cells that exhibited a clear tuning to motion-through-depth. Researchers concluded that the responses to motion-through-depth

found in MT could be explained as a combination of tuning for frontoparallel motion and static disparities (Maunsell & Van Essen, 1983a).

However, recent studies have revisited the issue of 3D motion tuning in MT. Rokers et al. (2009) established that human MT/MST complex exhibits a strong selective adaptation to 3D motion direction. Subsequent electrophysiological work determined that neurons in MT do have selectivity for 3D motion direction (Sanada & DeAngelis, 2014; Czuba et al., 2014). Chapter 5 explains the origin of this selectivity at the level of individual neurons by describing an encoding model for 3D motion direction.

1.7 Summary

This thesis furthers the understanding of the visual perception of the dynamic three-dimensional environment and the neural mechanisms underlying that perception. It focuses primarily on 3D motion, developing simple models of sensorimotor control and neural coding. These contributions reveal principles of motion perception critical to the understanding of human visual perception in the natural world.

Chapter 2

Continuous psychophysics: Target-tracking to measure visual sensitivity

This work was published in the Journal of Vision. Bonnen, K., Burge, J., Yates, J., Pillow, J., & Cormack, L. K. (2015). Continuous psychophysics: Target-tracking to measure visual sensitivity. Journal of Vision. 15(3):14.

Author contributions: K.B., J.Y., J.B., and L.C. conceived and designed research; K.B., J.B. and L.C. performed experiments; K.B., J.B., J.P. and L.C. analyzed data; K.B., J.B., J.P. and L.C. interpreted results of experiments; K.B., J.B. and L.C. prepared figures; K.B., J.B., J.P. and L.C. drafted manuscript; K.B., J.B., J.Y., J.P. and L.C. edited and revised manuscript; all authors approved final version of manuscript.

We introduce a novel framework for estimating visual sensitivity using a continuous target-tracking task in concert with a dynamic internal model of human visual performance. Observers used a mouse cursor to track the center of a 2D Gaussian luminance blob as it moved in a random walk in a field of dynamic additive Gaussian luminance noise. To estimate visual sensitivity, we fit a Kalman filter model to the human tracking data under the assumption that humans behave as Bayesian ideal observers. Such observers optimally combine prior information with noisy observations to produce an estimate of target position at each time step. We found that estimates of human sensory noise obtained from the Kalman filter fit were highly correlated with traditional psychophysical measures of human sensitivity ($R^2 > 97\%$). Because each frame of the tracking task is effectively a "mini-trial", this technique reduces the amount of time required to assess sensitivity compared with traditional psychophysics. Furthermore, because the task is fast, easy, and fun, it could be used to assess children, certain clinical patients, and other populations that may get impatient with traditional psychophysics. Importantly,the modeling framework provides estimates of decision variable variance that are directly comparable with those obtained from traditional psychophysics. Further, we show that easily-computed summary statistics of the tracking data can also accurately predict relative sensitivity (i.e. traditional sensitivity to within a scale factor).

KEYWORDS: psychophysics, vision, Kalman Filter, manual tracking

2.1 Introduction

If a stimulus is visible, observers can answer questions such as "Can you see it?" or "Is it to the right or left of center?" This fact is the basis of psychophysics. Since Elemente der Psychophysik was published in 1860 (Fechner, 1860), an enormous amount has been learned about perceptual systems using psychophysics. Much of this knowledge relies on the rich mathematical framework developed to connect stimuli with the type of simple decisions just described (e.g. Green & Swets, 1966). Unfortunately, data collection in psychophysics can be tedious. Forced-choice paradigms are aggravating for novices, and few but authors and paid volunteers are willing to spend hours in the dark answering a single, basic question over and over again. Also, the roughly one bit per second rate of data collection is rather slow compared with other techniques used by those interested in perception and decision

making (e.g. EEG).

The research described here is based on a simple intuition: if a subject can accurately answer psychophysical questions about the position of a stimulus, he/she should also be able to accurately point to its position. Pointing at a moving target – manual tracking – should be more accurate for clearly visible targets than for targets that are difficult to see. We show that this intuition holds, and that sensitivity measures obtained from a tracking task are directly relatable to those obtained from traditional psychophysics. Moreover, tracking a moving target is easy and fun, requiring only very simple instructions to the subject. Tracking produces a large amount of data in a short amount of time, because each video frame during the experiment is effectively a "mini-trial".

In principle, data from tracking experiments could stand on their own merit. For example, if a subject is able to track a 3 c/d Gabor patch with a lower latency and less positional error than a 20 c/d Gabor patch of the same contrast, then functionally, the former is seen more clearly than the latter. It would be nice, however, to take things a step further. It would be useful to establish a relationship between changes in tracking performance and changes in psychophysical performance. That is, it would be useful to directly relate the tracking task to traditional psychophysics. The primary goal of this paper is to begin to establish this relationship.

We designed complimentary tracking and forced choice experiments such that both experiments: i) used the same targets, ii) contained external noise that served as the performance-limiting noise. We used stimuli that were Gaussian luminance blobs targets corrupted with external pixel noise (Figure 2.1; see Methods for details).

The main challenge was to extract a parameter estimate from the tracking task that was analogous to a parameter traditionally used to quantify performance in a psychophysical task. In a traditional 2AFC psychophysical experiment for assessing position discrimination, the tools of signal detection theory are used to obtain an estimate of the signal-to-noise ratio along a hypothetical decision axis. With reasonable assumptions, the observation noise associated with position estimates can be determined.

For a tracking experiment, recovering observation noise requires a model of tracking performance that incorporates an estimate of the precision with which a target can be localized. General tracking problems are ubiquitous in engineering and

the optimal control theory of simple tracking tasks is well established. For cases like our tracking task, the Bayesian optimal tracker is the Kalman filter (Kalman, 1960). The Kalman filter explicitly incorporates an estimate of the performance-limiting observation noise as a key component. The next few paragraphs provide a brief discussion of the logic behind a Kalman filter. The purpose of the discussion is to make clear how observation noise affects a Kalman filter's tracking performance

In order to track a target, the Kalman filter uses the current observation of a target's position, information about target dynamics, and the previous estimate of target position to obtain an optimal (i.e. minimum mean squared error) estimate of true target position on each time step. Importantly, the previous estimate has a (weighted) history across previous time steps built-in. How these values (the noisy observation, target dynamics, and the previous estimate) are combined is dependent on the relative size of the two sources of variance present in the Kalman filter: 1) the observation noise variance (i.e. the variance associated with the current sensory observation) and 2) the target displacement variance (i.e. the variance driving the target position from time step to time step).

When the observation noise variance is low relative to the target displacement variance (i.e. target visibility is high), the difference between the previous position estimate and the current noisy observation is likely to be due to changes in the position of the target. That is, the observation is likely to provide reliable information about the target position. As a result, the previous estimate will be given little weight compared to the current observation. Tracking performance will be fast and have a short lag.

On the other hand, if observation noise variance is high relative to target displacement variance (i.e. target visibility is low), then the difference between the previous position estimate and the current noisy observation is likely driven by observation noise. In this scenario, little weight will be given to the current observation while greater weight will be placed on the previous estimate. Tracking performance will be slow and have a long lag. Thus, the Kalman filter qualitatively predicts the data patterns observed in this set of experiments, under the assumption that increasing blob width reduces target visibility, thereby increasing observation noise.

In our analysis, we fit human tracking data with a Kalman filter, but this approach is different from traditional Kalman filter applications. Typically the

Kalman filter is used to estimate unknown target positions given a set of noisy observations, but here we flip the Kalman filter and learn noise parameters given known target positions. We allowed the model's observation noise parameter, R, to vary as a free parameter. The parameter value (observation noise variance) that maximizes the likelihood of the fit under the model is our estimate of the target position uncertainty that limits the tracking performance of the observer.

In the results that follow, we show that using a Kalman filter to model the human tracking data yields essentially the same estimates of position uncertainty as do standard methods in traditional psychophysics. The correlations between the results of the two paradigms are extremely high, with over 97% of the variance accounted for. We also show that more easily computed statistical summaries of tracking data (e.g. the width of the peak of the cross-correlation between stimulus and response) correlate almost as highly with traditional psychophysical results. To summarize, an appropriately constructed tracking task is a fun, natural way to collect large, rich datasets, and yield essentially the same results as traditional psychophysics in a fraction of the time.

2.2 General Methods

Observers

Three of the authors served as observers. All had normal or corrected-to-normal vision. Two of the three had extensive prior experience in psychophysical experiments. All the observers participated with informed consent and were treated according the principles set forth in the Declaration of Helsinki of the World Medical Association

Stimuli

The target was a luminance increment (or "blob") defined by a 2-dimensional Gaussian function embedded in dynamic Gaussian pixel noise. We manipulated the spatial uncertainty of the target by varying the space constant (standard deviation, hereafter referred to as 'blob width') of the Gaussian keeping the luminous flux (volume under the Gaussian) constant. Examples of these are shown in Figure 2.1. The space constants were 11, 13, 17, 21, 25, and 29 arcmin; the intensity of the pixel noise was clipped at 3 standard deviations, and set such that the maximum

value of the 11 arcmin Gaussian plus three noise standard deviations corresponded to the maximum output of the monitor. We used this blob target (e.g. as opposed to a Gabor patch) because, for the tracking experiment, we wanted a target with an unambiguous bright center at which to point.

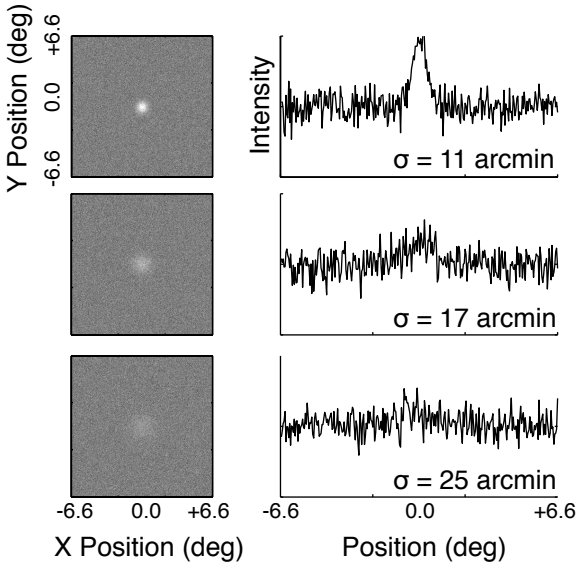

Figure 2.1: Examples of the stimuli are shown in the left column, and cross-sections (normalized luminance vs. horizontal position) are shown on the right.

In the tracking experiment, the target moved according to a random (Brownian) walk for 20 s (positions updated at 60 Hz) around a square field of noise about 6.5 deg (300 pixels) on a side. To specify the walk, we generated two sequences of Gaussian white noise velocities (v_x, v_y) with a one pixel per frame standard deviation. These were summed cumulatively to yield a sequence of x,y pixel positions. Also visible was a 2x2 pixel (2.6 arcmin) square red cursor that the observer controlled with the mouse.

Apparatus

The stimuli were displayed on a Sony OLED flat monitor running at 60 Hz. The monitor was gamma-corrected to yield a linear relationship between luminance and pixel value. The maximum, minimum, and mean luminances were 134.1, 1.7, and 67.4 cd/m2 respectively.

All experiments were run using custom code written in MATLAB and used the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). A standard USB mouse was used to record the tracking data, and a standard USB keyboard was used to collect the psychophysical response data.

Experiments 1 and 2 (the tracking and main psychophysics experiments) were run using a viewing distance of 50 cm giving 45.5 pixel per degree of visual angle. Experiment 3, a supplementary psychophysical experiment on the effect of viewing duration, was run using a viewing distance of 65.3 cm giving 60 pixels per degree. In both cases, the observer viewed the stimuli binocularly using a chin cup and forehead rest to maintain head position.

2.3 Experiment I – Tracking

In the tracking experiment, observers tracked a randomly moving Gaussian blob with a small red cursor using a computer mouse. The data were fit with a Kalman filter model of tracking performance. The fitted values of the model parameters provide estimates of the human uncertainty about target position (i.e. observation noise).

Methods

Each tracking trial was initiated by a mouse click. Subjects tried to keep the cursor centered on the target for 20 s while the target moved according to the random walk. The first five seconds of one such trial are shown in Movie 1.

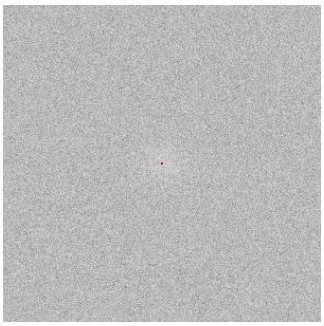

Figure 2.2: **Movie 1**. A 5 sec example of an experimental trial (actual trials were 20 sec long). The luminance blob performed a two-dimensional random walk. Each position was the former position plus normally distributed random offsets (s.d. = 1 pixel) in each dimension (x,y). The subject was attempting to keep the red cursor centered on the blob.

A block consisted of ten such trials at a fixed blob width. Subjects ran one such block at each of the six blob widths in a single session. Each subject ran two sessions and within a session, block order (i.e. blob width) was randomized. Thus each subject completed 20 tracking trials at each blob width, for a total of 24,000 samples (400 seconds at 60 Hz) of tracking data per blob width. As we later show, this is more data than required to produce reliable results (see Appendix 2.7 for an analysis of the precision of tracking estimates vs. sample size). However, we wanted large sample sizes so that we could compare the data with traditional psychophysics with high confidence.

Results

The tracking task yields time series data: the two-dimensional spatial position of a target (left panel of Figure 2.3; black curve) and the position of the tracking cursor (red curve). The remaining panels in Figure 2.3 show the horizontal and vertical components of the time series data in the left panel as a function of time. Subjects were able to track the target. The differences between the two time series (true and tracked target position), and how these differences changed with target visibility (blob width), constitute the dependent variable in the tracking experiment.

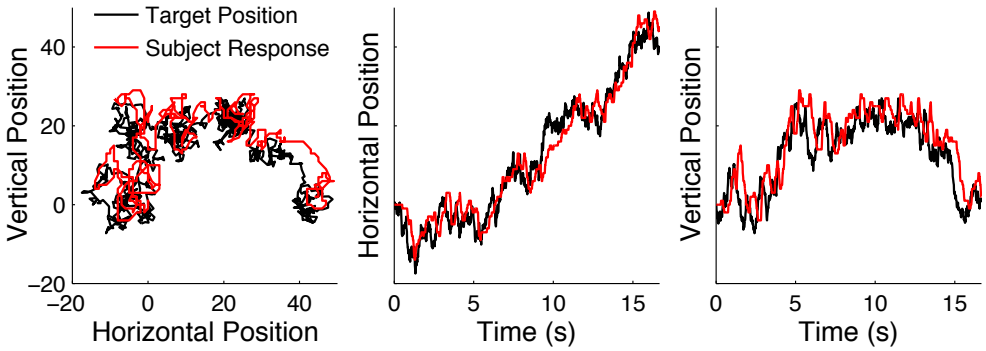

Figure 2.3: Target position and subject response for a single tracking trial. (Left plot) 2D tracking segment with subject response in red and the target position in black. The subject tracked a target starting at (0,0). (Middle and right plot) Horizontal and vertical components over time of the trace in the left plot.

A useful tool for quantifying the relationship between target and response time series is the cross correlogram (CCG; see e.g. Mulligan et al. 2013). A CCG is a plot of the correlation between two vectors of time series data as a function of the lag between them. Figure 2.4 shows the cross-correlation as a function of lag for each individual tracking trial sorted by blob width (i.e. target visibility). Each panel shows CCGs per trial in the form of a heat map (low to high correlation mapped from red to yellow) sorted on the y axis by blob width during the trial. Each row of panels is an individual subject. Because our tracking task has two spatial dimensions, each trial yields a time series for both the horizontal and vertical directions. The first and second columns in the figure show the horizontal and vertical CCGs, respectively, and the black line traces the maximum value of the CCGs across trials. As blob width increases (i.e. lower peak signal-to-noise), the response lag increases, the peak correlation decreases, and the location of the peak correlation becomes more variable. As there were no significant differences between horizontal and vertical tracking in this experiment,the rightmost column of Figure 2.4 shows the average of the horizontal and vertical responses. Clearly, the tracking gets slower and less precise as the blob width increases (i.e. target visibility decreases).

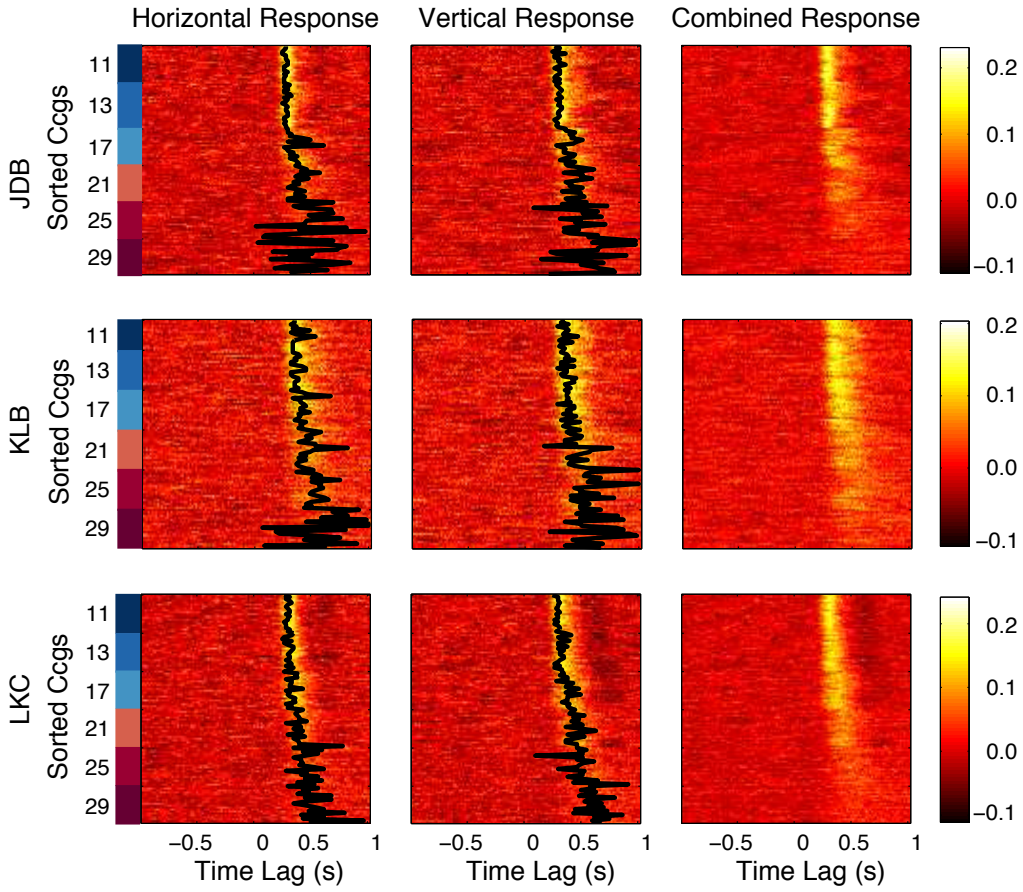

Figure 2.4: Heatmaps of the cross-correlations between the stimulus and response velocities. Each row of a sub-panel represents an individual tracking trial, and the trials have been sorted by target blob width (measured in arcmin and labeled by color blocks that correspond with the curve colors in Figure 2.5) with the easiest to see at the tops of each sub-panel. The black lines trace the peaks of the cross-correlagrams. The right column shows the average of horizontal and vertical response correlations within a trial.

Figure 2.5 shows a plot of the average CCG across trials for each blob width for each of the three subjects (a re-plot of the data from Figure 2.4, collapsing across trial within each blob width). The CCGs sort by blob width: as blob width increases, the height of the CCG peak decreases, the lag of the CCG peak increases,

and the width of the CCG increases. These results show that tracking performance decreases monotonically with the signal-to-noise ratio. This result is consistent with the expected result in a traditional psychophysical experiment. That is, as target visibility decreases, the observer's ability to localize a target should also decrease.

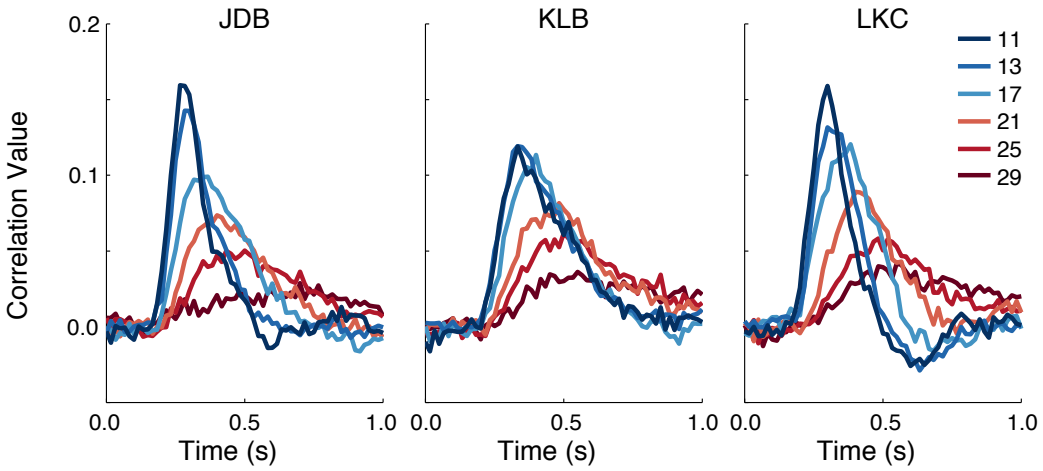

Figure 2.5: Average CCGs for blob width (curve color, identified in the legend by their σ in arcminutes) for each of the three observers (panel). The peak height, location of peak, and width of curve (however measured) all sort neatly by blob width, with the more visible targets yielding higher, prompter, and sharper curves. This shows that there is at least a qualitative agreement between measures of tracking performance and what would be expected from a traditional psychophysical experiment.

In order to quantify tracking performance in a way that can be directly related to traditional psychophysics, we fit a Kalman filter model to the data and extracted the observation noise variance (filter parameter R) as a measure of performance. Figure 2.6 illustrates the details of the Kalman filter in the context of the tracking task. Our experiment generated two position values at each time step in a trial: 1) the true target position (x_t) on the screen and 2) the position of the observer's cursor ($\hat{x}_t$), which was his or her estimate of the target position (plus dynamics due to arm kinematics, motor noise and noise introduced by spatiotemporal response properties of the input device). The remaining unknowns in the model are the noisy sensory observations, which are internal to the observer and

cannot be measured directly. These noisy sensory observations are modulated by a single parameter; the observation noise variance (R). We fit the observation noise variance (R) of a Kalman filter model (per subject) by maximizing the likelihood of the human data under the model given the true target positions (see Appendix 2.8 for details). Note that we have assumed for the purpose of this analysis that the aforementioned contributions of arm kinematics, motor noise, and input device can be described by a temporal filter with fixed properties.

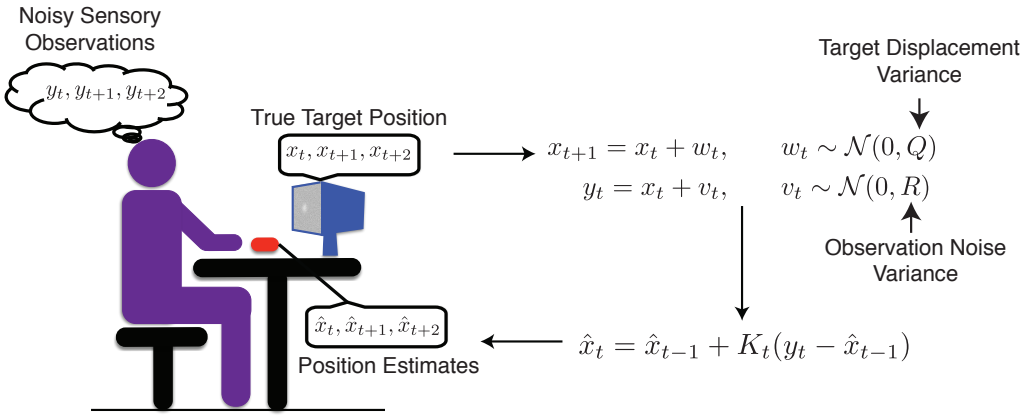

Figure 2.6: Illustration of the Kalman filter and our experiment. The true target positions and the estimates (cursor positions) are known, while the sensory observations, internal to the observer, are unknown. We estimated the variance associated with the latter, denoted by R, by maximizing the likelihood of the position estimates given the true target positions by adjusting R as a free parameter.

For a given observer, this maximum-likelihood fitting procedure was done simultaneously across all the runs for a given blob width throwing out the first second of each run. This yielded one estimate of R for each combination of observer and blob width. Error distributions on R were computed via bootstrapping (i.e. resampling was performed on observers' data by resampling whole trials).

This approach is different from traditional Kalman filter applications. Typically, the Kalman filter is used in situations when the noisy observations are known. The filter parameters (R) are estimated and then the filter can be used to generate estimates ($\hat{x}_t$) of the true target positions (x_t). In our case, the noisy observations

cannot be observed and we estimate the observation noise variance (filter parameter R) given the true target positions (x_t), the target position estimates ($\hat{x}_t$), and the target displacement variance (filter parameter Q). Thus, we essentially use the Kalman filter model in reverse, treating x_t and $\hat{x}_t$ as known instead of y_t, in order to accomplish the goal of estimating R.

We attempt to convey an intuition about what the fitting accomplishes in Figure 2.7. The top-left panel shows an example trace of subject position error (i.e. subject response minus target location). This position error reflects observation noise (and presumably some motor noise and apparatus noise). The bottom-left panel shows three possible traces of position error generated by simulating from the model – the black trace using an approximately correct value of R (such as that on which our analysis converges), and two others (offset vertically for clarity) using incorrect values. Note that, visually, the standard deviations of the red and green traces are too large and too small, respectively. However, the standard deviation of the black curve is approximately equal to the standard deviation of the blue curve (the human error trace). This point is made more clear by examining the distributions of these residual position values collapsed across time (right column). Note that the black distribution has roughly the same width as the blue distribution, while the others are too big or too small. This is essentially what our fitting accomplishes: finding the Kalman filter parameter, R, that results in a distribution of errors with a standard deviation that is "just right". (Brett, 1987).

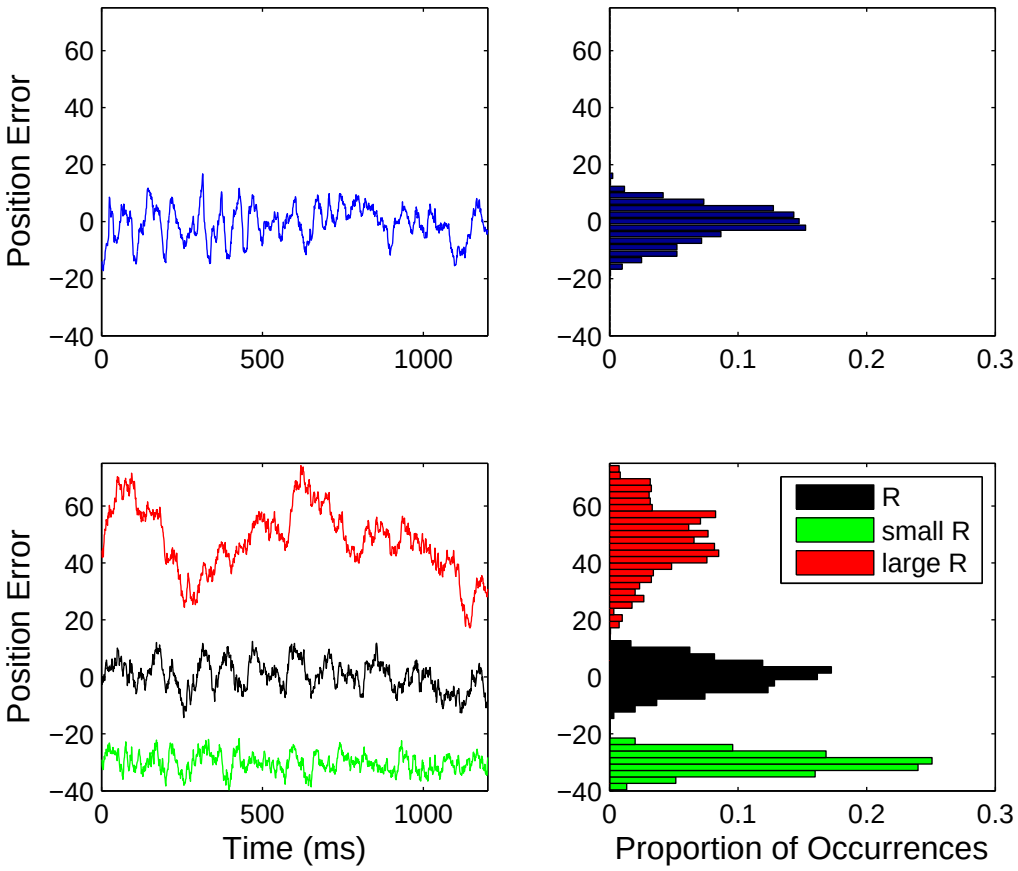

Figure 2.7: The left column shows the positional errors (response position - target position) over time of a subject's response (top) and 3 model responses (bottom, offset vertically for clarity); the black position error trace results from a roughly correct estimate of R. The right column shows the histograms of the positions from the first column. The distribution from the model output with the correct noise estimate (black), has roughly the same width as that from the human response (blue, top).

The results of this analysis are shown in Figure 3.2, which plots the square root of the estimated observation noise variance, $\sqrt{R}$, as a function of blob width for each of the three observers. The estimate of $\sqrt{R}$ represents an observer's un-

certainty about the target location. For the remainder of the paper we refer to $\sqrt{R}$ as the positional uncertainty estimate. The results are systematic, with the tracking noise estimate increasing as a function of blob width in the same way for all three observers. The results are intuitive, in that, as the width of the Gaussian blob increases, the precision with which an observer can estimate the target position decreases, yielding greater error in pointing to the target with a mouse. Qualitatively, they are similar to what we would expect to see in a plot of threshold vs. signal-to-noise ratio derived from traditional psychophysical methods.

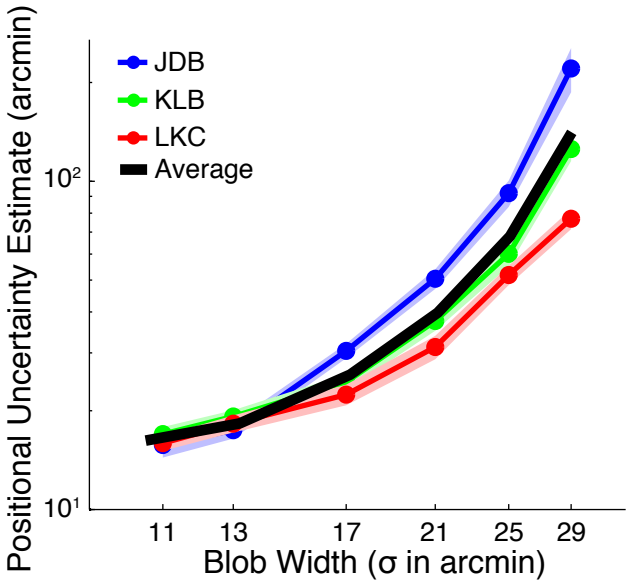

Figure 2.8: Positional uncertainty estimate from the Kalman filter analysis plotted as a function of the Gaussian blob width for three observers. Both axes are logarithmic. The pale colored regions indicate +/- s.e.m. computed by bootstrapping. The black line is the mean across the observers.

Discussion.

We used a Kalman filter to model performance in a continuous tracking task. The values of the best fitting model parameters provide estimates of the uncertainty with which observers localize the target. The results were systematic and agree qualitatively with the cross-correlation analysis, which is a more conventional way to analyze time-series data. Next, we determine the quantitative relationship between

estimates of positional uncertainty obtained from tracking and from a traditional psychophysical experiment.

2.4 Experiment II – Forced Choice Position Discrimination

In this experiment, observers attempted to judge the direction of offset of the same luminance targets used in the previous experiment. The results were analyzed using standard methods to estimate the (horizontal) positional uncertainty that observers had about target position.

Forced Choice Methods.

The apparatus was as described in General Methods. An individual trial is depicted in Figure 2.9. On each video frame throughout a trial, a new sample of Gaussian distributed noise, independent in space and time (e.g. white), was added to the target. The noise parameters were identical to those used in the tracking experiment. On each trial, the observer saw two 250 ms target presentations, separated by a 100 ms inter-stimulus interval. In one interval, the target always appeared in the center of the viewing area. In the other interval, the target appeared at one of nine possible stimulus locations (four to the left, four to the right and zero offset). The observer's task was to indicate whether the second interval target was presented to the left or right of the first interval target. Data were collected in blocks of 270 trials. Blob width was fixed within a block. Targets were presented 30 times at each of the nine comparison locations in a pseudo-random order. Each observer completed three blocks for each of the six target blobs, for a total of 4860 trials per observer (270 trials/block x 3 blocks/target x 6 targets/observer).

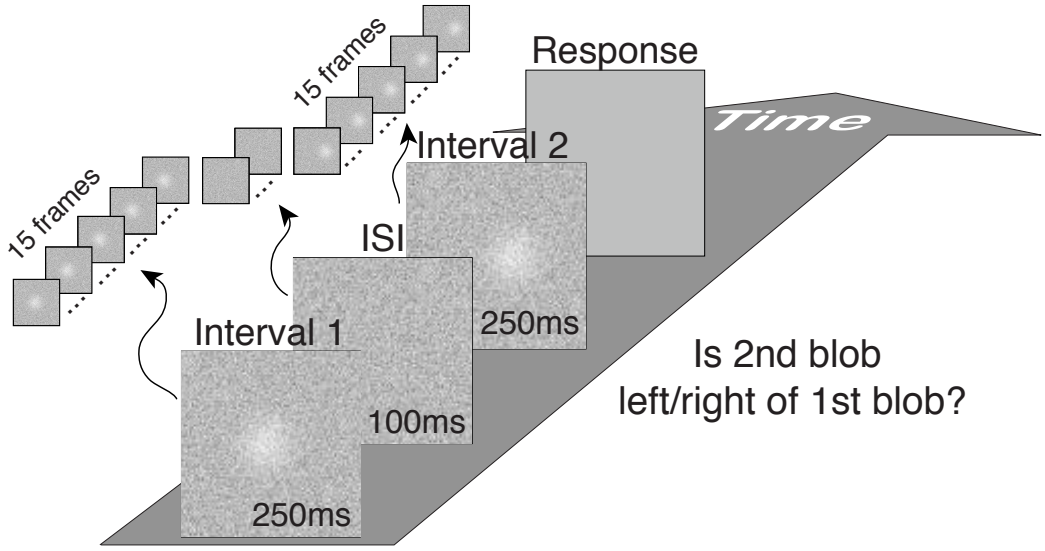

Figure 2.9: Timeline of a single trial. The task is a two interval forced choice task. The stimuli were Gaussian blobs in a field of white Gaussian noise. Subjects were asked to indicate whether the second blob was presented to the left/right of the first blob.

The data for each run were fit with a cumulative normal psychometric function (ϕ), and the spatial offset of the blob corresponding to d' = 1.0 point (single interval) was interpolated from the fit. The d' for single interval was used because it corresponds directly to the width of the signal+noise (or noise alone) distribution. Because $P_R = \phi\left(\frac{d'_{2I}}{2}\right) = \phi\left(\frac{d'}{\sqrt{2}}\right)$ where P_R is the percent rightward choices and d'_{2I} is the 2-interval d', threshold was defined as the change in position necessary to travel from the 50% to the 76% rightward point on the psychometric function.

Results

Thresholds as a function of blob width are shown in Figure 2.10. The solid data points are the threshold estimates from fitting all of an observer's data at a given blob width, and the error bands are +/- one standard error obtained by bootstrapping from the raw response data. The heavy black line shows the (arithmetic) mean for the three observers. The thresholds for all observers increase with increasing blob width, with a hint of a lower asymptote for the smallest targets. This is the

same basic pattern of data we would expect using an equivalent noise paradigm in a detection (e.g. Pelli 1990) or localization task, as the amount of effective external noise increases with increasing blob width.

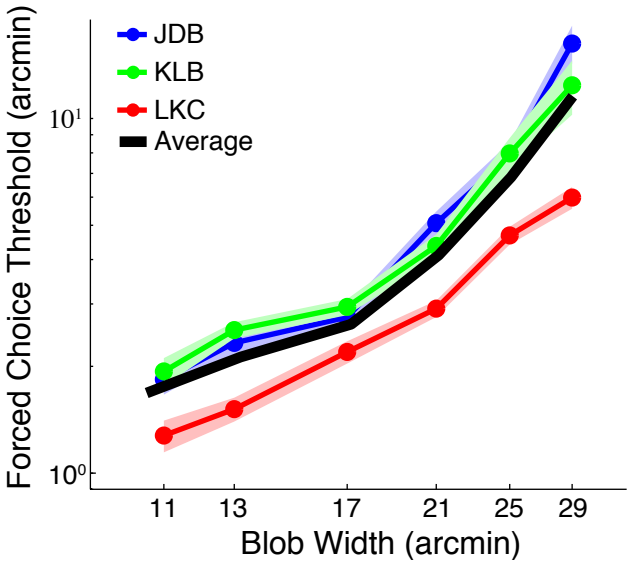

Figure 2.10: Forced choice threshold as a function of blob width. Each subjects average data are shown by the solid points, and the bands indicate bootstrapped s.e.m. Both axes are logarithmic. The solid black line shows the average across subjects.

Discussion.

The thresholds presented in Figure 2.10 correspond to a d' of 1.0, and thus represent the situation in which the relevant distributions along some decision axis were separated by their common standard deviation. Assuming that the position of the target distribution on the decision axis is roughly a linear transformation of the target's position in space, then this also corresponds to the point at which the targets were separated by roughly one standard deviation of the observer's uncertainty about their position. Thus, the offset thresholds serve as an estimate of the width of the distribution that describes the observer's uncertainty about the target's position. This is exactly what the positional uncertainty estimates represented in the tracking experiment. In fact, it would be reasonable to call the forced choice thresholds

"positional uncertainty estimates" instead. The use of the word threshold is simply a matter of convention in traditional psychophysics.

Figure 2.11 shows a scatterplot of the results from the tracking experiment (y coordinates) vs. those from the traditional psychophysics (x coordinates). The log-log slopes are 0.98 (LKC), 1.12 (JDB), and 1.02 (KLB). The corresponding correlations are 0.985, 0.996, and 0.980, respectively. Obviously, the results are in good agreement; the change in psychophysical thresholds with blob width is accounting for over 96% of the variance in the estimates obtained from the tracking paradigm, the high correlation indicates that the two variables are related by an affine transformation. In our case (see Figure 2.11), the variables are related by a single scalar multiplier. This suggests to us that the same basic quantity is being measured in both experiments.

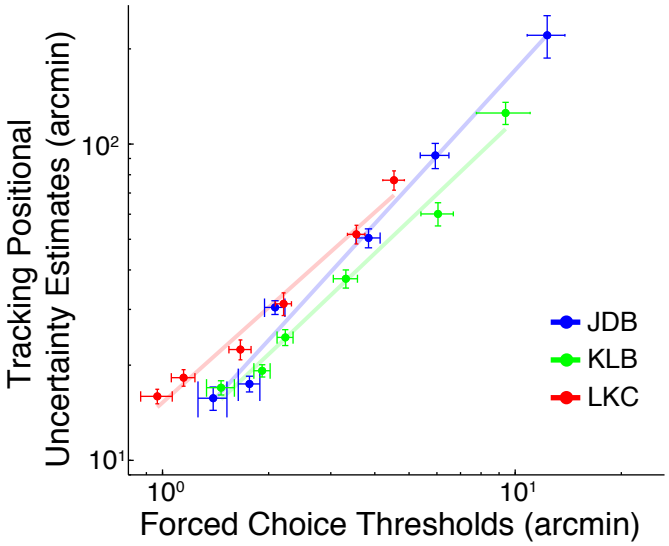

Figure 2.11: Scatter plot of the position uncertainty estimated from the tracking experiment (y axis) as a function of the thresholds from traditional psychophysics (x axis) for our 3 observers. The log-log slope is very close to 1 and the percentage of variance accounted for is over 96% for each observer.

There is, however, an offset of about one log unit between the estimates generated by the two experiments. For example if, for a given blob width, the 2AFC task yields an estimate of 1 arcmin. of positional uncertainty, the tracking

task would yield a corresponding estimate of 10 arcmin. The relative estimates are tightly coupled, but we would like to understand the reasons for the discrepancy in the absolute values. One obvious candidate is temporal integration, which would almost certainly improve performance in the psychophysical task relative to the tracking task.

2.5 Experiment III – Temporal Integration

One possible reason for the fixed discrepancy between the positional uncertainty estimates in the tracking task and the thresholds in the traditional psychophysical task is temporal integration. In the traditional task, the observers could benefit by integrating information across multiple stimulus frames (up to 15 per interval) in order to do the task. If subjects integrated perfectly over all 15 frames, threshold would be $\sqrt{15}$ times lower than the thresholds that would be estimated from 1 frame. The positional uncertainty estimated in the tracking task is the positional uncertainty associated with a single frame. Thus, it is possible that approximately half of the discrepancy between the forced choice and tracking estimates of positional uncertainty is due to temporal integration in the forced choice experiment.

It's also important to consider how the tracking task may be affected by temporal integration. In practice, if an observer's sensory-perceptual system is performing temporal integration then they are responding to a spatially smeared representation of the moving stimulus – a motion streak – instead of the instantaneous stimulus. Temporal integration per se is not modeled in our implementation of the Kalman filter, but its presence in the data would result in an overestimate of observation noise. This effect of temporal integration might further add to the discrepancy between the measurements of positional uncertainty.

In this experiment, we sought to measure our observers' effective integration time and the degree to which this affected the psychophysical estimates of spatial uncertainty.

Methods.

The methods for this experiment were the same as for Experiment II (above), except that the duration of the stimulus intervals was varied between 16.7 ms (one frame) and 250 ms (15 frames) while blob width was fixed. The inter-stimulus interval

remained at 100 ms. Observers KLB and JDB ran at a 17 arcmin blob width, and LKC ran at 21 arcmin (values that yielded nearly identical thresholds for the three observers in experiment 2). These were run using the same Sony OLED monitor, but driven with a Mac Pro at a slightly different viewing distance (see General Methods).

Results.

Figure 2.12 shows the offset thresholds as a function of stimulus duration. As in Figure 2.10, the data points are the interpolated thresholds (d'=1) from the cumulative normal fits, the error bands show +/- 1 standard error estimated by bootstrapping, and the solid black line show the mean thresholds across subject. Thresholds for all observers decreased with increased stimulus duration at the expected slope of $1/\sqrt{(n)}$ (dashed line for reference) until flattening out at roughly 50 to 100ms or 3 to 5 frames (Watson, 1979; Nachmias, 1981).

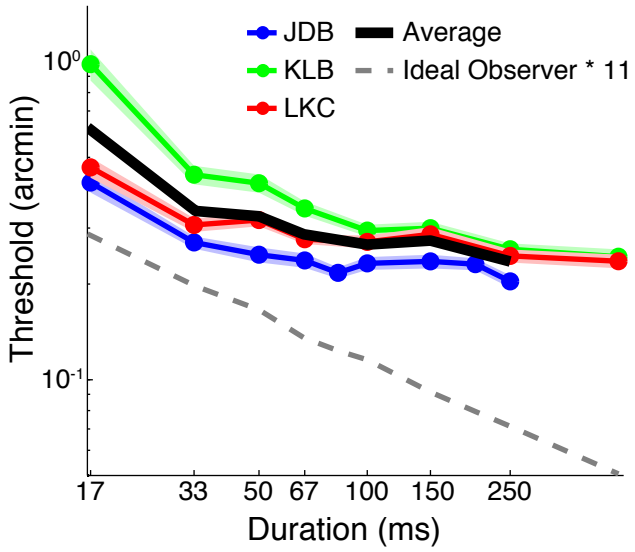

Figure 2.12: Threshold as a function of stimulus duration. Each subjects average data are shown by the solid points. Both axes are logarithmic. Data points and error bands are as in Figure 2.10. The gray line displays the performance of an ideal observer shifted up by a multiplier (11).

Discussion.

The thresholds at single frame durations approximate what thresholds would be if observers could not benefit from temporal integration in the psychophysical task. As we argued earlier, moreover, the tracking task could not have benefited from temporal integration; if anything, using multiple frames would cause the uncertainty estimates from the tracking task to be too high. It would therefore be conservative to correct the psychophysical thresholds from Experiment II upward by a factor corresponding to the ratio between the single frame and 15 frame thresholds from Experiment III. This turns out to be about a factor of 2, and would reduce the absolute difference between the tracking and psychophysical estimates from a factor of 10 to about a factor of 5.

An important next step in understanding temporal integration is to perform a comparable experiment in the tracking task (i.e. manipulating the rate at which the stimulus moves). Such a follow-up study would will further clarify the relationship between the forced choice task and the tracking task, as well as solidify the appropriate stimulus for a psychophysics tracking task.

2.6 General Discussion

In this paper, we have shown that data from a simple tracking task can be analyzed in a principled way that yields essentially the same answers that result from a traditional psychophysical experiment using comparable stimuli in a fraction of the time. In this analysis, we modeled the human observer as a dynamic system controller – specifically a Kalman filter (known from sensor calibration, e.g.). The Kalman filter is typically used to produce a series of estimated target positions given an estimate of the observation noise. We, in contrast, used the Kalman filter to estimate the observation noise given a series of estimated target positions generated by observer during our experiments.

The conceptualization of a human as an element of a control system in a tracking task is not a novel concept. In fact, this seems to be one of the problems that Kenneth Craik was working on at the time of his death – two of his manuscripts on the topic were published posthumously by the British Journal of Psychology (Craik 1947; Craik 1948). Because circuits or, later, computers, are generally much better

feedback controllers than humans, there has been less interest in the specifications of human-as-controller with a few exceptions: studies of pilot performance in aviation, motor control, and eye movement research (in some ways a sub-branch of motor control, in other ways a sub-branch of vision).

It is clear that the job of a pilot, particularly when flying with instruments, is largely to be a dynamic controller that minimizes the error between an actual state and a goal state. For example, the goal state might be a particular altitude and heading assigned by air traffic control. The corresponding actual state would be the current heading and altitude of the airplane. The error to be minimized is the difference between the current and goal states as represented on the aircraft's instruments. It comes as no surprise, then, that a large literature has emerged in which the pilot is treated as, in Craik's terms, an engineering system that is itself an element within a larger control system. However the pilot's sensory systems are not generally considered a limiting factor; pilot errors are never due to poor acuity (to our knowledge) but rather due to attentional factors related to multitasking or, occasionally, sensory conflict (visual vs. vestibular) resulting in vertigo. As such, while tracking tasks are often studied in the aviation literature, is not done to assess a pilots' sensory (or basic motor) capabilities.

The motor control literature involving tracking tasks can be divided into three main branches: eye movement control (e.g. Mulligan et al., 2013), manual (arm and hand) control (e.g. Berniker & Kording, 2008; Wolpert & Ghahramani, 1995), and, to a lesser extent, investigations of the interaction between the two (e.g. Brueggemann, 2007; Burge et al., 2008b; Burge et al., 2010; van Dam & Ernst, 2013). Within the motor control literature, there are several examples of the use of the Kalman filter to model a subject's tracking performance. Some of these focus almost exclusively on modeling the tracking error as arising from the physics of the arm and sensorimotor integration (Berniker & Kording, 2008; Wolpert & Ghahramani, 1995). Others provide a stronger foundation for our work by demonstrating how changing the visual characteristics of a stimulus affects human performance in a manner that can be reproduced by manipulating parameters of the Kalman filter (Burge et al., 2008b). Taken together this body of literature provides strong support for the idea that the human ability to adapt to and track a moving stimulus is consistent with the performance of a Kalman filter. We extend this literature by using the Kalman filter to explicitly estimate visual sensitivity.

In the results section, we showed a strong empirical relationship between the data from tracking and forced-choice tasks. To further this comparison, it would be useful to know what optimal (ideal observer) performance would be. Obviously, if ideal performance in the two tasks were different, then we wouldn't expect our data from experiments I and II to be identical, even if the experiments were effectively measuring the same thing. In other words, if the two experiments yielded the same efficiencies, then we would know they were measuring exactly the same thing. Of course this unrealizable in practice because the tracking response necessarily comprises motor noise (broadly defined) in addition to sensory noise, whereas the motor noise is absent in forced-choice psychophysics due to the crude binning of the response. What we can realistically expect is to see efficiencies from tracking and forced-choice experiments that are highly correlated but with a fixed absolute offset reflecting (presumably) motor noise and possibly other factors.

The ideal observer for the forced-choice task is based on signal detection theory (e.g. Ackermann & Landy, 2010; Geisler, 1989; Green & Swets, 1966). To approximate the ideal observer in a computationally efficient way, we used a family of templates identical to the target but shifted in spatial location to each of the possible stimulus locations. These were multiplied with the stimulus (after averaging across the 15 frames in each interval). The model observer chose the direction that corresponded to the maximum template response, defined as the product of the stimulus with the template (in the case of the zero offset template, then the model observer guessed with $p(right) = 0.5$). The stimuli and templates were rearranged as vectors so that the entire operation could be done as a single dot product as in Ackermann & Landy (2010). The ideal observer was run in exactly the same experiment as the human observers, except that the offsets were a factor of 10 smaller, which was necessary to generate good psychometric functions because of the model's greater sensitivity.

The left panel of Figure 2.13 shows the ideal observer's threshold as a function of blob width (black line), along with the human observers' data from Figure 3.2. The gray line shows the ideal thresholds shifted upward by a factor of 20. The results are as expected: the humans are overall much less sensitive than ideal, they approach a minimum threshold on the left, increase with roughly the same slope as the ideal in the middle, and then begin (or would begin) to accelerate upward as the target becomes invisible. A maximum efficiency of about 0.25% (a 1:20 ratio of

human to ideal d') is approached at middling blob widths, which is consistent with previous work using grating patches embedded in noise (Simpson et al., 2003).

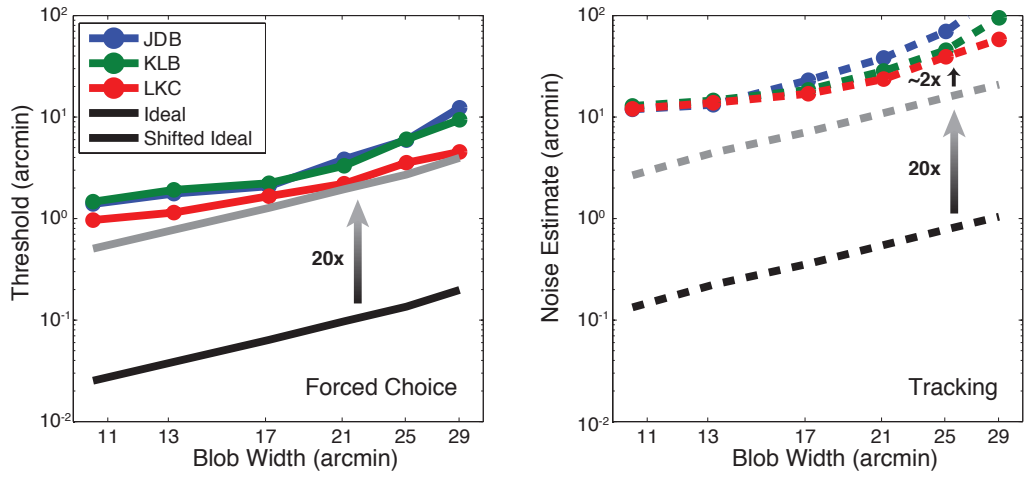

Figure 2.13: Relationship between human observers and an ideal observer. Force Choice Human threshold estimates (left) and tracking noise estimates (right) are replotted (blue, green and red lines). The ideal observers are depicted in black and the shifted ideal in grey. A multiplier of 20 results in the ideal observer approaching the human performance in both experiments.

In the tracking task, the ideal observer's goal was to estimate the location of the stimulus on each stimulus frame. To implement this, a set of templates identical to the stimulus but varying in offset in one dimension around the true stimulus location was multiplied with the stimulus each frame. The position estimate for each frame was then the location of the template producing the maximum response. The precision with which this observer could localize the target was simply the standard deviation of the position estimates relative to the true target location (i.e. the standard deviation of the error). Note that as the ideal observer had no motor system to add noise, this estimate corresponds specifically to the measurement noise in the Kalman filter formulation. It also corresponds to the ideal observer for a single-interval forced-choice task observer given only one stimulus frame per judgment.

The right panel of Figure 2.13 shows the ideal observer's estimated sensory

noise (dashed black line) as a function of blob width, along with the corresponding estimates of spatial uncertainty based on the Kalman filter fit to the human data replotted from Figure 2.13. The slope is the same as for the forced-choice task. The dashed gray line is the ideal threshold line shifted upward by a factor of 20 (the same amount as the shift in the left panel). After a shift reflecting efficiency in the forced choice ask, there is roughly a factor of 2 difference remaining. As previously mentioned, this is not surprising because the observer's motor system must contribute noise to the tracking task but not in the forced choice task.

We have constructed a principled observer model for the tracking task that yields comparable results to traditional forced-choice psychophysics, establishing the validity of the tracking task for taking psychophysical measurements. Here, we introduce simpler methods of analysis for the tracking task that provide an equivalent measure of performance. We show that the results from an analysis of the CCGs (introduced earlier) are just as systematically related to the forced-choice results as are those from the Kalman filter observer model.

The left panel of Figure 2.14 shows CCGs (data points) for observer LKC (replotted from Figure 2.5, right), along with the best fitting sum-of-Gaussians. Though Gaussians are not theoretically good models for impulse response functions, we used them as an example for their familiarity and simplicity. Based on visual inspection they seem to provide a rather good empirical fit to the data. We used a sum of two Gaussians (the second one lagged and inverted), rather than a single Gaussian, in order to model the negative (transient) overshoot seen in the data from the three smallest blob widths for LKC and the smallest blob width for JDB. For all other cases, the best fit resulted in a zero (or very near zero) amplitude for the second Gaussian.

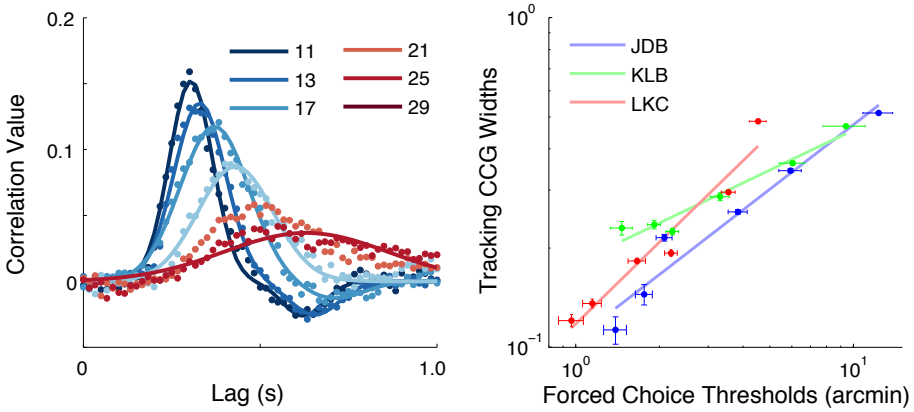

Figure 2.14: The left panel depicts the CCGs for subject LKC sorted by blob width (identified in the legend by their σ in arc minutes). The right panel shows the forced-choice estimates vs. the CCG widths (of the positive-going Gaussians) from the tracking data. Error bars correspond to s.e.m.

The right panel of Figure 2.15 shows the standard deviations of the best fit positive Gaussians from the left panel plotted as a function of the corresponding forced-choice threshold estimates. As with the Kalman filter estimates, the agreement is very good indicating that the tracking data yield basically the same answer as the forced-choice data regardless of analysis.

Two further points can be made about the simple Gaussian fits to the CCGs. First, the best-fit values for the three parameters (amplitude, lag or mean, and standard deviation) are very highly correlated with one another despite being independent in principle. Shown in Figure 2.15 are the best-fit parameter values plotted against one another pairwise. The relationships are plotted (from left to right) for amplitude vs. lag, lag vs. width and width vs. amplitude; the corresponding correlation coefficients are shown as insets. Clearly, it would not matter which parameter was chosen as the index of performance.

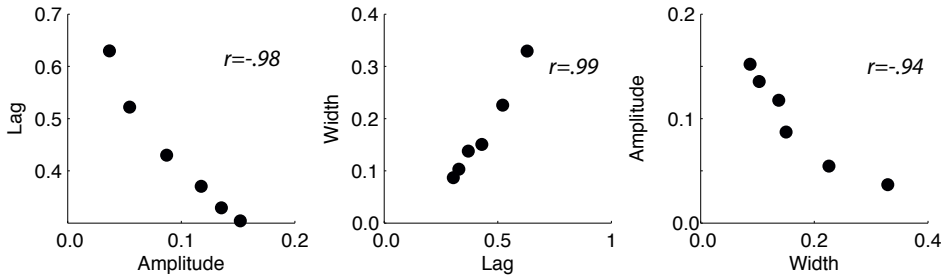

Figure 2.15: Parameters (amplitude, lag, and width) are very highly correlated. From the left to right, the panels represent: amplitude vs. lag, lag vs. width and width vs. amplitude. The correlation coefficients that correspond to each of these relationships are inset in each panel. These parameters are calculated from observer LKC's data.

As an aside, including the second Gaussian (negative) in fitting the CCG is unnecessary. The results are essentially identical when only a single positive Gaussian is used fit to the CCGs.

In conclusion, we have presented a simple dynamic tracking task and a corresponding analysis that produce estimates of observer performance or, more specifically, estimates of the uncertainty limiting observers' performance. These estimates correspond quite closely with the estimates obtained from a traditional forced-choice psychophysical task done using the same targets. Compared with forced-choice stimuli, this task is easy to explain, intuitive to do for naive observers, and fun. Informally, we have run children as young as 5 on a more game-like version of the task, and all were very engaged and requested multiple "turns" at the computer. We find it likely that this would apply more generally, not only to children, but also to many other populations that have trouble producing large amounts of psychophysical data. Finally, the "tracking" need not be purely spatial; one could imagine tasks in which, for example, the contrast of one target was varied in a Gaussian random walk, and the observers' task was to use a mouse or a knob to continuously match the contrast of a second target to it. In conclusion, the basic tracking paradigm presented here produces rich, informative data sets that can be used as fast fun windows onto observers' sensitivity.

2.7 Appendix I: Convergence of Kalman Filter Uncertainty Estimate

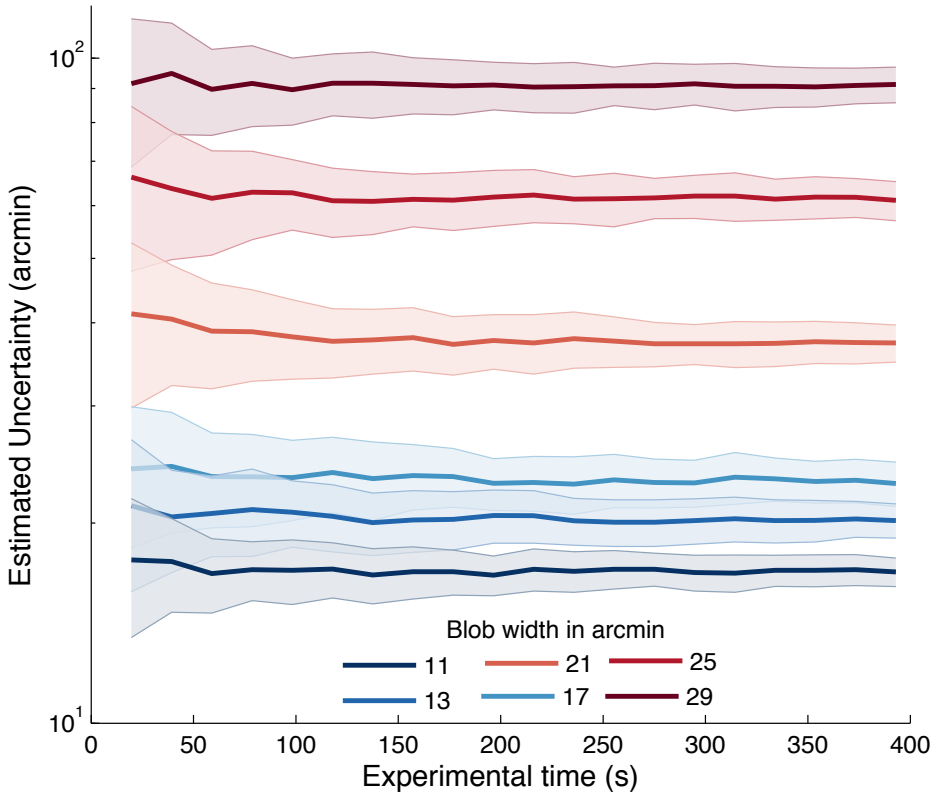

Figure A1. Estimated uncertainty (R) vs. experimental time used to estimate R. Error bounds show ±s.e.m. Blob width is indicated by curve color and identified in the legend by its σ in arc minutes.

Figure A1 demonstrates the time course of the convergence of the Kalman Filter Uncertainty Estimate on one subject's tracking data. Each of the solid lines represents the average estimated uncertainty (R) for a particular stimulus width produced by performing bootstrapping on the fitting procedure as we increase the total experimental time used to estimate R. The clouds around these estimates rep-

resent the standard error. It requires relatively little experimental time to produce reliable estimates of uncertainty using our Kalman filter fitting procedure.

2.8 Appendix II: Kalman Filter for Maximum-Likelihood Fitting Procedure

In this work we use a Kalman filter framework to estimate subjects' observation noise variance (R, see Figure 2.6) and therefore also position uncertainty which is defined as $\sqrt{R}$. The two time series produced by the experimental tracking paradigm – target position (x_t) and subject response ($\hat{x}_t$) – are used in conjunction with the Kalman filter in order to fit observation noise variance by maximizing $p(\hat{\mathbf{x}}|\mathbf{x})$, the probability of the position estimates given the target position under the Kalman filter model.

Consider the tracking paradigm a simple linear dynamical system with no dynamics or measurement matrices:

$$x_{t+1} = x_t + w_t, \qquad w_t \sim \mathcal{N}(0, Q) \tag{2.8.1}$$

$$y_t = x_t + v_t, \qquad v_t \sim \mathcal{N}(0, R) \tag{2.8.2}$$

where the x_t represents the target position, and y_t represents the subjects' noisy sensory observations, which we cannot access directly (see Figure 2.6).

Given a set of observations $y_{1:t}$ and the parameters $\{Q, R\}$, the Kalman filter gives a recursive expression for the mean and variance of $x_t|y_{1:t}$, that is, the posterior over x at time step t given all the observations $y_1, \ldots, y_t$. The posterior is of course Gaussian, described by mean $\hat{x}_t$ and variance P_t. The following set of equations perform the dynamic updates of the Kalman filter and result in target position estimates ($\hat{x}_t$).

$$S_t = P_{t-1} + Q \qquad \text{(prior variance)} \tag{2.8.3}$$

$$K_t = S_t(S_t + R)^{-1} \qquad \text{(Kalman gain)} \tag{2.8.4}$$

$$\hat{x}_t = \hat{x}_{t-1} + K_t(y_t - \hat{x}_{t-1}) \qquad \text{(posterior mean)} \tag{2.8.5}$$

$$P_t = K_t R \qquad \text{(posterior variance)} \tag{2.8.6}$$

We use this definition (eq. B.1-2) and the Kalman filter equations (eq B.3-B.6) to write $p(\hat{\mathbf{x}}|\mathbf{x})$. First, we find the asymptotic value of P_t and then use that to

simplify and rewrite the Kalman filter equations in matrix form.

Since Q and R are not changing over time the asymptotic value of the posterior variance P_t as $t \to \infty$ can be calculated by solving $P = (P+Q)R/(P+Q+R)$ for P, which yields:

$$P_\infty = \frac{-Q + \sqrt{Q^2 + 4QR}}{2} \tag{2.8.7}$$

$$= \frac{Q}{2}(\sqrt{1 + 4RQ^{-1}} - 1). \tag{2.8.8}$$

In order to further simplify, we will assume $P_0 = P_\infty$, i.e., the initial posterior variance will approach some asymptotic posterior variance. A Kalman filter asymptotes in relatively few time steps. In practice, our observers seem to as well, but to be safe we omitted the first second of tracking for each trial to insure that the observers' tracking had reached a steady state. Then the prior variance S, Kalman gain K and posterior variance P are constant. Thus, the dynamics above can be simplified to:

$$\hat{x}_t = (1-K)\hat{x}_{t-1} + Ky_t, \tag{2.8.9}$$

where K is the fixed constant:

$$K = (Q+P)(Q+P+R)^{-1} \tag{2.8.10}$$

This makes $\hat{\mathbf{x}}$ a simple auto-regressively filtered version of $\mathbf{y}$. The dynamics can be expressed in matrix form:

$$D\hat{\mathbf{x}} = K\mathbf{y}, \tag{2.8.11}$$

where D is a bi-diagonal matrix with 1 on the main diagonal and $K-1$ on the below-diagonal:

$$\mathbf{D} = \begin{bmatrix} 1 & & & \\ K-1 & 1 & & \\ & \ddots & \ddots & \\ & & K-1 & 1. \end{bmatrix} \tag{2.8.12}$$

By substituting for $\mathbf{y}$ and multiplying by $\mathbf{D}^{-1}$, this can be rewritten as:

$$\hat{\mathbf{x}} = K\mathbf{D}^{-1}\left(\mathbf{x} + \mathbf{v}\right) \tag{2.8.13}$$

Equation 2.8.13 in conjunction with equation 2.8.10 gives the expression relating the two time series $\hat{\mathbf{x}}$ and $\mathbf{x}$, to the unknown R. We can use this to write $p(\hat{\mathbf{x}}|\mathbf{x})$:

$$p(\hat{\mathbf{x}}|\mathbf{x}) \quad \sim \quad \mathcal{N}(K\mathbf{D}^{-1}\mathbf{x},\ K^2R\mathbf{D}^{-1}\mathbf{D}^{-\top}). \tag{2.8.14}$$

The log likelihood, $log(p(\hat{\mathbf{x}}|\mathbf{x}))$ (below), is used in order to perform the maximum-likelihood estimation of R.

$$\begin{aligned} log\left(p(\hat{\mathbf{x}}|\mathbf{x})\right) &= log(\mathcal{N}(\hat{\mathbf{x}}|K\mathbf{D}^{-1}\mathbf{x}, K^2R\mathbf{D}^{-1}\mathbf{D}^{-\top}) && (2.8.15)\\ &= -\frac{n}{2}log(2\pi) - \frac{1}{2}log\left|K^2R\mathbf{D}^{-1}\mathbf{D}^{-\top}\right| && (2.8.16)\\ &- \frac{1}{2}\left(\hat{\mathbf{x}} - K\mathbf{D}^{-1}\mathbf{x}\right)^{\top}\left(K^2R\mathbf{D}^{-1}\mathbf{D}^{-\top}\right)^{-1}\left(\hat{\mathbf{x}} - K\mathbf{D}^{-1}\mathbf{x}\right) && (2.8.17)\end{aligned}$$

(Note: coefficients $\mathbf{D}$ and K are defined in terms of Q and R). The log likelihood for a particular blob width ($\sigma = s$) for a given subject is evaluated by taking the sum over all trials with $\sigma = s$ of $p(\hat{\mathbf{x}}|\mathbf{x})$. In our analysis, maximum-likelihood estimation of R is performed for each blob width in order to investigate how the observer's positional uncertainty ($\sqrt{R}$) changes with increasing blob width (decreasing target visibility). *MATLAB implementation available from authors upon request.*

Chapter 3

Dynamic mechanisms of visually-guided 3D motion tracking

This work was published in the Journal of Neuroscience. Bonnen, K., Huk, A. C., & Cormack, L. K. (2017). Dynamic mechanisms of visually guided 3D motion tracking. Journal of Neurophysiology, 118(3), 1515-1531.

Author contributions: K.L.B., A.C.H., and L.K.C. conceived and designed research; K.L.B. and L.K.C. performed experiments; K.L.B. and L.K.C. analyzed data; K.L.B., A.C.H., and L.K.C. interpreted results of experiments; K.L.B. and L.K.C. prepared figures; K.L.B. and L.K.C. drafted manuscript; K.L.B., A.C.H., and L.K.C. edited and revised manuscript; K.L.B., A.C.H., and L.K.C. approved final version of manuscript.

Abstract

The continuous perception of motion-through-depth is critical both for navigation and interacting with objects in a dynamic three dimensional (3D) world. Here we used 3D tracking to simultaneously assess the perception of motion in all directions, facilitating comparisons of responses to motion-through-depth to frontoparallel motion. Observers manually tracked a stereoscopic target as it moved in a 3D Brownian random walk. We found that continuous tracking of motion-through-depth was selectively impaired, showing different spatiotemporal properties compared to frontoparallel motion tracking. Two separate factors were found to contribute to this selective impairment. The first is the geometric constraint that motion-through-depth yields much smaller retinal projections than frontoparallel motion, given the same object speed in the 3D environment. The second factor is the sluggish nature of disparity processing, which is present even for frontoparallel motion tracking of a disparity-defined stimulus. Thus, despite the ecological importance of reacting to approaching objects, both the geometry of 3D vision and the nature of disparity processing result in considerable impairments for tracking motion-through-depth using binocular cues.

3.1 Introduction

The perception of motion-through-depth is crucial to human behavior. It provides information necessary for tracking moving objects in the three-dimensional (3D) world so that we can, for example, duck to avoid being hit. However, the perception of motion and depth are typically examined independently. Both have become powerful model systems for investigating how information is processed in the brain (Julesz, 1971; Newsome & Pare, 1988; Shadlen & Newsome, 2001). However, significantly less work has considered the perception of motion and depth as part of one unified perceptual system for processing position and motion information from the 3D world.

In this study subjects continuously followed objects moving in a Brownian random walk through a 3D environment. The target tracking task we employ here provides a rich and efficient paradigm for examining visual perception and visually-guided action in the 3D world (Bonnen et al., 2015). Manual tracking responses can be collected at a much higher temporal resolution compared to the binary decisions in trial-based forced choice psychophysics. Our previous work has demonstrated that target tracking provides measures of visual sensitivity that are comparable to those obtained using traditional psychophysical methods (Bonnen et al., 2015). This prior work relied on the underlying logic that tracking should be more accurate for a clearly visible target than for targets that are difficult to see. Here we extend this logic to investigate 3D motion perception: tracking should be more accurate for clearly visible motion than for motion that is more difficult to see. Tracking also takes advantage of the natural human ability to follow objects in the environment. Forced choice tasks typically require that subjects view a single motion stimulus and make a binary decision about that motion, which they then communicate with a button press or other discrete behavioral response that is often arbitrarily mapped onto the visual perception or decision. While this traditional approach has yielded much information about motion processing in the visual system, tracking allows us to examine motion perception in finer temporal detail in the context of a task that is also more natural for observers.

Our first experiment examined tracking performance for Brownian motion in a 3D space. Subjects were instructed to track the center of a target (by pointing at it with their finger) as it moved in a 3D Brownian random walk. Tracking performance

was impaired for motion-through-depth relative to horizontal and vertical motion. Thus the impaired processing of motion-through-depth observed in discrete, trial-based tasks generalizes to naturalistic, continuous visually-guided behavior (Tyler, 1971; McKee et al., 1990; Nienborg et al., 2005; Brooks & Stone, 2006; Katz et al., 2015; Cooper et al., 2016). Follow-up experiments isolated the sources of the deficits for tracking motion-through-depth. Experiment II and III show that the deficit is partially due to the geometry of motion-through-depth relative to an observer. However, this did not account for the longer latencies for tracking motion-through-depth compared to frontoparallel motion. We hypothesized that this remaining difference was a signature of disparity processing (Wheatstone, 1838); previous work has shown behavioral delays for static disparities and slower temporal dynamics for neural responses (Braddick, 1974; Norcia & Tyler, 1984; Cumming & Parker, 1994; Nienborg et al., 2005). Experiment V examined whether the longer latencies observed in Experiments I and II can be attributed to disparity processing. When disparity processing was imposed on frontoparallel motion tracking using dynamic random element stereograms, we found impaired tracking performance that better matched the temporal characteristics of motion-through-depth tracking.

In summary, we found that the diminished performance in depth motion tracking can be explained by a combination of two factors: a geometric penalty, because 3D spatial signals give rise to 2D retinal signals (projections of the 3D motion), and a disparity processing penalty, because the combination of signals across the two eyes gives rise to different temporal dynamics.

3.2 General Methods

Observers

Three observers served as the subjects for all of the following experiments. All had normal or corrected-to-normal vision. Written informed consent was obtained for all observers in accordance with The University of Texas at Austin Institutional Review Board. Observers were treated according to the principles set forth in the Declaration of Helsinki of the World Medical Association. Two of the three observers were authors, and the third (subject 3) was naive concerning the purposes of the experiments.

Apparatus

Stimuli were presented using a Planar PX2611W stereoscopic display. This display consists of two monitors (with orthogonal linear polarization) separated by a polarization-preserving beam splitter (Planar Systems; Beaverton, OR). Subjects wore simple passive linearizing filters to view binocular stereo stimuli. In all experiments, subjects used a forehead rest to maintain constant viewing distance. In Experiment V subjects were fully head-fixed using both a chin cup and a forehead rest. Each monitor was gamma-corrected to produce a linear relationship between pixel values and output luminance.

A Leap Motion controller was used to record the manual tracking data (Leap Motion Inc.; San Francisco, CA). It uses two IR cameras and an infrared light source to track the position of hands and fingers. This device collected measurements of the (x,y,z) position of the observer's pointer finger over time (see Appendix A for an evaluation of the spatiotemporal characteristics of this device).

All experiments and analyses were performed using custom code written in MATLAB using the Psychophysics Toolbox (Pelli, 1997; Brainard, 1997; Kleiner et al., 2007). Subpixel motion was achieved using the anti-aliasing built into the "DrawDots" function of the Psychophysics Toolbox. During trials, observers controlled a cursor by moving their pointer finger above the Leap Motion controller. The experiments were performed with observers sitting at a viewing distance of 85 cm, except the final experiment (Experiment V) in which viewing distance was 100 cm.

Stimuli

In all experiments, subjects tracked the center of a target as it moved in a random walk, controlling a visible cursor with their finger. Each dimension of the random walk (horizontal, vertical and depth) was defined as follows:

$$x_{t+1} = x_t + w_t, \qquad w_t \sim \mathcal{N}(0, \sigma^2) \tag{3.2.1}$$

with time steps corresponding to .05 ms (20Hz). A trial consisted of 20 seconds of tracking.

For Experiments I - IV, the target and cursor were luminance-defined circles (61.5 cd/m^2, .8° diameter; and 71.3 cd/m^2, .3° diameter, respectively) on a gray

background (52.4 cd/m^2). Luminance was measured with a photometer (PR 655, Photo Research; Syracuse, NY) through the beamsplitter and a polarizing lens. For Experiment V the target was a disparity-defined square (width = .8°) created by a dynamic random element stereogram (DRES) (Julesz & Bosche, 1966; Norcia & Tyler, 1984). Both the target and the background were composed of Gaussian pixel noise clipped at 3 standard deviations and set to span the range of the monitor output (mean = 52.4 cd/m^2, max = 102.8 cd/m^2, min = 2.043 cd/m^2; see Figure 3.12 for example). The cursor was a small red square (.1°).

Looming and focus cues (i.e. accommodation and defocus) are both known to be cues for motion-through-depth, but were not rendered in these stimuli. These cues would have been very small for real-world versions of our stimuli (see the General Discussion for more details).

Analysis

Each trial resulted in a time series of target and response positions. In order to examine tracking performance, we calculated a cross-correlogram (CCG) for each trial of the target velocity and the response velocity for the relevant directions of motion (see e.g. Mulligan et al. 2013; Bonnen et al. 2015). A CCG shows the correlation between the target and response velocities as a function of time lag (horizontal offset between target and response time series). An average CCG was computed per subject across all the trials in a condition. The CCGs loosely resembled causal filters or impulse response functions. In fact, for a linear system and white noise, the CCG is an estimate of the impulse response function.

The shape of the average CCG characterizes the overall latency and spatiotemporal fidelity of the tracking response. Some basic features of these CCG response functions, i.e. peak, lag, width, provide simple measures of performance in each condition. The peak is the maximum correlation value. The lag is temporal offset (s) at the peak correlation value. The width refers to the width of the CCG at half the peak correlation (i.e. height). The lag provides a measure of response latency while peak and width are related to the spatiotemporal fidelity. The median was chosen as a summary statistic for these features. While the average CCG was robust, outliers were observed on individual trials, particularly in low amplitude conditions. In order to be consistent across all conditions in all experiments and

avoid ad-hoc methods for excluding outliers, we chose to report the median of the features (peak, lag, width).

For each condition the median and its 95% confidence intervals were estimated via bootstrapping. A bootstrapped data set was generated by resampling the original data set (e.g. the peaks for the horizontal direction for subject 1 in Experiment I) with replacement. The median was calculated for that bootstrapped data set. This was repeated N times ($N = 1000$). The median and 95% confidence intervals of those N medians are reported (see Figures 3.3, 3.7, 3.9, 3.11, 3.14). In many places the 95% confidence intervals may be hidden behind the data because they are relatively small intervals.

We performed several planned comparisons of features across conditions in our experiments. For these comparisons, the effect size (effectively a d') was calculated on the medians:

$$\frac{|m_1 - m_2|}{\sqrt{s_1^2 + s_2^2}} \tag{3.2.2}$$

where m_1 and m_2 are the medians of the respective conditions and s_1 and s_2 are the standard deviations. Because the data contained outliers and did not meet the assumption of normality, we could not perform traditional Student's t-tests for these comparisons. To evaluate significance we sampled (N=100,000) to obtain the distribution of differences between the medians in the two conditions in question. We report the cumulative probability that the difference is <= 0 as our significance value, where the difference is taken in the direction of the effect (effectively a 1-tailed t-test for medians).

3.3 Experiment I. 3D Tracking

Observers tracked the center of a target as it moved in a 3-dimensional Brownian random walk. Analysis of the resulting time series (target path and subject's response path) revealed a selective impairment for tracking motion-through-depth.

Methods

In this experiment, observers were asked to track the center of a luminance-defined stereoscopic target as it moved in a 3-dimensional Brownian random walk ($\sigma = 1$ mm) using their finger to control a visible cursor. Note that when referring to motion in each of the three dimensions we will use the terms horizontal motion, vertical motion, frontoparallel motion (referring to horizontal and/or vertical motion), and motion-through-depth in order to remain consistent with existing literature. The term 3D motion will refer more generally to motion in all directions. The cursor motion was rendered to match the motion of the observer's finger in space, such that when the subject moved their finger one centimeter in a direction, the cursor appeared to move one centimeter in that direction. Observers completed 50 such trials in blocks of 10.

Each trial yielded a pair of x-y-z-time series: the position of the target in 3D space (i.e., the stimulus), and the position of the cursor (i.e., the observer's response). For each trial, we computed a cross-correlogram (CCG; see e.g. Mulligan et al. 2013; Bonnen et al. 2015) of the target velocity and the response velocity for the horizontal, vertical and depth components. We have reported the average across all trials. In Experiment 1, the CCG functions are well fit by a skewed Gabor function (e.g. Geisler & Albrecht 1995), a sine function windowed by a skewed Gaussian (a Gaussian with two σ values, σ_1 above the mean and σ_2 below the mean):

$$f(t) = a * e^{-\frac{(t-\mu)^2}{2\sigma_1^2}} * sin\left(2\pi\omega * (t-\mu)\right), \quad t \geq \mu \tag{3.3.1}$$

$$f(t) = a * e^{-\frac{(t-\mu)^2}{2\sigma_2^2}} * sin\left(2\pi\omega * (t-\mu)\right), \quad t < \mu \tag{3.3.2}$$

where t is the function domain; a is the amplitude, μ is the mean and the offset of the sine wave, and σ_1, σ_2 are the standard deviations, σ_1 for $t \geq \mu$ and σ_2 for $t < \mu$; for the sine function: ω is the frequency of the sine wave. The location of this function's maximum value (i.e. the lag) is equal to $\mu + \sigma_1^2$. The skewed Gabor function was fit to CCGs using least squares minimization. The proportion of the variance explained is used to measure the goodness of fit. This measure is calculated by leave-one-out cross validation. All but one trial was used to perform the fit, then the correlation between the left out trial and the fit is calculated This

value is squared to find the variance explained. This is repeated 50 times, once for each trial and the average is reported.

Results

Figure 3.1 shows tracking time series data for one subject from an example trial (20 seconds). Subjects were able to track the target (gray lines) in each of the cardinal directions (relative to the observer): horizontal (left, blue), vertical (middle, red), and depth (right, black).

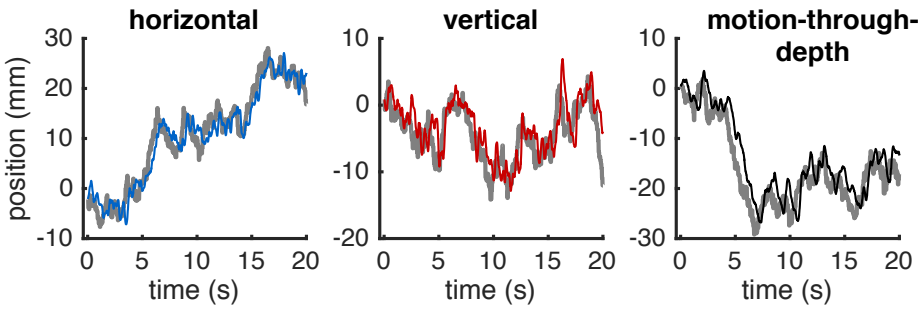

Figure 3.1: Example of data generated by target tracking. These data were taken from a single trial completed by subject 1. Each subplot shows the position for a particular cardinal direction (horizontal, vertical, and depth) over time. In every panel the thick grey line represents the target position. The thinner line in each panel represents the subject's tracking response (horizontal - blue, vertical - red, depth - black).

Figure 3.2 shows the mean CCGs (dots) and 95% confidence intervals (cloud) for each of the three subjects, across all trials using the same color conventions as Figure 3.1. Skewed Gabor functions were fit to the CCGs (see Methods for details). The solid lines in Figure 3.2 correspond to the fits. The proportion of the variance explained across subjects and cardinal directions ranges from 73% to 88% with an average of 82% (see Appendix B: Table 3.1 for goodness of fit values per CCG).

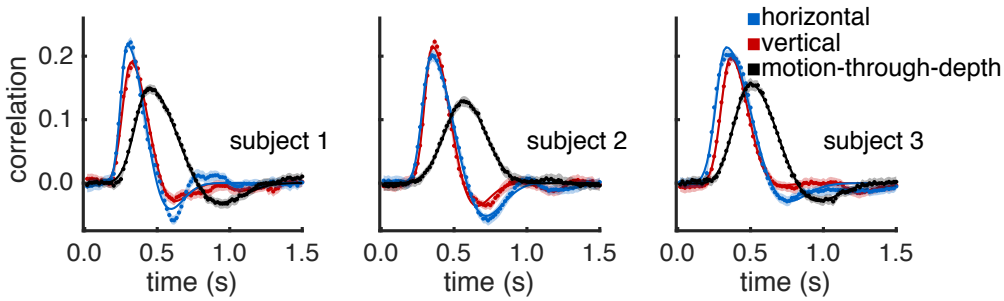

Figure 3.2: Experiment I – 3D tracking. Tracking motion-through-depth is impaired relative to tracking frontoparallel motion. Average cross-correlograms (CCGs) are plotted (dots) for all 3 subjects for horizontal (blue), vertical (red), and motion-through-depth (black). The skewed-Gabor fits of those CCGs are plotted as solid lines. Here and in similar figures to follow, error clouds represent 95% confidence intervals on the data, though these error clouds are often not distinguishable from the data and/or fits. Notice the pronounced difference in tracking performance between frontoparallel motion (horizontal/vertical) and motion-through-depth.

All subjects show a significant impairment for tracking motion-through-depth compared to horizontal and vertical motion. In particular, the depth CCG for each subject has a decreased peak correlation, increased lag (of that peak correlation) and increased width (at half-peak) compared to either of the frontoparallel CCGs (horizontal and vertical). The differences in these features indicate a longer response latency and decreased spatiotemporal precision for tracking motion-through-depth in Experiment I.

Figure 3.3 shows the median lag (first row), peak (second row) and width (third row) for the horizontal, vertical, and depth CCGs for each observer (error bars indicate 95% confidence intervals, see General Methods for how these features are computed). We compared the features of motion-through-depth tracking performance to horizontal motion tracking performance to confirm our previous observations that the depth motion CCGs exhibit increased lags, decreased peaks, and increased widths ($p<1e\text{-}5$ across all comparisons; see Table 3.2 for effect sizes and significance values). These differences are indicative of decreased performance in tracking motion-through-depth across all features.

While there are differences between horizontal and vertical tracking CCGs,

they are all relatively small and idiosyncratic to the observer. For example, notice that S1 shows a slightly lower peak and longer lag for vertical tracking as compared to horizontal. Informally we have observed that these individual differences are relatively stable across time and experimental condition (over the course of 2-3 years of experiments in our lab). However, our primary interest is in the general differences in performance between frontoparallel and depth tracking. Therefore we take horizontal tracking to be representative of frontoparallel tracking for the purposes of comparison.

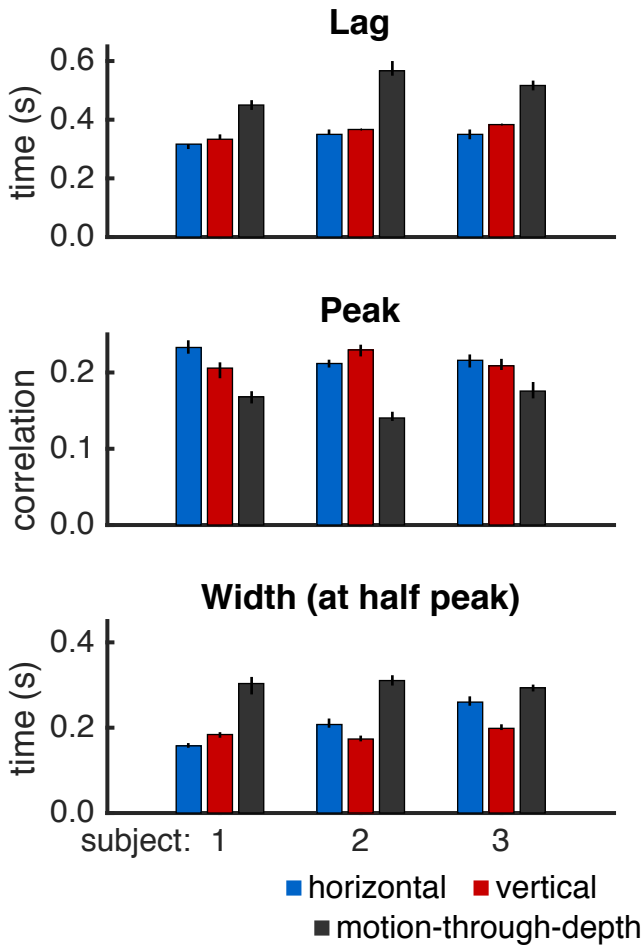

Figure 3.3: Summary of features of tracking performance in Experiment I, calculated from CCGs shown in Figure 3.2. Features (top panel - lags; middle panel - peaks; lower panel - width at half peak) indicate consistently better performance (shorter lags, higher peak correlation values, and smaller CCG widths) for tracking frontoparallel motion as compared to motion-through-depth for a target moving in a Brownian random walk. Bar height indicates median values and error bars show 95% confidence intervals.

Discussion

In Experiment I, observers tracked a target as it moved in a 3-dimensional Brownian random walk. Observers showed impaired tracking performance for tracking the

motion-through-depth compared to the frontoparallel components of motion.

Two potential explanations for this relative impairment are: 1) The egocentric geometry of motion-through-depth results in a smaller signal-to-noise ratio (i.e. the size of the visual signals are much smaller for motion-through-depth vs. frontoparallel motion); and 2) The perception of motion-through-depth involves additional mechanisms in order to make use of binocular signals (e.g. interocular velocity differences, changing disparities). Those additional mechanisms have different spatiotemporal signatures in the context of tracking.

The following experiments examine the contribution of these two explanations to the impairment of motion-through-depth tracking, with Experiments II-IV focused primarily on the ramifications of geometry and Experiment V focused on the role of disparity processing.

3.4 Experiment II. Geometry of 3D motion as a constraint on motion-through-depth tracking performance

The magnitude of the retinal signal resulting from an environmental motion depends on the direction of the motion relative to the observer (see Figure 3.4). In fact, when the viewing distance is large relative to inter-pupillary distance (allowing the small angle approximation), frontoparallel motion and a motion-through-depth result in retinal projections with a relative size of $\sim 1 : \frac{ipd}{d}$, where d is viewing distance and ipd is interpupillary distance. This approximation assumes that the viewing distance is large compared to the interpupillary distances ($d >> ipd$) and that $x \approx 0$ – both are true during the motion-through-depth condition of the following experiment (d=85cm, ipd=6.5cm, x=0). The ratio is calculated by similar triangles, as shown in the diagram in the middle panel of Figure 3.4. For this experiment the ratio is 1 : .08, meaning the motion-through-depth signal is less than 10% of frontoparallel signal. This geometric reality greatly reduces the signal-to-noise ratio (SNR) for tracking motion-through-depth. We examined the ramifications of this geometry in two ways: 1) a set of simulations that used a Kalman filter observer and manipulated signal size and 2) a set of experiments that manipulated signal size for frontoparallel motion tracking and then compared it to motion-through-depth

tracking performance.

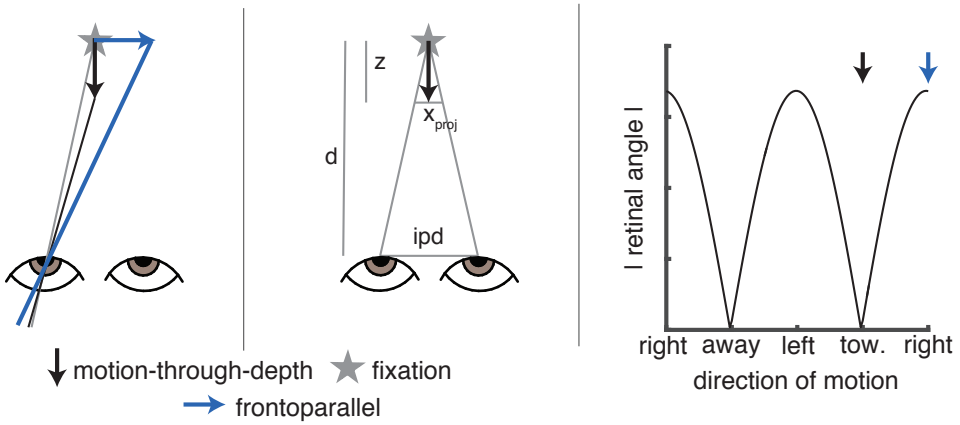

Figure 3.4: Frontoparallel motion and depth motion produce differently sized retinal signals. (Left Panel) For an environmental motion vector that remains the same size regardless of direction, the magnitude of the resulting motion on the retina (measured as the absolute angular difference, i.e. the difference between the grey line and the black/blue lines) is smaller for motion-through-depth (black) than for horizontal motion (blue). (Middle Panel) The approximation of the ratio of the size of the retinal projections for frontoparallel motion vs motion-through-depth is calculated by similar triangles, making the assumption that $d >> ipd$ and x=0. From this diagram we see that $\frac{ipd}{d} = \frac{x_{proj}}{z}$. Let $z = 1$, then $x_{proj} = \frac{ipd}{d}$. (Right Panel) The magnitude of the motion on the retina (retinal angle, black line) is periodic over the environmental motion direction (left/right are large and towards/away are small). Arrows show the two cases illustrated in the far left panel.

For this set of experiments/simulations and all remaining experiments, we shift to reporting σ in arcminutes, because we are now concerned with size of the motion falling on the retina and this is traditionally reported in degrees (or arcminutes) of visual angle. The standard ratio used to convert from mm of motion on the screen to arcminutes is simply: $60 * 2 * atand(\frac{v}{2d})$, where $atand$ is the arctangent in degrees, v is the motion on the screen, and d is the viewing distance.

Kalman filter observer

We simulated tracking at different signal sizes using a Kalman filter observer (ideal observer for the behavioral tracking paradigm; Baddeley et al., 2003; Bonnen et al., 2015). It makes a set of testable predictions for how manipulating SNR should affect measures of performance (e.g. CCGs, peak, lag, and width).

Two equations form a simple linear dynamical system: the random walk of a stimulus ($x_{1:T}$, see equation 3.2.1) and the corresponding noisy observations ($y_{1:T}$) of an observer:

$$y_t = x_t + v_t, \qquad v_t \sim \mathcal{N}(0, R) \tag{3.4.1}$$

where x_t is a target position, y_t is the noisy observation of x_t, σ^2 is the parameter that controls the motion amplitude of the stimulus, R is the observation noise variance, and v_t corresponds to a random variable drawn from $\mathcal{N}(0, R)$ at time t. (See appendix of Bonnen et al., 2015 for additional details). In this formulation $\sqrt{\frac{\sigma^2}{R}}$ is the signal-to-noise ratio. The Kalman filter provides an optimal solution for estimating x_t given $y_{1:t}$, σ^2 and R. Using the Kalman filter as an observer, we simulated responses to a random walk with different values of σ while R remained fixed, effectively manipulating the SNR. Figure 3.5 summarizes the results of this simulation. The values of σ were chosen to correspond to those in Experiment II (below). R was set to 200 arcminutes so that the CCG in the maximum SNR condition had a peak response comparable to the empirical results observed in Experiment II (below). Changing σ systematically changes the optimal Kalman gain (K) which specifies how much a new noisy observation is weighted relative to the previous estimate. The CCGs reflect those changes in the optimal Kalman gain, showing a decreased peak correlation (see lower middle panel of Figure 3.5) and increased width of the CCG (see lower right panel of Figure 3.5). The lag is unaffected (see lower left panel of Figure 3.5.)

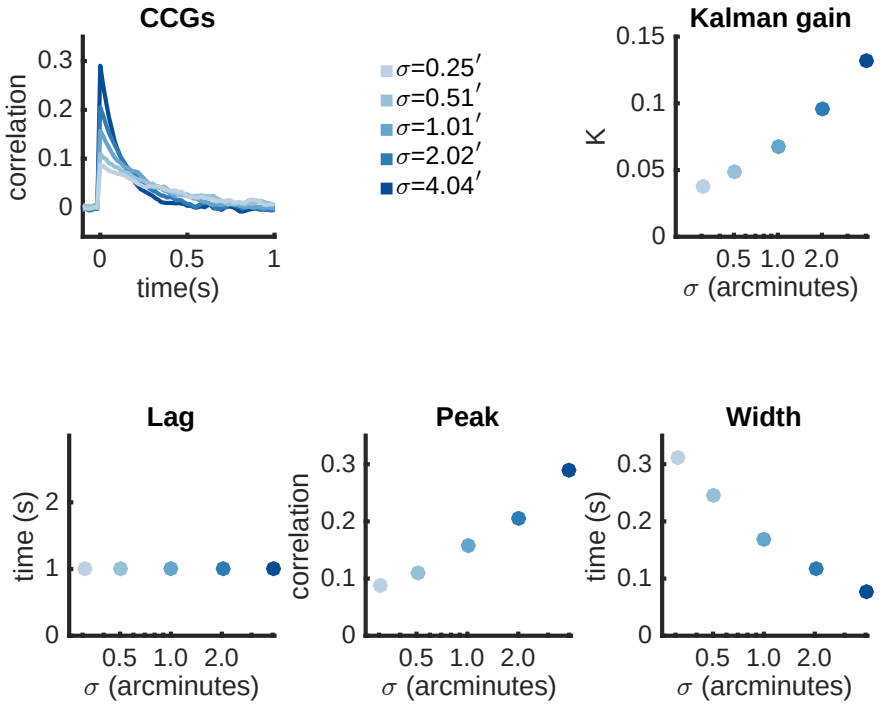

Figure 3.5: Performance of a Kalman filter observer. (Upper Left Panel) The change in the optimal Kalman gain (K) as a function of motion amplitude (σ). Larger values of σ result in larger values of K which results in a higher weighting of new observations. (Upper Right Panel) Average CCGs were calculated for simulated Kalman filter observer responses at each of the values of σ. (Lower Left Panel) Crucially, lag was independent of *sigma*. (Lower Middle Panel) Higher values of σ resulted in higher peak correlations indicating a higher spatiotemporal fidelity. (Lower Left Panel) Higher values of *sigma* resulted in lower widths at half height, indicating a higher temporal precision in the response.

Changes in SNR due to geometry predict some of the differences between frontoparallel motion tracking and motion-through-depth tracking observed in Experiment I: a drop in the CCG peak and an increase of the width of the CCG, but not the change in lag. The following experiment examines the effects of manipulating SNR on human performance, compares human performance to the Kalman filter observer, and makes a comparison between tracking frontoparallel motion and motion-through-depth.

Methods

As in Experiment I, observers tracked the center of the target using their finger to control the cursor. A trial consisted of 20 seconds of tracking the target. In this experiment, observers always tracked 1-dimensional Brownian random walks as in Equation 3.2.1. We manipulated σ, thus controlling the distribution of the size of the motion steps across blocks of trials.

There were 2 distinct conditions: frontoparallel motion tracking and motion-through-depth tracking. In the frontoparallel motion tracking, observers tracked the target as it moved left and right on the x-axis. The motion corresponded to a one dimensional Brownian random walk at $\sigma = 4.04, 2.02, 1.01, .51, .25$ arcminutes; or in pixels at $\sigma = 3.49, 1.75, 0.87, 0.44, 0.22$ – well within the subpixel capabilities of the Psychophysics Toolbox. Cursor responses were also confined to the x-axis. In the motion-through-depth tracking, observers tracked the target as it moved back and forth along the z axis. The horizontal projections of the depth motion corresponded to a one-dimensional Brownian random walk at $\sigma = .51$ arcminutes. Cursor responses were confined to the z-axis.

The geometry of the stimulus/cursor was drawn to match the observers' motion in space, such that when the subject moved their finger one centimeter in a direction, the cursor appeared to move one centimeter in that direction. Observers completed 20 trials in blocks of 10 at each value of σ (indicated previously). The experiment was block-randomized.

Results

Figure 3.6 summarizes the results from this set of experiments in the form of CCGs, while Figure 3.7 plots the lags (top row), the peak correlations (middle row), and the width-at-half-peak (bottom row) for each condition.

Frontoparallel motion tracking: Tracking performance decreases with decreasing motion amplitude, as evidenced by the systematic changes in the CCGs (blue) in Figure 3.6. This manifests specifically as an increased peak and decreased width for an increased σ, with little to no change in the lags (slopes reported in Table 3.3). The changes to frontoparallel motion tracking performance with the manipulation of SNR are inconsistent with the impaired motion-through-depth tracking found in Experiment I in important ways: 1) Decreasing motion amplitude does

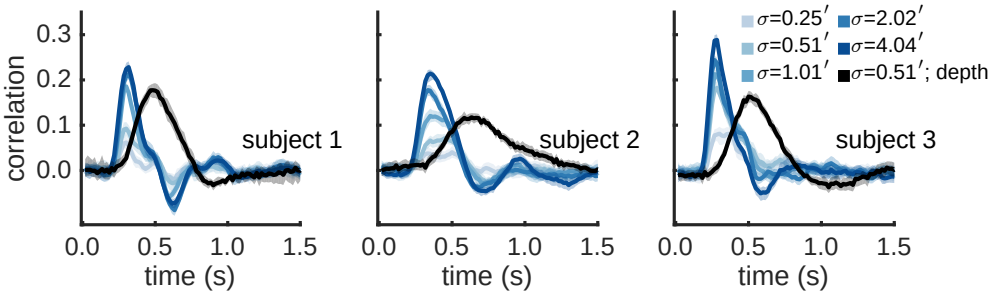

Figure 3.6: Experiment II – Manipulating motion amplitude demonstrates that impairments to motion-through-depth tracking are not purely the result of the smaller SNR. Average CCGs for frontoparallel motion tracking across the 5 different motion amplitudes are shown in blue. Decreased motion amplitude results in decreased tracking performance, primarily manifesting as a decreased peak, with no appreciable change in lag. These behavioral results are comparable to the predictions made in the simulations (see Figure 3.5). The average CCG for the motion-through-depth tracking condition is shown in black. Motion-through-depth tracking performance has an obviously increased lag compared to frontoparallel motion tracking performance; i.e. the peak of the depth CCG is right-shifted compared to the frontoparallel CCGs. The motion amplitude in the motion-through-depth condition matches the 2nd smallest motion amplitude in the frontoparallel motion conditions. However, for subjects 1 and 2 the peak of the depth CCG does not match the peak of the frontoparallel condition with comparable motion amplitude. The peak of the depth CCG is actually more consistent with a higher motion amplitude. Error clouds represent 95% confidence intervals on the data.

not shift the lag of responses; and 2) The horizontal projections corresponding to motion-through-depth tracking in the original experiment have a σ of $\sim$ .25 arcminutes. Performance tracking motion-through-depth in Experiment I, as measured by peak correlation (medians of .17, .14 and .18 for each subject respectively) is better matched by higher σ values in the frontoparallel motion tracking condition experiment.

Relationship of results to Kalman observer The observed behavioral changes with the manipulation to SNR are consistent with those predicted by the Kalman observer: little change in lag, a drop in the CCG peak and an increased CCG width (see Figures 3.5, 3.6, and 3.7). We should note that the simulations predict a stronger relationship between width and *sigma* value, than we have reported. We believe that this is largely due to limitations in the Kalman filter as an ideal observer of human behavior. Human observers do not have perfect signal transmission of position/motion through the visual system. Previous work has described monocular/binocular temporal filters that would impart some of the features observed in our CCGs, particularly the negative lobes (Neri, 2011; Nienborg et al., 2005). This difference in shape due to temporal filters would likely have an effect on our measures of width.

Motion-through-depth tracking: Subjects performed an additional motion-through-depth tracking condition in order to generate data directly comparable to the frontoparallel motion tracking condition in this experiment, i.e. visual signal size matched the second smallest motion amplitude condition performed during frontoparallel motion tracking (σ = .51 arcminutes). This CCG corresponding to this condition is plotted in black in Figure 3.6. The motion amplitude of this condition makes it directly comparable to the frontoparallel condition marked by the second lightest blue in Figures 3.6 and 3.7. We compared the peak, lag and width of the CCGs for these two conditions (see Appendix B, Table 3.4 for effect sizes and significance values). Again, motion-through-depth tracking performance exhibits an increased lag not present in the frontoparallel motion tracking (p<1e-5). For subjects 1 and 2, the peak correlation for motion-through-depth tracking was better matched by a higher value of σ in the frontoparallel motion condition (p<1e-5), but for subject 3 there was no significant difference (p=.07).

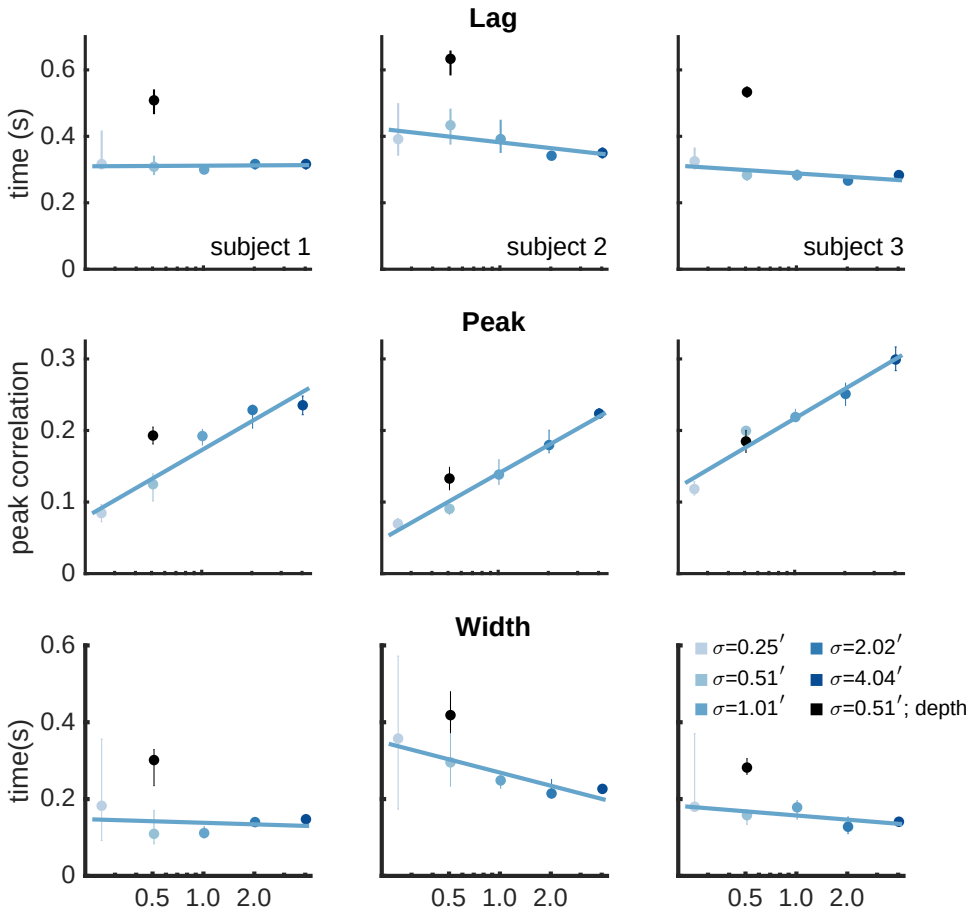

Figure 3.7: Summary of features of tracking performance for frontoparallel motion tracking and motion-through-depth tracking shown in Figure 3.6. Color corresponds to condition, error bars indicate 95% confidence intervals, and the lines correspond to least squares fits of the frontoparallel data. (Top Row) Median lags, (Middle Row) median peak correlations, and (Bottom Row) median width values for all 3 subjects. Peak correlation increases as a function of σ. Lag changes very little. Width has a negative relationship with σ. See Appendix B, Table 3.3 for slope values. With one exception (subject 3, peak), the point corresponding to depth tracking is clearly afield from the line describing the frontoparallel data.

Discussion

Changing motion amplitude (as we have done in this experiment) is in essence changing the range of retinal signal sizes and thus the overall SNR, given some fixed level of internal noise. Tracking performance (as measured by peak, i.e. spatial fidelity) does decrease with decreasing motion amplitude and it is reasonable to conclude that the difference in visual signal size contributes to the impairment to motion-through-depth tracking.

However, a direct comparison of frontoparallel motion and motion-through-depth tracking performance when visual signal size is equivalent (at $\sigma = .51$ arcminutes) revealed substantial differences in performance. Motion-through-depth tracking had an increased latency and for some subjects different spatial fidelity. Based on these results, we conclude that the impairment to tracking depth motion observed in Experiment I is not completely explained by the difference in visual signal size due to geometry.

Experiment III and IV address concerns about whether differences in performance are due to differences in motor control. Experiment III changes the cursor motion in the frontoparallel motion condition so that it matches the cursor motion to visual signal size ratio of the motion-through-depth condition. This accounts for much of the remaining difference in the spatial fidelity of the tracking response. Experiment IV examines whether physical differences in arm motion direction play a significant role in tracking performance.

3.5 Experiment III. Frontoparallel cursor motion consistent with motion-through-depth tracking

The one component of the geometry not equivalent across frontoparallel motion tracking and motion-through-depth tracking in Experiment II was the cursor gain. We intentionally drew the cursor to match the observers' motion in space, so that when the subject moved their finger one centimeter in a direction, the cursor appeared to move one centimeter in that direction. Consider the motion-through-depth and frontoparallel motion conditions where retinal motion amplitude was equivalent ($\sigma = .51'$). Because retinal signals of equal size result in much larger environmental motions in depth, subjects had to move more on average in the motion-through-

depth conditions. That difference in movement also provides finer control of the cursor and may cause delays during depth motion tracking. In this experiment, subjects performed frontoparallel motion tracking using a cursor with the same gain as motion-through-depth tracking (i.e. larger finger motion was required).

Methods

As in Experiment II, observers tracked the target as it moved in a one-dimensional Brownian random walk along the x-axis (at $\sigma = .51$ arcminutes). Like the regular frontoparallel motion tracking condition, cursor responses were confined to the x-axis. However, the gain of the cursor motion was set to match the gain associated with motion-through-depth tracking, such that a larger motor movement was required to produce a relatively small cursor movement. Each subject completed 20 trials in this condition.

Results

Changing the cursor gain to match the visual signal size does have an impact on tracking performance. Figure 3.8 illustrates that frontoparallel, gain-corrected tracking performance more closely resembles motion-through-depth tracking (replotted in black) than frontoparallel tracking only (see Figure 3.6 for reference).

Figure 3.9 shows the remaining differences between performance for motion-through-depth tracking versus gain-corrected, frontoparallel motion tracking. Even in the gain-corrected frontoparallel motion tracking, there was still a difference in lag compared to motion-through-depth tracking, while the difference in CCG peak height diminished substantially. The CCG width only remained significantly different only in the case of subject 1.

Discussion

The impaired performance in motion-through-depth tracking originally demonstrated in Experiment I is partially explained by geometry as illustrated in Experiments II and III. However there remains at least an unexplained lag (or increased latency) in the motion-through-depth tracking response. One possibility that must be eliminated is that this temporal delay is simply due to motor differences between fron-

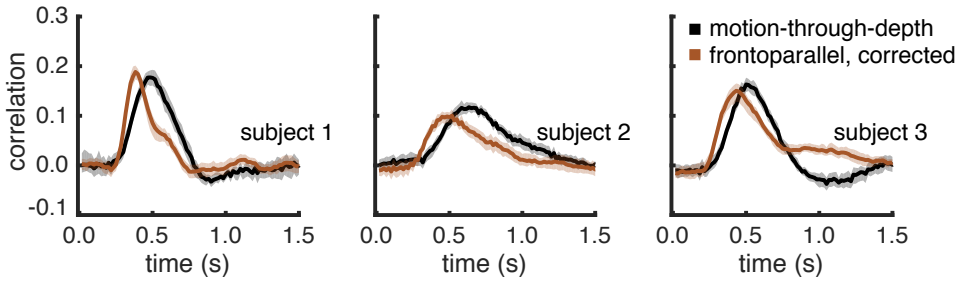

Figure 3.8: Experiment III – Cursor motion consistent with visual signal size (i.e. gain-corrected) still cannot account for the difference in latency for motion-through-depth tracking. The average CCG for the condition with gain-corrected cursor motion is shown in brown. The average motion-through-depth tracking CCG from Experiment II (see Figure 3.6) is replotted in black for each subject. Error clouds represent 95% confidence intervals on the data. The gain-correction accounts for some of the difference between frontoparallel motion and motion-through-depth tracking seen in Experiment I. However, there remains a increased lag for motion-through-depth tracking ($p<1e\text{-}3$ for all subjects; see Appendix B, Table 3.5).

toparallel motor movements (left, right, up, down) and depth motor movements (forward and backward).

3.6 Experiment IV: Frontoparallel Cursor Motion and Cursor Control Consistent with Motion-Through-Depth Tracking

During tracking, subjects move their finger left and right, up and down, and back and forth. Each of the cardinal directions is tied to one of these arm/finger motions. It is possible that some or all of the latency differences are due to the different motor demands of moving the arm/finger back and forth. In order to determine the contribution of these motor differences, subjects performed a frontoparallel motion tracking task in which we manipulated whether a subject controlled XY cursor motion with XY finger motion (as in all the previous experiments) or controlled XY cursor motion with XZ finger motion (as one does when using a mouse or trackpad where motion towards the screen moves the cursor up the screen).

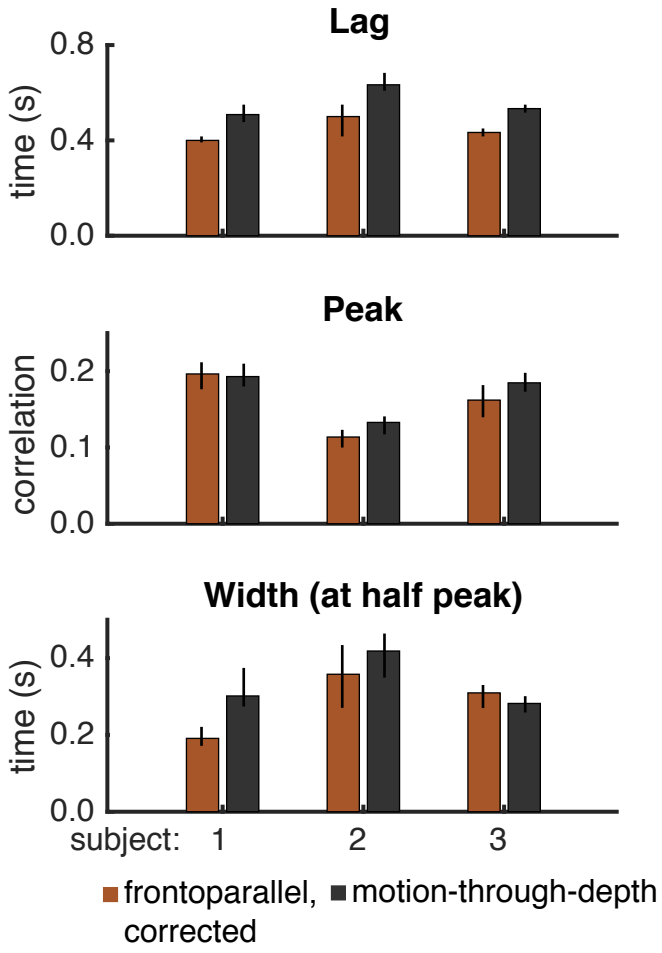

Figure 3.9: Summary of features of tracking performance for gain-corrected frontoparallel motion tracking and motion-through-depth tracking. (Top Panel) Median lags, (Middle Panel) peak correlations, and (Bottom Panel) median widths, for motion-through-depth (black), and gain-corrected frontoparallel (brown) tracking for each of the subjects. A pronounced and consistent difference in lags remains between motion-through-depth tracking and frontoparallel motion tracking. In the case of peak correlation and width, corrected gain accounts for the majority of the difference. Error bars represent 95% confidence intervals.

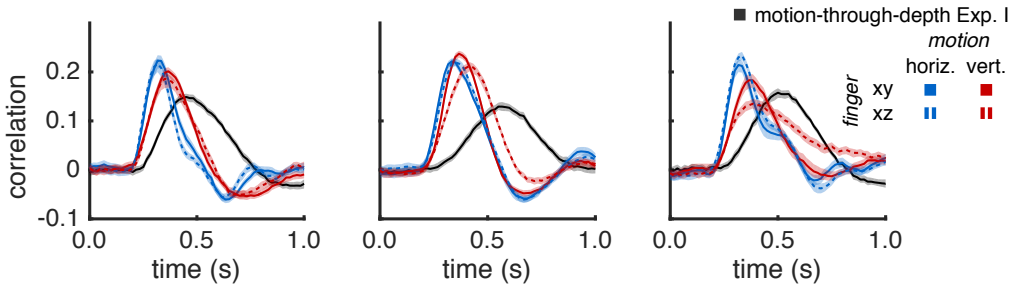

Figure 3.10: Experiment IV: Manipulating the finger motion axis (XY vs XZ) cannot account for the difference between frontoparallel and motion-through-depth tracking. Average CCGs are shown for XY motion tracking using XY finger motion (solid lines) and using XZ finger motion (dashed lines) where the horizontal motion CCGs are in blue and the vertical motion CCGs are in red. The average depth CCG from Experiment I is replotted in black for convenient comparison. Error clouds represent 95% confidence intervals on the data.

Methods

Observers tracked the target as it moved in a two-dimensional Brownian random walk ($\sigma = 4.04$ arcminutes) in xy-space. Cursor responses were confined to the x and y axes. In the XY condition, finger motion along the y-axis controlled cursor motion on the y-axis. In the XZ condition, finger motion along the z-axis controlled cursor motion on the y-axis. Each trial was 20 seconds and two observers performed 50 trials in each condition.

Results

Average CCGs were calculated for the horizontal and vertical motion (see Figure 3.10). Note that subject 1 shows nearly identical performance across the two conditions. The main question here is whether the motor demands of moving forward and backward could result in the impairments observed in Experiment I. Thus, the main comparison here is between/ the vertical motion - XZ tracking performance and the motion-through-depth tracking.

Features of tracking performance are calculated for motion-through-depth, vertical - XZ, and vertical - XY. The comparison of interest is between vertical - XZ and motion-through-depth, but vertical - XY is provided for reference. The comparison between features of vertical - XZ tracking and motion-through-depth

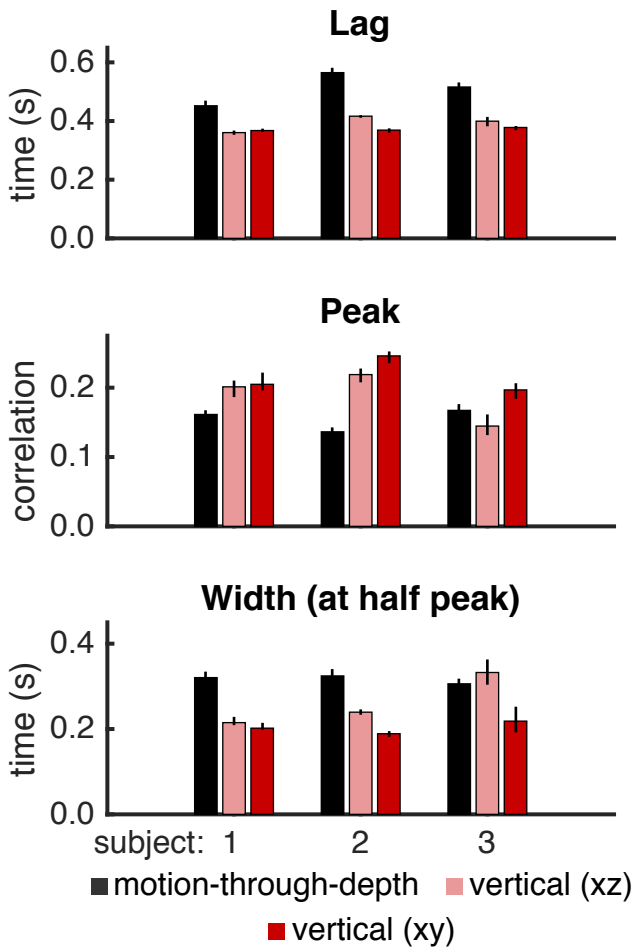

Figure 3.11: Summary of features of tracking performance from depth (black), vertical - XZ (light red), and vertical - XY (red) CCGs pictured in Figure 3.10. Features (top panel - lags; middle panel - peaks; lower panel - width at half peak) indicate consistent difference between motion-through-depth tracking and vertical - XZ tracking. Bar heights indicate median values and error bars show 95% confidence intervals.

tracking performance demonstrates large differences in tracking performance for most features/subjects (see Table 3.6). Again, there is a consistent difference in the lags.

Discussion

Tracking XY motion in XZ does not account for the huge differences in lags between frontoparallel motion and motion-through-depth tracking observed in Experiment I. The largest difference remaining between motion-through-depth tracking and frontoparallel motion tracking is that disparity computations are required to perform the motion-through-depth tracking. We hypothesized that the remaining response delay for motion-through-depth tracking in Experiment III is the consequence of disparity processing. This hypothesis is explored in Experiment V.

3.7 Experiment V. Disparity processing as a constraint on motion-through-depth tracking performance

Previous experiments cannot entirely account for the difference between frontoparallel motion tracking and motion-through-depth tracking performance. The remaining difference is primarily in the latency of the tracking response. However, frontoparallel motion tracking does not require processing of binocular signals, e.g. binocular disparities or IOVDs. In this experiment we imposed disparity processing on frontoparallel motion tracking. Subjects tracked disparity-defined target created by a dynamic random element stereogram as it moved in a 3-dimensional random walk. We also applied what was learned in Experiment II and III, adjusting the amplitude of the depth motion to increase its visual signal size, and matching the cursor gain across directions.

Methods

In this experiment, observers were asked to use their finger to track the center of a disparity-defined square target created by a dynamic random element stereogram (DRES, see General Methods), using a cursor (small red luminance square). The geometry of the stimulus/cursor was drawn as in Experiment III, so that the geometry of the cursor motion was matched across frontoparallel and depth motion

according to visual signal size. A trial consisted of 20 seconds of tracking the target as it moved in 3-dimensional Brownian random walk ($\sigma = 2.02'$ in horizontal and vertical dimensions, $\sigma = .79'$ in depth). The size of σ for depth was adjusted per subject, so that the average CCG height matched, however σ was the same for all three subjects. Observers completed 30 trials in blocks of 10.

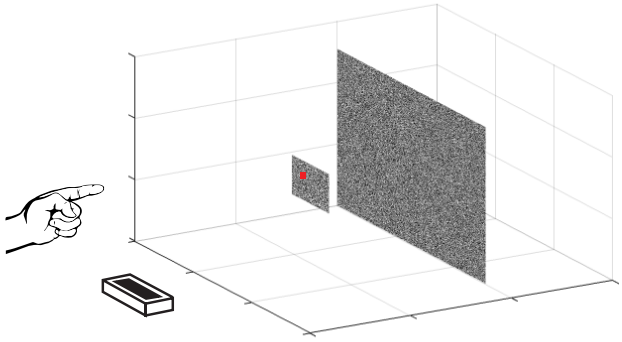

Figure 3.12: Example of the Dynamic Random Element Stereogram (DRES) stimulus. The target was constrained to be in front of the background. Both the target and the background were composed of Gaussian pixel noise that updated at 60Hz.

Results

Figure 3.13 summarizes the results from this experiment. Average CCGs are shown for frontoparallel motion and depth motion directions. The latency difference present previously is now negligible. The amplitude adjustment required to match the CCG peak height was a ratio of 1:6 for frontoparallel vs. depth. This is much smaller than ∼1:15.4 predicted by the size of the disparity signal alone ($1 : \frac{distance}{ipd}$, where viewing distance was 100cm and ipd was 6.5cm).

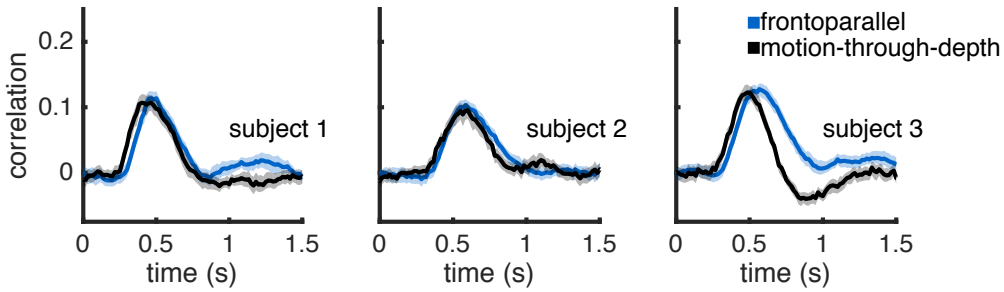

Figure 3.13: Experiment V – Imposing disparity processing on frontoparallel motion results in performance similar to motion-through-depth. Average CCGs for frontoparallel motion (blue) and motion-through-depth (black) during DRES tracking. Error clouds represent 95% confidence intervals on the data. The latency difference between the frontoparallel motion and motion-through-depth CCGs is negligible, or reversed (subject 3; see Table 3.7).

Figure 3.14 shows the peak, lag and widths for frontoparallel and depth motion. The lags for all subjects are either not significantly different or reversed, the peaks are not significantly different (by design) and the widths are not significantly different for subject 2 (see Table 3.7 for effect sizes and significance).

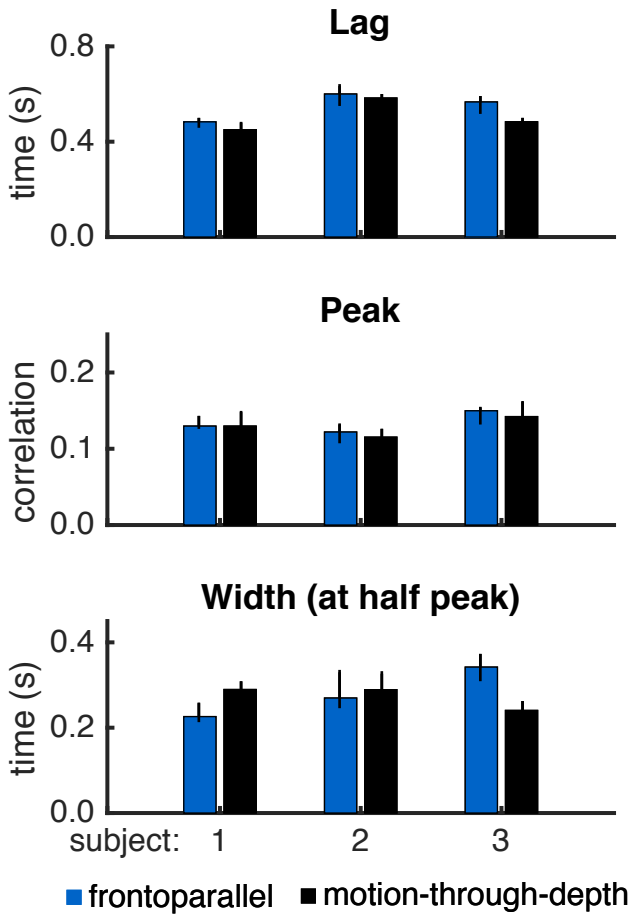

Figure 3.14: Summary of features of tracking performance from depth (black) and frontoparallel (blue) CCGs pictured in Figure 3.13. Bar heights indicate median values and error bars show 95% confidence intervals. Features (top panel - lags; middle panel - peaks; lower panel - width at half peak) are similar across motion-through-depth tracking and frontoparallel tracking. In particular the latency difference between frontoparallel motion tracking performance and motion-through-depth tracking performance is negligible or reversed (See Table 3.7).

Discussion

By creating a disparity-defined target we imposed binocular disparity processing on frontoparallel motion tracking and removed monocular cues and interocular velocity differences as potential sources of information. This resulted in nearly matched latencies between frontoparallel motion and motion-through-depth tracking.

Although motion-through-depth amplitude had to be adjusted to better match the CCG peak height across directions, it did not have to be adjusted as much as is predicted by the geometry of visual signal size. It is possible that motion along the depth axis is privileged in disparity processing (see General Discussion).

3.8 General Discussion

Our primary finding was that tracking performance involves an impairment for the perception of motion-through-depth relative to frontoparallel motion. This is consistent with limitations found for vergence vs. version responses during eye tracking of visual targets (Mulligan et al. 2013; see their Figure 4), demonstrating that this perceptual impairment is still present in the much 'quicker' oculomotor plant. After accounting for differences in visual signal size, this impairment to the perception of motion-through-depth is primarily a temporal difference, a lag in the response, which is attributed to disparity processing.

Throughout the course of this paper, we have examined and directly compared frontoparallel motion and motion-through-depth tracking performance. However the data in Experiment I was collected for target/cursor motion in all directions, not for the cardinal directions in isolation. The choice to perform the CCG analysis on the cardinal directions was an arbitrary one in many respects. We can describe the tracking performance in greater detail by systematically calculating a CCG for axes of motion along the sphere of possible directions. Figure 3.15 shows such an analysis for subject 1. The CCGs are calculated at 5°intervals around the XY, XZ, and YZ planes. Then the CCGs are plotted as a heatmap in polar coordinates where θ is the direction of motion the CCG was calculated on and ρ is the lag. The main 3D heatmap shows CCGs from the XY, XZ, and YZ planes simultaneously. The smaller 2D heatmaps show the analysis on each of the 3 planes. Note that this analysis is sign-invariant and thus there is 180°rotational symmetry.

The peaks of the CCGs on the 2D heatmaps form visible 'rings'. These rings are fairly circular for the XY (frontoparallel) plane, i.e. frontoparallel motion tracking. This is unsurprising given the relative consistency between the previously calculated vertical and horizontal CCGs. The bottom two rows show the same analysis for the YZ (sagittal) plane and the XZ (horizontal) plane. The elliptical nature of these heat maps clearly demonstrate the difference in depth vs frontoparallel while also revealing the the progression of tracking characteristics in between.

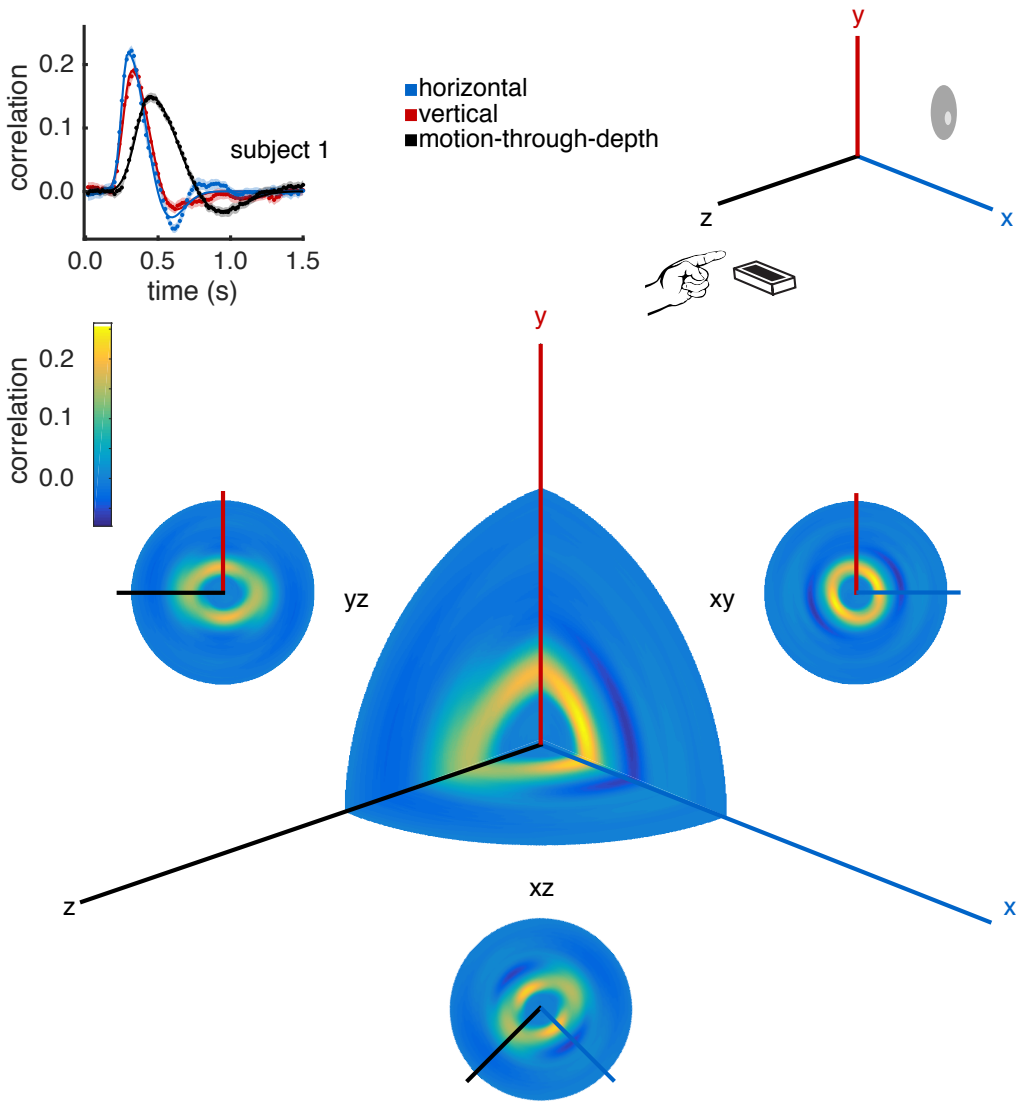

Figure 3.15: Tracking performance across many directions. Here we show an analysis of tracking performance extended to all possible motion directions for subject 1. (Upper Left) The CCGs for the cardinal motion directions are replotted from Experiment I, Figure 3.2. (Upper Right) A schematic of the tracking paradigm in Experiment 1. A subject tracks a circular target, reporting its position by controlling a cursor with their pointer finger. *(continued on the next page)*

Figure 3.15: *(continued)* (Lower) At the center, the main 3D heatmap shows CCGs from the xy, xz, and yz planes simultaneously, where the gray scale axis represents correlation, θ is the direction of motion, and ρ is the lag (See text for additional details). We also show the full CCG heatmap for each plane: xy, xz, and yz. Note the elongation of the peak correlation ridge near the z axis, and the presence of negative lobes on the xy plane, but not along the z direction in the other two planes.

Frequency-domain analysis of 3D motion tracking

Here we re-examined the results of Experiment I in the frequency domain. The frequency domain responses was computed on a trial-by-trial basis, and the resulting amplitude and phase responses were averaged within subject and condition to yield mean gains and phase lags as a function of temporal frequency. Figure 3.16 summarizes this analysis as a Bode plot for each of the three subjects. The top row shows the response gain as function of frequency and the bottom row the phase (absolute, unwrapped) as a function of frequency.

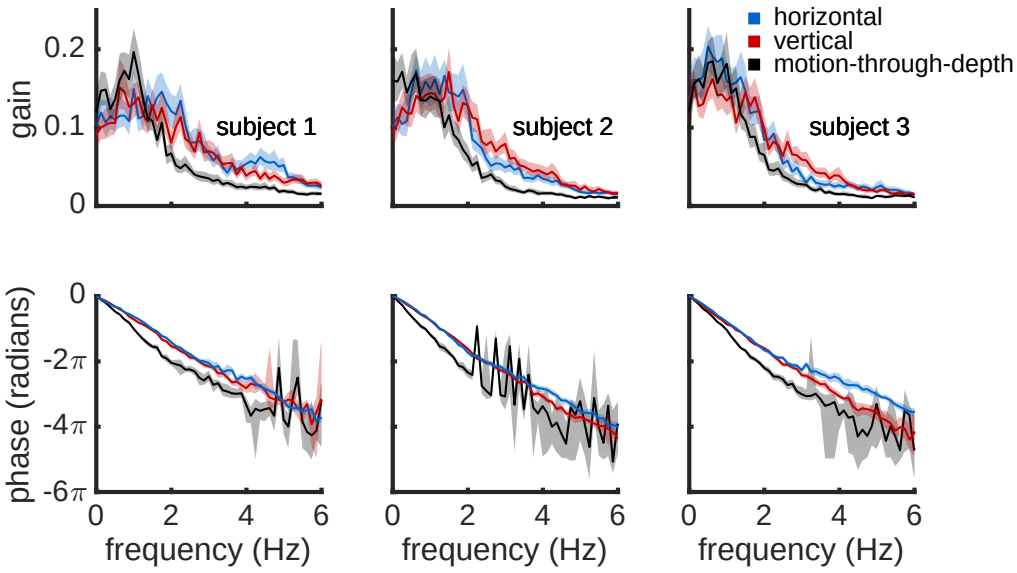

Figure 3.16: Bode plots showing the responses of the three subjects (columns) in the temporal frequency domain. The top row shows response gain and the bottom row shows response phase, both as a function of frequency. Consistent with the time-domain cross-correlation analysis, all subjects show a larger response lag for motion-through-depth tracking relative to frontoparallel tracking (bottom row). Moreover, response gain is higher for frontoparallel tracking above around 1.5 Hz and, for two of the three subjects, response gain for motion-through-depth tracking peaks at a lower temporal frequency than does that for frontoparallel tracking. Shaded regions show bootstrapped 95% confidence intervals.

Recall that the stimulus motion was a Brownian random walk in position and Gaussian white noise in velocity. Thus the stimulus velocities were also Gaussian white noise in the frequency domain. Although all frequencies are equally represented in the stimulus, the analysis presented in Figure 3.16 demonstrates that subjects primarily track the low frequencies and that this is even more pronounced for motion-through-depth tracking compared to frontoparallel tracking. Note also the consistently larger phase lags for motion-through-depth tracking where reliable responses were obtained. This result is supported by previous psychophysical and electrophysiological work that demonstrated poorer temporal resolution for dispar-

ity modulation (Norcia & Tyler, 1984; Lu & Sperling, 1995; Nienborg et al., 2005) compared to contrast modulation (Kelly, 1971; Kelly, 1976; Hawken et al., 1996; Williams et al., 2004). Furthermore, the inability to track higher frequency modulations also provides a reasonable explanation for why the correlation values in the reported CCGs are overall quite low.

The role of visual signal size in motion-through-depth tracking performance

Experiment II explored the role of visual signal size and SNR in tracking performance by manipulating frontoparallel motion and motion-through-depth tracking so that they had matched visual signal size. We concluded that subjects' depth tracking performance had an increased latency and an improved spatial fidelity (for 2 of 3 subjects) compared to the frontoparallel condition (see Figure 3.7). A follow-up experiment (III) suggested that the observed spatial improvement was actually related to the gain on cursor control, leaving just a latency difference between the characteristics of motion-through-depth tracking and frontoparallel motion tracking.

It is surprising that the overall spatial fidelity of motion-through-depth tracking performance and frontoparallel motion tracking performance is approximately equal (after we account for the differences of geometry). Classical demonstrations of "stereomotion suppression" Tyler (1971) led us to expect that subjects should show spatial fidelity deficits in motion-through-depth tracking performance relative to frontoparallel. However the differences in our experimental task provide an explanation. In Tyler (1971), subjects set the amplitude of a sinusoidal motion at the threshold of their perception. The moving bar oscillated sinusoidally either in depth or horizontally about a reference. Thresholds were consistently higher for depth motion across all frequencies, i.e. two eyes were less sensitive than one *at threshold*. Thresholds for frequencies above .5 Hz were consistently between .2 and .8 arcminutes. Similar threshold ranges have been found for static disparities (Badcock & Schor, 1985), which may be a better comparison since our motion stimulus is not a single sinusoid. In our experiment, subjects tracked a target moving in all directions with a visible cursor. They were instructed to keep the cursor center on the target in all dimensions (or one, depending on the condition). We examined distribution of disparity between the target and the cursor (see Figure 3.17) during the motion-

through-depth tracking task in Experiment II. Given a conservative threshold of 50 arc seconds (Tyler, 1971; Badcock & Schor, 1985), a high proportion of the trials are spent with the target and the cursor at supra-threshold relative disparities (83%, 86%, 84% for each subject respectively). This high proportion of supra-threshold disparities provides a plausible explanation for why we do not observe the deficits that might be predicted by previous work on disparity processing.

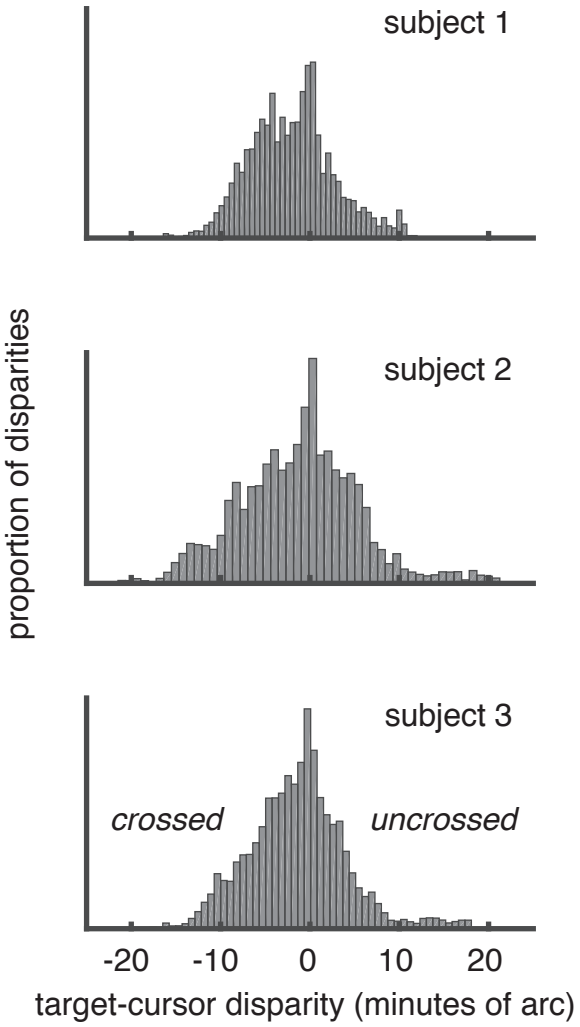

Figure 3.17: Histogram of relative disparity between target and cursor at each time step (.1667s) across all motion-through-depth tracking trials (see Experiment II) for each of the subjects. Given a conservative estimate of disparity threshold (50 arcsec), these histograms demonstrate that subjects spent a high proportion of the trials (83%, 86%, 84% for each subject respectively) with the target and the cursor at supra-threshold relative disparities.

Potential cue-conflicts: Accommodation, defocus and looming

There are several known cues for motion-through-depth which have not been rendered for these experiments (accommodation, defocus and looming). The absence of these cues has the potential to cause cue-conflict for motion-through-depth stimuli. However, based on further analysis of the results and comparisons to the perceptual thresholds for those cues, the presence of cue-conflicts is unlikely.

Figure 3.17 demonstrates that the bulk of relative disparities in our experiment were between -10 and +10 arcminutes, or roughly -.05 to +.05 diopters. Accommodation thresholds are conservatively ∼0.1D (Wang & Ciuffreda, 2006). Given the relative disparities between the cursor/target, and the assumption that during a trial subjects were looking at the target or the cursor (or somewhere in between), we can conclude that the majority of the time accommodation cues were sub-threshold. Similarly, the depth of field of the eye is typically reported as between 0.1D and 0.5D (Walsh & Charman, 1988) and thus the range of predicted disparities is well below that threshold.

In the case of looming, the extent of the motion relative to viewing distance is quite small. The maximum extent in depth (either towards or away) is 5 cm from the starting point; the mean is 2.4 cm. This translates to a maximum change in target size of 3.1 arcminutes and a mean of 1.5 arcminutes over the course of a 20 second trial. The change per stimulus update (20 Hz) was smaller: a maximum of .4 arcminutes, mean of .2 arcminutes. Looming has been studied primarily with stimuli moving in a sinusoid. Motion-through-depth based on looming cues alone are detectable when the amplitude of the oscillations are .5-2 arcminutes, depending on the frequency (Regan & Beverley, 1979). While the change over the course of the trial is in the perceptible range, the individual stimulus updates are not. Furthermore, our stimuli moved in a random walk, resulting in a looming cue that lacked a consistent change in size over time, which would probably result in higher thresholds for detection.

Binocular cues for perception of motion-through-depth

With the results of Experiments I-IV in mind, we considered the remaining impairment, which was primarily a difference in response latency. Even after accounting for the geometry inherent in depth motion, the perception of motion-through-depth

appears to involve binocular mechanisms that exhibit different spatiotemporal signatures in the context of tracking.

Experiment V examined the role of disparity processing (a binocular mechanism) in tracking. We generated a disparity-defined target using a Dynamic Random Element Stereogram (DRES) (Julesz & Bosche, 1966; Norcia & Tyler, 1984). Imposing disparity processing on frontoparallel motion tracking removed the latency differences between frontoparallel motion and motion-through-depth tracking, suggesting that the latency difference is a signature of binocular disparity processing.

Psychophysical and electrophysiological work has shown a poor temporal resolution for disparity modulation (Norcia & Tyler, 1984; Nienborg et al., 2005) compared to contrast modulation (Kelly, 1971, 1976). Psychophysical work on static disparities also shows evidence for a temporal delay (>100ms) for binocular disparity processing (Neri, 2011). Nienborg et al. (2005) provides an explanation for the poorer temporal resolution in processing binocular disparities that explains both the behavioral and neuronal temporal resolution deficits observed in previous work. Though the differences in temporal resolution between disparity modulation and contrast modulation apppears to suggest separate mechanisms for disparity tuning and contrast tuning, they can be explained by a binocular cross-correlation (i.e. disparity energy model, Ohzawa 1998). Models of disparity selectivity in neurons require the calculation of the cross-correlation between signals from the left and the right eye, temporally broadband monocular images that are already bandpass filtered. The result of the cross-correlation of pre-filtered signals is a low-pass response for binocular signals compared to the equivalent monocular signal. Thus poorer temporal resolution is expected for responses to disparity signals – this may be related to the temporal deficits observed for motion-through-depth tracking and disparity in particular in our experiments.

Neri (2011) also suggests that some of the temporal dynamics of disparity processing are due to a rigid order for processing in which coarse processing precedes and constrains the finer, more detailed processing, an idea which is supported by electrophysiological work (Norcia et al., 1985; Menz & Freeman, 2003). In fact, Samonds et al. (2009) demonstrates that disparity selectivity may continue to sharpen as much as 450-850 ms after stimulus onset. Qualitatively similar results have been found in V1 for orientation (Ringach et al., 2003), and spatial frequency (Bredfeldt & Ringach, 2002), although these sharpening effects appear to evolve over shorter

time scales than those found for disparity processing. Further work is needed examining the temporal dynamics of physiological responses to static and changing disparities, in order to better understand its connection to temporal dynamics in behavior.

This experiment and its conclusion focuses primarily on a single binocular cue: changing disparity. However, early studies of depth motion perception point out that there are two potential binocular sources of information for motion-through-depth: inter-ocular velocity differences (IOVD) and changing disparities (CD)(Regan & Beverley, 1973b). In principle these provide the same information. However, they differ in the order of operations resulting in either a binocular comparison of velocities (IOVD) or a temporal comparison of disparities (changing disparity). Researchers have debated which of these cues is predominant in the visual system (Cumming & Parker, 1994; Rokers et al., 2009; Czuba et al., 2011). Unfortunately, the nature of the target-tracking task is such that we cannot isolate the IOVD cue, like we isolated the CD cue in Experiment V.

It is also worth noting that the statistics of the random walks used across all the experiments in this work may not result in motion stimuli that are ideal for IOVD cues. The frontoparallel and depth noise velocities were white, meaning that the velocity at a given time point was not correlated with the time points around it. This means that the IOVD signal is not as predictable as the CD signal, which involves comparing positions that are correlated and is consistent with the notion that the IOVD signal doesn't have a huge effect on motion-through-depth tracking performance in this paradigm. Recent work suggests that the visual system might use different sources of binocular information depending on the relative fidelity of cues in a situation or the demands of a particular task (Allen et al., 2015).

Privilege for processing motion-through-depth in disparity-limited stimuli?

The stimulus used in Experiment V adjusted the depth motion amplitude so that the CCG peak height was matched between horizontal and depth motion directions. The same motion amplitude value was used across all three subjects. However, this value was not as high the ratio derived for the relationship between the magnitude of frontoparallel motion and the retinal projections of depth motion ($1 : \frac{ipd}{d}$, see

Introduction to Experiment II). The conclusion that we draw from this is that perhaps there is a privilege for processing binocular disparities associated with motion-through-depth.

However, there is little existing evidence to support this observation. Apparent motion studies using Julesz's dynamic random dot stereogram (Julesz & Bosche, 1966) for left-right motion detection and towards-away motion detection find comparable detection thresholds of approximately 5 Hz (Julesz & Payne, 1968; Norcia & Tyler, 1984). This is clearly not a privilege for depth motion, but unlike experiments with monocular cues, it does not find a deficit for motion-through-depth.

Regan & Beverley (1973a) examined detection of 'sideways' vs. depth motion in random dot stereograms (not dynamic - so there were still monocular cues). At the very slowest frequencies the detection thresholds were comparable in one subject but for the most part, monocularly viewed motion detection thresholds were better. In addition, early work with oscillating bars demonstrated better motion detection for monocularly viewed vs. stereoscopically viewed oscilating bars (Tyler, 1971; Regan & Beverley, 1973b), with the potential exception of the ± 5 arcminutes around fixation (Regan & Beverley, 1973b). These findings do not support the idea that there is a privilege for processing binocular disparities associated motions through depth. However, the stimuli used in these previous experiments were not purely disparity-defined. It is possible that the monocular cues present in the stimuli obscure the privilege for motions through depth during disparity processing.

More work is needed to show if there is indeed a privilege for motion-through-depth in disparity-defined stimuli, and in particular to establish how it changes from threshold to suprathreshold motions, across different types of motion stimuli (i.e. from motion that oscillates to random walks).

Conclusions

Despite the crucial importance of egocentric depth motion, we found significant impairment for depth motion perception as compared with frontoparallel motion perception. However, closer examination revealed that these deficits were relatively consistent with the geometry of the stimulus and the limitations of the binocular mechanisms used to perceive the motion.

3.9 Appendix A: Leap Motion Controller

A Leap Motion controller was used to collect measurements of 3D position (x,y,z in mm of our observers' fingers.) The Leap Motion controller is a 8cm x 1cm x 3cm USB device that uses two IR cameras and an infrared light source to track hands, fingers and 'finger-like' tools, reporting back positions and orientations. A line drawing of the device is shown in Figure 3.18. Leap Motion, Inc reports that the device has a field of view of 150°, with an effective range of 2.5cm - 60cm above the device (1in to 2ft) To acquire coordinates in Matlab we used an open source Matlab interface for the Leap Motion controller written by Jeff Perry (https://github.com/jeffsp/matleap).

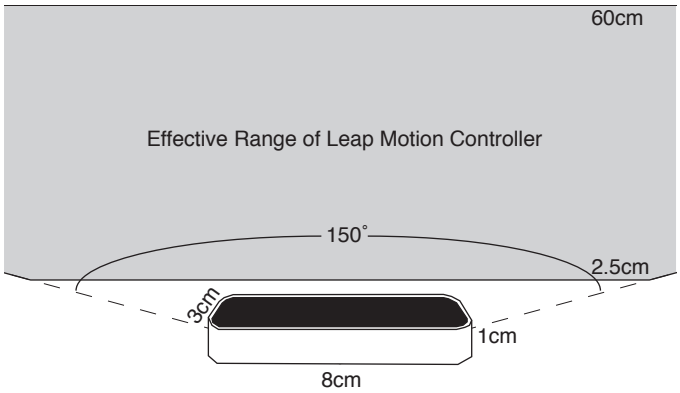

Figure 3.18: Leap Motion Controller. At the bottom of the figure is a line drawing of a Leap Motion controller. The controller is 8cm x 1cm x 3cm. It's effective range (colored in gray) is a conical frustrum above the controller.

We conducted two experiments to establish the precision of the Leap Motion controller in the context of our task. The first experiment evaluated the spatial precision of the Leap Motion controller. The second experiment measured both the temporal accuracy (lag) and precision.

Spatial precision of Leap Motion controller

Methods. The apparatus was the same as in the original experiment (see General Methods). Two subjects (subject 1 and subject 2 from above) were asked to point their pointer finger and remain stationary above the Leap Motion controller for

5 seconds. The same process was repeated for a wooden dowel. The dowel is recognized as a 'tool' and was fixed at typical finger height above the leap.

Results. Figure 3.19 shows x-y-z position over time (blue, red, and black respectively) for the index fingers of two subjects and the fixed wooden dowel. The mean x-y-z drift in millimeters for the S1, S2 and the wooden dowel was $(0.443, 0.445, 0.439)$, $(-0.128, 0.084, -0.023)$, and $(0.005, -0.006, 0.010)$ respectively, while standard error (also mm) was $(0.015, 0.016, 0.013)$, $(0.008, 0.015, 0.004)$, and $(0.0002, 0.0008, 0.0010)$. As expected, the dowel was considerably more stable than the human subjects, demonstrating the the any noise or drift in the Leap Motion controller itself is well below the level of motor noise exhibited by human subjects.

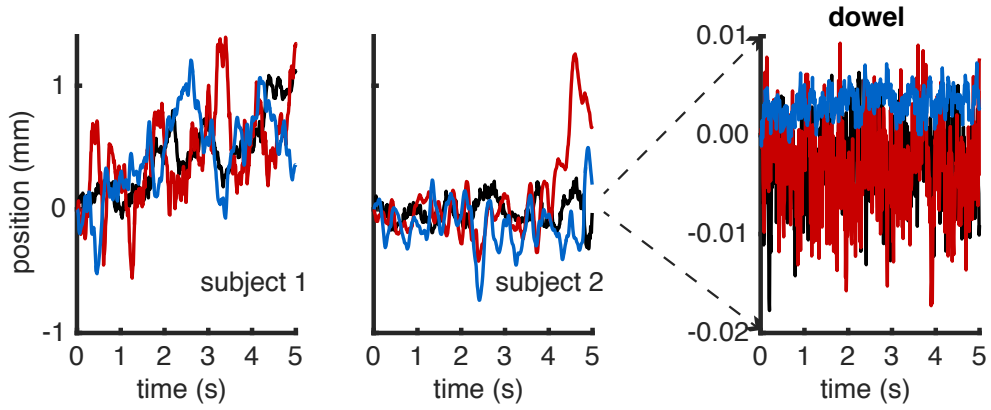

Figure 3.19: Measurement of drift of stationary fingers and a fixed wooden dowel. Each panel shows the x (blue), y (red), and z (black) drift in position over time during 5 second period in which either the subject (S1, S2) was instructed to remain stationary or the dowel was fixed above the Leap Motion controller. Clearly, the intrinsic spatial noise level of the Leap Motion controller is much smaller than the steadiness of the observers' hands.

Temporal lag and precision of Leap Motion controller

Apparatus. A schematic of the setup is shown on the left panel of Figure 3.20. The basic apparatus was the same as in the original experiment (see General Methods). Two photocells (VDT Sensors Inc., Hawthorn, CA) were used. The first photocell was placed against the lower of the two Planar monitors. The second photocell was placed opposite a beam of light generated by a laser pointer (green). The Leap

controller was placed underneath the beam. Subjects were given the occluder (small flat piece of plastic attached to a ring) to wear on their pointer finger in order to block the beam of light. Both photocells were connected to an oscilloscope (Agilent DSO-X 2014A; Agilent Technologies; Santa Clara, CA).

Procedure. Just before each trial, subjects arranged their hand so that the occluder was about to block the beam of light. Once a trial began, any position farther than a threshold of 2 mm forward from the initial position triggered the screen to flip from black to white. The subjects would then move their hand forward, blocking the light beam. The oscilloscope was then used to measure the difference between the onset of motion in physical space (or when the beam of light was blocked) vs the onset of motion on the screen (or when the screen flipped from black to white). Oscilloscope traces from a sample trial are shown on the left side of Figure 3.20 with the beam photocell in blue and the screen photocell in red. The measurement taken each trial was the difference between the location in time (s) of the step down of the screen photocell and the step up of the beam photocell. Two subjects (S1 and S2 from before) performed 10 trials for each device. When S1 was using the Leap Motion controller as a subject, S2 was the experimenter, taking measurements from the oscilloscope and vice versa. For comparison we used exactly the same procedure to evaluate a bluetooth trackpad (Apple Magic Wireless Trackpad) and a more standard USB mouse (Logitech).

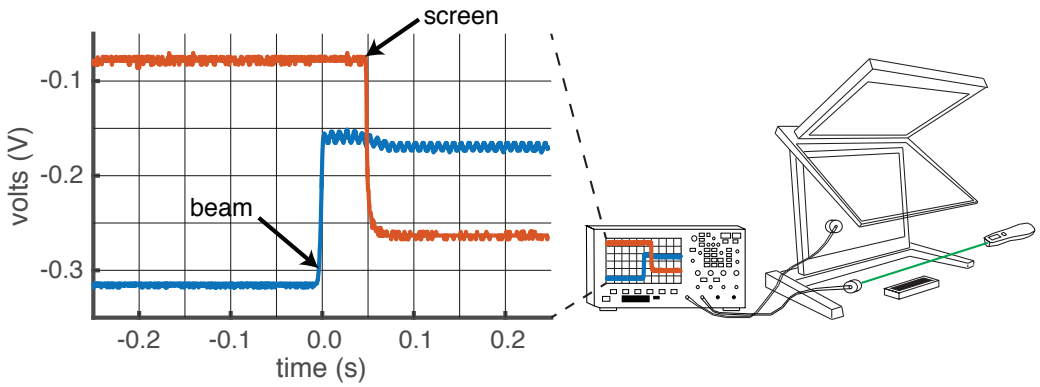

Figure 3.20: Schematic of photocell arrangement and oscilloscope readings. (Right) The first photocell was placed on the lower Planar monitor. The second was placed to one side of the Leap Motion controller with a beam from a laser pointer pointed directly at the collector. A forward hand movement broke the laser beam and, via the Leap, also triggered the software to flip the screen from black to white. (Left) Sample oscilloscope output from a single trial. The oscilloscope reports the voltage over time from the screen photocell (red) and the beam photocell (blue). When the subject moves their finger forward, the occluder worn on the subject's finger blocks the laser pointer. This causes the blue trace to step up and the movement triggers the code to change the screen from black to white causing the red trace to step down. The time difference between these two steps is the measurement of interest.

Results. Figure 3.21 shows the results for all three devices. Results were consistent across both subjects. The USB mouse was the fastest from motion to screen change at 31ms and 34ms for S1 and S2 respectively, followed by the Leap Motion controller at 56ms and 66ms, and finally the bluetooth trackpad at 102ms and 92ms. Although the Leap was not the fastest input device, it was clearly within the latency range of common input devices.

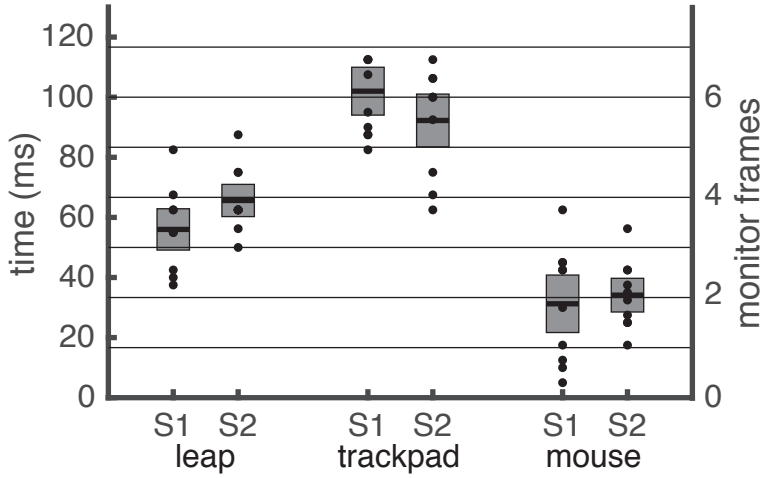

Figure 3.21: Lag and precision of Leap Motion controller, bluetooth trackpad and USB mouse. We measured the lag and temporal precision of 3 devices: Leap, trackpad and mouse for two subjects (S1 and S2). The above plot shows that temporal lag (milliseconds on the left, frames on the right) for the mean (horizontal line) as well as each trial (black dot). There are ten trials per condition per subject but some data points are overlapping. The mean is denoted by the thick black horizontal line and the standard deviation by the gray box.

Leap Motion controller refresh rate

Leap Motion, Inc reports that the device has a refresh rate of 115Hz. Each sample collected from the leap has a unique ID, so this can be tested. We wrote a Matlab script that sampled from the Leap controller counting the unique frames. We ran this script 10 times for 5 seconds each. The Leap Motion controller's update rate was 114Hz in each of these tests.

3.10 Appendix B: Statistical Tests

subject	1	2	3
horizontal	.82	.83	.88
vertical	.79	.83	.83
depth	.83	.73	.84

Table 3.1: The proportion of variance explained by the fits shown in Figure 3.2.

subject	1	2	3
Lag	2.78, p<1e-5	3.46, p<1e-5	3.10, p<1e-5
Peak	1.83, p<1e-5	1.75, p<1e-5	1.09, p<1e-5
Width	2.63, p<1e-5	0.84, p<1e-5	0.66, p<1e-5

Table 3.2: Comparison of frontoparallel motion tracking (horizontal, blue in figures 1, 2, & 3) and motion-through-depth tracking (black in figures 1, 2, & 3) in Experiment 1. Summary of the effect sizes and significance values for the difference of medians.

subject	1	2	3
Lag	.001, R^2=.03	-.025, R^2=.55	-.015, R^2=.54
Peak	.058, R^2=.94	.057, R^2=.99	.060, R^2=.95
Width	-.006, R^2=.05	-.050, R^2=.86	-.016, R^2=.56

Table 3.3: Linear fits of lag, peak and width for changing amplitude in Experiment II. Summary of the slope and R^2.

subject	1	2	3
Lag	1.84, p<1e-5	2.19, p<.1e-5	7.73, p<1e-5
Peak	1.77, p<1e-5	1.50, p<.1e-5	0.40, p = .07
Width	2.16, p=1e-5	0.39, p=.02	1.69, p<1e-5

Table 3.4: Comparison of motion-through-depth tracking and frontoparallel motion tracking at $\sigma = .51$ arcminutes in Experiment II. Summary of the effect sizes and significance values for the difference of medians.

subject	1	2	3
Lag	2.55, p<1e-5	0.55, p=1e-3	0.87, p<1e-5
Peak	0.08, p=0.40	0.58, p=.01	0.48, p=.09
Width	1.54, p<1e-5	0.23, p=0.15	0.22, p=.05

Table 3.5: Comparison of gain-corrected frontoparallel motion tracking and motion-through-depth tracking performance in Experiment III. Summary of the effect sizes and significance values for the difference of medians.

subject	1	2	3
Lag	.60, p<1e-5	2.45, p<1e-5	1.70, p<1e-5
Peak	0.89, p<1e-5	1.97, p<1e-5	0.52, p=.01
Width	1.02, p<1e-5	0.67, p<1e-5	0.22, p=.03

Table 3.6: Comparison of vertical tracking with XZ finger motion and motion-through-depth tracking performance in Experiment IV. Summary of the effect sizes and significance values for the difference of medians.

subject	1	2	3
Lag	0.29, p=.04	.10, p=.20	1.00, p=.002
Peak	1e-3, p=0.61	0.19, p=.38	0.21, p=.22
Width	0.15, p=.03	0.05, p=.21	0.84, p<1e-5

Table 3.7: Comparison of frontoparallel motion tracking and motion-through-depth tracking performance in Experiment V. Summary of the effect sizes and significance values for the difference of medians.

Chapter 4

Transcending the trial: Linking continuous behavior, ongoing neural activity, and the time course of natural stimuli

This review is in preparation. Authors: Huk, A. C., Bonnen, K., He, B.

Abstract

The vast majority of experiments examining perception and behavior are conducted using experimental paradigms which adhere to a rigid trial structure – each trial consists of a brief and discrete series of events, and is regarded as independent from all other trials. The assumptions underlying this structure ignore the reality that natural behavior is rarely discrete, brain activity follows multiple time courses which do not necessarily conform to the trial structure, and the natural environment has statistical structure and dynamics that exhibit long-range temporal correlation. Modern advances in statistical modeling and analysis offer tools that make it feasible for experiments to move beyond the rigid IID (independent and identically distributed) trial structure. Here we review literature that serves as evidence for the feasibility and advantages of moving beyond trial-based paradigms in order to understand the neural basis of perception and cognition. Furthermore, we propose a synthesis of these efforts, integrating the characterization of natural stimulus properties with measurements of continuous neural activity and behavioral outputs within the framework of sensory-cognitive-motor-loops. Such a framework provides a basis for the study of natural statistics, naturalistic tasks, and/or slow fluctuations in brain activity, which should provide starting points for important generalizations of analytical tools in neuroscience and subsequent progress in understanding the neural basis of perception and cognition.

4.1 Introduction

In neuroscience, our conception of experiments is invariably built from the notion of the trial – a brief and discrete series of events that allow the experimenter to select certain input parameters and then measure the resulting output of the nervous system. Each subsequent trial can be executed in conceptual isolation from the prior ones, typically with statistically-independent input parameters. The outputs of the nervous system occurring during or right after each of these trials are then subjected to classical analyses derived from well-established tools, such as signal detection theory. The reliance on conventional experimental paradigms and analyses reflects a preference for apparent simplicity and control for the experiment and analysis, over the ecological validity of the tasks and stimuli used to probe the brain. Here, we examine emerging approaches for quantitative neuroscience experiments that acknowledge that natural behavior is rarely discrete and that brain activity follows multiple time courses which do not necessarily obey experimenter-imposed trial structure. We conclude that the synthesis of such approaches has the potential to progress our understanding of neural computation and how neural activity supports perception and behavior without a loss of quantitative rigor.

In this article, we focus on two primary reasons to move beyond these conventional paradigms. First, consider the fact that your own reading of this paper has not involved a series of events that could be well described as brief, independent trials, but which still arises from coordinated patterns of sensory input, neural processing, mental functions, and motor behavior. Likewise, riding your bike to work, searching for a lost key, or deciding whether to continue reading this – from simple sensorimotor behaviors to the highest forms of metacognition – involve continuous chains of sensory-cognitive-motor loops of processing that continue over time frames longer and less well-defined than that of a conventional experimental trial. Here, we explain that analytical tools exist for characterizing these sequences of behaviors. We therefore argue that continuous sensory-cognitive-motor loops are not merely tractable, but should be thought of as the most appropriate framework for studying many forms of behavior, perception, and cognition that are currently either shoehorned into trials (or not studied due to the difficulty in doing so).

The second main reason for moving beyond near-exclusive reliance on trial-based analyses is that they do not reflect the realistic structures and dynamics that

exist and occur in the environment and the brain. The statistics of the inputs that define the environment in which the brain evolved and normally functions do not necessarily follow the standard (e.g., statistically IID, Gaussian, et cetera) assumptions often made to simplify quantitative analyses. Instead, the sensory environment is typically broadband, with visual patterns exhibiting multiple spatial frequencies, and natural soundscapes exhibiting multiple temporal frequencies. The distribution of such spatial and temporal frequencies (i.e., its power) is often well approximated as $1/f^{\beta}$ (where β typically ranges from 0 to 2), implying that sensory systems typically exhibit a wide set of time scales and temporal dynamics, quite distinct from the usual single-frequency (or otherwise tightly restricted) nature of experimental inputs. Although the presentation of a single sinusoidal input allows for powerful and intuitive analyses derived from systems identification approaches, frequency-based analytic tools are capable of handling more complex inputs, and broad spectrums with both fast/brief and slow/long timescales are present both in natural stimuli and in recorded patterns of brain activity.

More broadly, we take this opportunity to explain an experimental and analytic framework that stands to make continuous behaviors and mental processes, ongoing brain activity, and natural statistical structure more tractable and more integrated. The value of this approach is not just ecological validity. By taking on the continuous, broadband, and generally more complex nature of both sensory inputs and neural activity, analytic tools actually become more powerful. They become capable of capturing elements typically left over as unexplained variance ("noise"), such as slow temporal fluctuations in neural activity that do not obey the faster timing of individual trials. Experiments also gain efficiency, placing the subject or nervous system in contexts with a higher effective duty cycle, with far less time spent in secondary phases, such as the dreaded intertrial interval. And perhaps most reassuringly, the tools for thinking this way not only already exist (and have been applied to neuroscience in certain situations), but are also quantitatively relatable to many more familiar analyses. We now lay out the case for considering this approach in more detail, and then synthesize a generic framework for analysis and a corresponding prescription for experimental design.

4.2 Moving from discrete to continuous paradigms in the study of sensorimotor transformations

Signal detection theory (SDT) is central to the study of the relationship between neural activity and cognitive function (Green & Swets, 1966). In the context of sensory systems, it posits that each sensory stimulus is represented in the form of a scalar "internal response", which reflects the intensity of the sensory stimulus, but which is perturbed by noise. This noise, often working in tandem with low stimulus intensities, places the internal response for a particular trial in an ambiguous regime: it is unclear how much of the internal response is driven by the stimulus, and how much is driven by noise on that particular trial.

As suggested by its name, signal detection theory is most straightforwardly applied to the challenge of detecting a weak signal. In such tasks, an observer (i.e., a human or a trained animal) is presented a stimulus, and their task is to indicate whether the stimulus was present or not. Most readers will be familiar with the comparison of each trial's internal response to an (also unobserved) internal criterion, as well as the four possible resulting outcomes (hits, misses, correct rejections, and false alarms). It is also common to apply this framework to tasks other than simple detection; for example, to the identification of a single stimulus (e.g., was a motion display of varying strength moving more to the left or to the right), and to two alternative forced choice tasks (e.g., which of the two moving patches was faster)? A variety of classic and more modern texts provide excellent primers for detailed and thorough mathematical treatments in many extended domains, although the majority of applications of SDT are of the simpler cases (Green & Swets, 1966; Nevin, 1969; Banks, 1970; Stanislaw & Todorov, 1999; Swets, 2014). What we will focus on here are the core assumptions when signal detection theory is related to neural activity.

In detection or simple identification/discrimination tasks, connecting SDT to measurements of neural activity is superficially straightforward. The internal response is presumably the appropriate neural activity, which is described as the noisy spike count during stimulus presentation. Thus, neurophysiological recordings can be thought of as providing direct access to the internal representations that are treated as unobserved variables in purely behavioral experiments, and which are usually estimated from analysis in such contexts.

One limitation of signal detection theory is that it is couched in discrete, singular terms. A single value of noisy internal response is compared against a criterion value. Note that in this instance, the internal response might be the spike count for the entire stimulus presentation, which therefore collapses any temporal dynamics within the trial. Although convenient and appropriate prima facie for tasks in which the sensory event is correspondingly brief and/or discrete, sensory decisions in the real world often take variable amounts of time to complete, and the sensory stimuli themselves can have noise that changes over time, and which therefore engender a process of evidence accumulation. It therefore seems worthwhile to consider extensions of signal detection theory to explicitly capture the temporal dimension.

The best-known temporal extension of signal detection theory is the drift-diffusion model (e.g., Wald 1947; Ratcliff 1978; Smith & Ratcliff 2004; J. Palmer et al. 2005; Ratcliff & McKoon 2008). Loosely, diffusion can be thought of as signal detection over time, in which each instantaneous temporal instance has a corresponding noisy internal representation. Over time, these repeated "pulls" from a signal-detection theory type mechanism are accumulated over time. When this accumulated evidence reaches a requisite level (the "decision bound") the decision is made; the actual behavioral response is then generated and, in the simplest case, assumed to be a relatively rapid process that is brief and stereotyped relative to the decision phase. The rate of this accumulation can depend on the strength of the sensory stimulus, and thus diffusion to bound can be formulated to make a prediction for both the accuracy and speed of decisions as a function of different stimulus conditions.

The value of accumulating evidence should be intuitive, but drift diffusion is only one way that the brain could benefit from evaluating evidence over time. Any mechanism that interrogates more than the initial impulse of a stimulus is capable of producing increases in accuracy with additional time. Classical drift diffusion is specified in continuous time and has no leak of the integration mechanism, but both discrete and leaky variants of accumulator models are often capable of fitting speed and accuracy data from psychophysical tasks (Usher & McClelland, 2001; Ditterich, 2006). Indeed, large bodies of literature have focused on distinguishing between these models, and substantial effort has been put into elaborating these models to include additional mechanisms such as competing accumulators (Smith & Vickers,

1988; Mazurek et al., 2003; Reddi et al., 2003), nonstationary models (Burbeck & Luce, 1982; Smith, 1995), and trial-to-trial parameter variability (Ratcliff, 1978; Ratcliff & Rouder, 1998, 2000; Ratcliff, 2002; Smith & Ratcliff, 2004). It has recently been argued that the majority of such elaborated models are not falsifiable given standard tasks and data (Jones & Dzhafarov, 2014).

The ambiguities associated with testing between various formulations of decision making mechanisms comes in part from the relatively limited amount of data collected on each trial. While the drift-diffusion model acknowledges the noisy time-varying internal process involved in sensory processing, the matching experimental paradigm (i.e., some sort of forced choice task) still only results in a single discrete behavior at the end of that internal process. This has the advantage of producing data that are simple to analyze (i.e., whether the choice was accurate, and when the response was made), but standard paradigms that wait for the end of the trial to record discrete behavioral outputs are by construction not able to directly shed light upon the noisy time-varying internal process meant to be studied within these paradigms.

Logically, an alternative approach would be to measure a series of behavioral observations in response to a presented stimulus. This time series could then be used to better model and understand the noisy internal processes that underlie sensory information processing. This is unwieldy if one envisions extending standard tasks to include multiple intermediate reports, but is in fact straightforward if one instead steps outside forced-choice tasks. One such class of tasks that provides a time series of behavioral observations are tracking tasks (e.g. Baddeley et al. 2003; Burge et al. 2008b; Mulligan et al. 2013; Bonnen et al. 2015, 2017). In these tasks, subjects track targets with their eyes or by pointing with their finger. These tasks are more natural for subjects and generate a large amount of behavioral data in a relatively short period of time.

Behavior in such tasks can be modeled by simple dynamic linear systems (i.e. state-space models, see equations 4.2.1 - 4.2.2) and their solutions (e.g. Kalman filter, see equation 4.2.3):

$$x_t = F_t * x_{t-1} + w_t; \quad w_t \sim N(0, Q_t) \tag{4.2.1}$$

$$y_t = H_t * x_t + v_t; \quad v_t \sim N(0, R_t) \tag{4.2.2}$$

where x_t is the stimulus parameter tracked by the subject at time t, F_t is the process transition matrix, w_t is the process noise, y_t is the noisy internal response, H_t is the observation model that maps the true state space to the observation space, v_t is the internal noise. Here we assume Gaussian noise models for both process and internal noise (the former of which can be enforced in stimulus design). Under this assumption, the Kalman filter provides the Bayes-optimal estimator:

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t \left(y_t - H_t \hat{x}_{t|t-1} \right) \tag{4.2.3}$$

where $\hat{x}_{t|t}$ is an estimate of x_t. and K_t is the Kalman gain. Tracking tasks in conjunction with Kalman filter models form the basis for a more detailed study of the temporal dynamics of sensory processing. Typically, the Kalman filter is used to produce estimates of x_t, given the noisy measurements y_t. However, by flipping the estimation framework over to become a fitting problem, Bonnen et al. (2015) showed how the state-space model and Kalman filter solution can also be used to estimate the parameters associated with the noisy internal response, given the stimulus parameter (i.e. the true state x_t) and the behavioral response (i.e. the estimates, $\hat{x}_{t|t}$).

There are significant similarities between the state-space model of tracking and the drift-diffusion model of forced choice tasks. Notice that the noisy internal response is also a component of the Kalman filter model; the internal noise is part of the noisy measurement of some underlying state of the world. In drift diffusion the internal noise affects the accumulated evidence and is related to the behavioral outcome when the process hits the decision bound; In the Kalman filter framework, the Kalman filter solution gives an equation for relating the behavioral estimates to the noisy internal response over time. The advantage of the Kalman framework is that the time series of noisy internal responses is not related to a single behavioral outcome but rather to series of behavioral estimates.

Here we have laid out the math for a linear state-space model with Gaussian noise and its Kalman filter solution – these are the standard assumptions for the Kalman filter but it is worth noting that a variety of extensions of the original Kalman filter exist which extend the solution to nonlinear and non-Gaussian state-space models (Sorenson, 1985; Uhlmann, 1992; Julier & Uhlmann, 1997). Work across a range of subfields has used the Kalman filter to model neural data. The

brain-machine interface community uses Kalman filters to perform decoding on neural activity to control cursors, robotic arms, etc (e.g. Carmena et al. 2003; Wu et al. 2004). There is behavioral and theoretical evidence for the existence of the nervous system approximating the function of Kalman filters in implementing state-space estimators (Denève et al., 2007; Makin et al., 2015). The application we propose here is distinct in being more pragmatic and more general, putting forth that the Kalman filter (and other filtering solutions to state-space models) can be used as framework for relating behavior and neural activity in much the same way that neuroscience has leveraged signal detection theory. Throughout this section we have laid out the advantages to such a framework, in particular that it models sensory information processing as a noisy time-varying process and furthermore that the behavioral observations collected in such a framework would be collected at a temporal resolution more aligned with the underlying neural dynamics.

The development of continuous paradigms for relating behavior and neural activity is further motivated by a number of current revisitations of the time course of neural activity and its correlation with behavior in the context of trial-based paradigms (Lundqvist et al., 2016; Bolkan et al., 2017; Schmitt et al., 2017; Churchland et al., 2010; Goris et al., 2014; Yates et al., 2017). Many classic lines of work have been re-opened to reveal that even the simplest forms of temporal computations linking sensory and motor stages remain unclear. For example, current debate surrounds whether persistent neural activity actually exists at the single trial level during oculomotor working memory tasks and memory-guided saccade tasks or whether more transient bursts with variable times are the substrate of oculomotor working memory (Lundqvist et al., 2016; Bolkan et al., 2017; Schmitt et al., 2017). It is also contentious as to whether ramping activity during simple perceptual decisions is a straightforward neural correlate of evidence accumulation or whether it reflects alternate dynamics, a mixture of simpler factors, and/or secondary signals not functionally necessary for performing the task (Gold & Shadlen, 2007; Freedman & Assad, 2016; Huk et al., 2017).

Finally, the development of continuous paradigms is motivate by the need to move beyond the stimulus and analysis constraints present in conventional trial-based paradigms. A variety of phenomena have accumulated that are poorly understood and not well-integrated, but which are all related to long time-scale fluctuations in natural stimuli, cortical dynamics, and perceptual behavior. Most of

these do not fit nicely into assumptions of short time scales and temporal statistical independence. We believe that the development of continuous paradigms relating behavior and neural activity will be critical to continued progress in understanding such phenomena. The following section examines these phenomena in greater detail.

4.3 Removing the IID assumption: Natural stimuli, ongoing brain activity, and serial dependencies in perception and behavior

A repeated finding using trial-based paradigms is that human perception and behavior exhibit long-range temporal correlation, manifesting as trial-to-trial correlation in the perceptual judgments or behavioral outputs. For example, in simple tasks such as threshold-level detection or reaction time tasks, as well as reproduction of a particular level of force or a particular time interval, the trial-to-trial fluctuations of hit rate, reaction time, force output, and time-interval output exhibit long-range temporal correlation such that their power spectra follow a $P \propto 1/f^{\beta}$ form, where P is power, f is temporal frequency, and β is a scaling parameter typically between 0 and 1 (Gilden et al., 1995; Gilden, 2001; Monto et al., 2008). Such a $1/f$-type power spectrum indicates that performance many trials ago is still correlated with that in the current trial, with the magnitude of this relation falling off with increasing time interval. Interestingly, long-range temporal correlation in reaction time fluctuations are modulated by task difficulty (Clayton & Frey, 1997) in a manner similar to task modulations of long-range temporal correlations in neural activity (He et al., 2010), suggesting that slow fluctuations in neural activity may underlie trial-to-trial behavioral dependence – a point we elaborate on below.

Although long-range temporal correlation in human behavioral output has been long described, (positive) trial-to-trial serial dependence in human perception was only recently discovered (Chopin & Mamassian, 2012; Fischer & Whitney, 2014; Liberman et al., 2014), and remains controversial (Maus et al., 2013; Fritsche et al., 2017). Presumably, this is due to the fact that perception is also strongly influenced by adaptation – negative trial-to-trial correlation which may cancel out positive serial dependence, resulting in the net effect varying across experiments depending on the exact paradigm and subject population. Nonetheless, there is

now strong evidence suggesting the existence of positive trial-to-trial correlations in both perceptual and behavioral outcomes.

What is the neural basis of trial-to-trial serial dependence in perception and behavior? A recent study found that in an orientation judgment task, orientation signals in V1 measured by fMRI were positively correlated from trial to trial, similar to the perceptual decisions made by subjects; in addition, both the behavioral and neural serial dependence were spatially specific (St John-Saaltink et al., 2016). More broadly, ongoing brain activity at the level of population signals recorded by local field potentials (LFP) (Manning et al., 2009; Milstein et al., 2009), electrocorticography (ECoG) (K. J. Miller et al., 2009; He et al., 2010), MEG/EEG (Dehghani et al., 2010; Lin et al., 2016), and fMRI (Bullmore et al., 2001; He, 2011) exhibit long-range temporal correlations, manifesting as power spectra following a $P \propto 1/f^{\beta}$ form, with β typically between 0 and 2 (He, 2014). This long-range temporal correlation in neural activity extends to at least the time scale of several minutes (with the corresponding $1/f$-type power spectrum extending down to below 0.005 Hz; He 2011; Lin et al. 2016), and is thus well positioned to produce trial-to-trial correlations in ongoing neural activity with standard trial-based behavioral paradigms. Thus, long-range temporal correlation in neural activity is a natural cause for serial dependence in perception and behavior.

A now-extensive literature describes the rich network structures embedded in spontaneous fMRI signals (for reviews see Buckner et al. 2013, Petersen & Sporns 2015, and Raichle 2015). Spontaneous fMRI signals, which have a frequency range of < 0.5 Hz, correlate with the low-frequency (< 5 Hz) component of neural field potentials, named "slow cortical potentials" (SCPs) (He et al., 2008; Pan et al., 2013). Like the spontaneous fMRI signals, ongoing fluctuations in the SCPs are also coherent within intrinsic large-scale brain networks (He et al., 2008). Both types of signals contain very slow fluctuations in the order of seconds to minutes, and exhibit long-range temporal correlations (He et al., 2010; He, 2011) that are well poised to drive trial-to-trial serial dependence in perception and behavior. Consistent with this idea, studies have demonstrated that pre-stimulus spontaneous fMRI and SCP activity influence threshold-level perception and behavioral output (Boly et al., 2007; Fox et al., 2007; Hesselmann et al., 2008; Monto et al., 2008; Li et al., 2014; Baria et al., 2017). However, much work remains to be done to directly probe the connection between slow fluctuations in fMRI signals and SCPs, with time scales extending far

beyond a typical trial, and trial-to-trial correlations in perception and behavior.

Along a separate, but likely related vein, it is well known that many natural stimuli exhibit temporal or spatial power spectra following a $P \propto 1/f^{\beta}$ form, with β commonly ranging between 0 and 2. In the visual domain, natural movies typically follow a $P(f) \propto 1/f^{\beta}$ type temporal power spectrum (Dong & Atick, 1995). In the auditory domain, loudness and pitch fluctuations of natural soundscapes, such as urban and rural environmental noise (De Coensel et al., 2003), speech and music (Voss & Clarke, 1975), also exhibit 1/f-type temporal power spectra. Thus, the temporal dynamics of natural stimuli exhibit long-range temporal correlation, in a manner similar to trial-to-trial fluctuations of human behavioral output as well as slow, ongoing neural activity recorded by fMRI or SCP.

Might there be a relationship between these three phenomena: statistical structures of natural stimuli, trial-to-trial correlations in perception and behavior, and long-range temporal correlations in ongoing neural activity? As mentioned earlier, slow fluctuations in ongoing neural activity are well positioned to contribute to serial dependence in perception and behavior. However, the other link -- between neural activity and perception/behavior on the one hand and natural stimuli on the other hand – has proven more elusive. This is partly because natural stimuli are less analytically tractable than simpler, artificial stimuli with narrower temporal / spatial frequency bandwidth or the sorts of Gaussian and/or IID assumptions often made in trial-based frameworks. However, tools for analyzing neural activity in response to natural stimuli are developing quickly, such as assessing similarity in evolving neural dynamics between repeated presentations of the same temporally-extended natural stimulus (Hasson et al., 2010), and encoding models relating multiple stimulus parameters to neural activity at each time point (Naselaris et al., 2011). In addition, mathematically constructed artificial stimuli that capture the 2nd-order statistical structures (i.e., power spectrum, autocorrelation) of natural stimuli but are nonetheless precisely controlled have proven to be a powerful tool for probing how the nervous system processes statistical structures present in natural stimuli (e.g., S. E. Palmer et al. 2015; Lin et al. 2016). For instance, long-range temporal correlations exhibited by MEG activity recorded from humans not only reflect long-range temporal correlations in stimulus input but also predict individual subject's ability to discriminate different levels of temporal correlations in the stimulus input (Lin et al., 2016).

Lastly but certainly not least, being able to make valid predictions about environmental stimuli confers an obvious evolutionary advantage. So far, studies on predictive processing based on statistical regularities in stimulus input have typically adopted simple, artificial stimuli that involve repeated presentation of items or sequences (e.g., Bekinschtein et al. 2009; Yaron et al. 2012; Gavornik & Bear 2014). And most trial-based frameworks enforce that there is nothing predictable about the next trial based on the preceding ones. Yet, the long-range temporal correlations prevalent in natural stimuli suggest that natural stimuli have a substantial degree of predictability, and it seems plausible that the nervous system has evolved to capitalize on such dependencies to make predictions about its environment in order to best react to it or act upon it. Thus, a key question for future studies is how predictive processing based on natural statistical structures is implemented in the brain.

Importantly, tools for addressing these questions in both stimulus design and data analysis already exist. As mentioned earlier, temporally varying natural stimuli often exhibit temporal power spectra following a $P \propto 1/f^{\beta}$ form, with β typically ranging between 0 and 2. This 2nd-order statistical structure (i.e., power spectrum, or its closely related auto-correlation function) is what confers temporal redundancy or predictability for the continuous natural stimuli. When β is in the range of [0, 1] (in fact, anywhere between -1 and 1), the corresponding time-domain stimulus input forms a stationary sequence (technically referred to as "fractional Gaussian noise" or fGn). When such a sequence has zero-mean (the mean can be added back after estimation), the mathematically optimal linear prediction of order K for the upcoming item in the sequence $\hat{x}_n$ based on past samples $(x_{n-1}, x_{n-2}, ..., x_{n-k})$ is written as:

$$\hat{x}_n = \sum_{k=1}^{K} a_k x_{n-k}, \tag{4.3.1}$$

where the vector $a = (a_1, a_2, ..., a_k)$ is to be chosen (or estimated) so as to minimize the average squared prediction error. Linear algebra leads to an explicit theoretical solution for $\underline{\hat{a}}$ (Scharf & Demeure, 1991):

$$\underline{\hat{a}}^{theory} = \underline{\underline{R}}_K^{-1} r_x \tag{4.3.2}$$

where r_x is the covariance sequence of process x, and $\underline{\underline{R}}_K$ denotes the KxK square matrix, with entry $(\underline{\underline{R_K}})_{p,p'} = r_x(|p - p'|)$ for $p, p' \in \{1, ..., K\}^2$.

When β is in the range (1,2) (in fact, anywhere between 1 and 3), the corresponding time-domain stimulus input forms a nonstationary sequence (technically referred to as "fractional Brownian motion" or fBm). fBm sequences are cumulative sums of their corresponding fGn sequences, whose β exponents differ by 2. The mathematically optimal prediciton for an upcoming item in an fBm sequence can be estimated by the sum of the current item in the fBm sequence and the optimal prediction for the upcoming item in the corresponding fGn sequence.

Together, these tools allow the mathematical calculation of the optimal prediction for the upcoming stimulus input given the past history of any stimulus sequence exhibiting a $P \propto 1/f^{\beta}$-type power spectrum (where $\beta \in [-1, 3]$).

Employing this mathematical framework, a recent study created a set of stimulus sequences exhibiting $1/f^{\beta}$-temporal power spectra, where $\beta \in [0, 2]$. To dissociate sensory processing of the current stimulus input from predictive processing of the upcoming stimulus input, these sequences converged onto the same value for the penultimate item, while their different history prescribed different values for the optimally predicted upcoming item. The actually presented last item was randomly drawn from a fixed distribution and subjects gave surprise ratings for this last item based on the previous stimulus history. Using psychophysics, the authors established that human subjects can indeed capitalize on these natural statistical structures to make valid predictions (Maniscalco et al., 2018). In addition, concurrent MEG recordings revealed that slow, arrhythmic activity in the SCP range reflected integration of stimulus sequence history over time, and that such history integration contained in slow neural activity predicted the mathematically expected value of the upcoming stimulus input, providing a concrete computational mechanism, implemented in the human brain, for forming predictions based on natural statistical structures.

4.4 Conclusion: Moving to naturalistic and continuous stimuli, behavior, and neural measurements without a loss of quantitative tractability

The time seems ripe to loosen current adherence on trial-based paradigms for understanding the neural basis of perception and cognition. In this review, we have discussed: (1) analytic frameworks that are amenable to continuous input-output relations, while still being relatable to the signal-detection framework; (2) the existence of behavioral and neural responses that do not conform to the time scales of individual trials, and thus violate trial-based independence assumptions; and (3) that the statistics of natural stimuli also span timescales distinct from trials, and thus the nervous system typically functions with inputs and goals that are not comprehensively probed with standard experimental paradigms.

Here, we propose a synthesis of these sorts of efforts into a generic framework that characterizes the broadband properties of stimuli (as opposed to attempting to simplify these properties), and which measures continuous neural activity and behavioral outputs (instead of summarizing neural activity with simple statistics and/or considering binary behavioral outputs). With these philosophies of stimulus-task design and measurement in place, the analytic framework appears within reach. We conclude by identifying three key areas for continued development. First, the Kalman filter framework initially proposed by Bonnen et al. (2015), implements the simplest proof-of-concept assuming the stimulus is a random walk composed of Gaussian noise. A variety of extensions of the original Kalman filter exist, extending the solution to nonlinear and non-Gaussian state-space models (Sorenson, 1985; Uhlmann, 1992; Julier & Uhlmann, 1997). Future work will need to both identify appropriate stimuli and the corresponding Kalman filter solutions. Second, tools for analyzing temporally-continuous neural data will need to be adapted. There are several promising instances of such tools, from the generalized linear model (GLM) framework used to characterize spike trains, to the frequency-based tools used for field potential recordings. However, linking these tools together and to behavior will better integrate this endeavor. Third, the loop needs to be closed, with the aforementioned analytic developments pointing to a class of broadband and/or continuous stimuli and behavioral measures that are not just appropriate

but maximally efficient and/or insightful.

In summary, the convenience of chopping inputs and outputs up into trials makes a lot of sense for allowing straightforward analyses of both brain function and behavior. However, at this time, enough pressure has accumulated to suggest that strict adherence to the trial is fated to providing only a partial and somewhat artificial understanding of how intelligent actions are generated by the brain. Somehow, the brain grapples with slow internal fluctuations in its own activity, slow external fluctuations in sensory stimuli, and the need to not just respond continuously, but to do so in a predictive manner. The analysis of natural statistics, naturalistic tasks, and/or slow fluctuations in brain activity should no longer be seen as niche enterprises, but rather as the starting points for important generalizations of our tools and subsequent understanding.

Chapter 5

Neural coding of 3D motion

This work is in preparation. Authors: Bonnen, K., Czuba, T., Kohn, A., Cormack, L. K., Huk, A. C.

Patterns of neural activity represent information in the brain, but interpreting the precise meaning of such patterns (the neural code) remains a challenge. Inquiry into neural coding typically involves characterizing the patterns of neural activity due to particular stimuli (encoding) and/or estimating a stimulus from a pattern of neural activity (decoding). This approach has been successful in studying sensory systems where sensory features are often related to neural activity via canonical bell-shaped (Gaussian) tuning functions. Despite its successes, the encoding-decoding framework often overlooks transformations that occur between the physical environment and the signal introduced to the nervous system via sensory transduction. For the visual system, this transformation is the projection of a dynamic 3D environment onto the 2D retina in each eye. Here we show that this environment-to-retina transformation fundamentally reshapes the neural representation of sensory information. The resulting mapping generally does not give rise to canonical Gaussian tuning functions thought to be the fundamental form for neural tuning to visual features. The tuning that *does* arise from incorporating the environment-to-retina transformation explains overlooked (but substantial) anomalies in both neurophysiological recordings of spatiotemporal sensitivity in primate visual cortical neurons, and in psychophysical studies that have reported strikingly non-veridical perception of motion-through-depth. Furthermore, decoding analyses reveal that the encoding of 3D motion direction information in MT relies on relatively small differences in neural tuning (e.g. differences in speed preference, tuning bandwidth, and ocular dominance) to visual input from the left eye versus the right eye. Our findings highlight that non-intuitive insights come from extending work on neural coding in ways that recognize the nervous system's ultimate goal of inferring the properties of the environment.

5.1 Encoding

A major goal of the nervous system is to infer the properties of the physical environment given patterns of neural activity. To do this, patterns of neural activity must carry information about the properties of the physical environment. For the primate visual system, the challenge is to extract and represent the structure of the environment over time from the dynamic two-dimensional patterns of light that fall upon the two retinae. Much of the work examining this process in the visual system has focused on monocular frontoparallel stimuli and their properties (e.g. orientation, 2D motion direction, 2D speed: Hubel & Wiesel 1959; Adelson & Bergen 1985; Albrecht & Geisler 1991; Albright 1984; Maunsell & Van Essen 1983b; Newsome & Pare 1988; Simoncelli & Heeger 1998; Rust et al. 2006). While this has led to enormous progress in understanding the extraction and representation of spatiotemporal structure in the primate visual system, these simplified cases are fundamentally studies of retinal stimulation, rather than studies of the representation of a dynamic 3D environment. As a result, even when binocular information is considered (e.g. binocular disparity), generalization to a principled understanding of how the primate visual system represents the dynamic 3D environment is missing from the current understanding.

Here we build a theoretical framework to explain the encoding of environmental spatiotemporal structure, extending current approaches which implictly treat the stimulus as a direct proxy for the pattern of stimulation upon the sensory receptor surface (in this case, the retinae). As a simple case, we focus on motion direction within the x-z plane (i.e., towards/away/right/left, which we will refer to as 3D motion). For the better-studied case of frontoparallel motion (i.e., up/down/left/right; x-y), neurons in many brain areas (including MT, the middle temporal area) have approximately Gaussian-shaped tuning for frontoparallel direction and (log-)Gaussian tuning for speed (Maunsell & Van Essen, 1983b; Paradiso, 1988; Nover et al., 2005). These are canonical tuning curve shapes seen in many other domains, and analyses of this encoding benefit from the fact that stimuli containing purely x-y motion project isomorphic patterns of stimulation upon both retinae (simple inverted images).

But to relate these canonical tuning forms to the representation of information about the actual physical environment, requires building an encoding model that maps the dynamic 3D environment to the binocular retinal stimulation, and

only then applies the known tuning forms to the resulting retinal projections. The tuning curves for 3D motion direction that result from this environmental (as opposed to retinal) encoding framework are strikingly non-canonical (i.e. distinctly non-Gaussian, see Figure 5.1). The implication is that neural responses, when interpreted with respect to the 3D environment, cannot be conceived of with existing Gaussian idealizations. To test the validity of incorporating the environment-to-retina transformation into neural representations, we tested the ability of this framework to explain MT tuning curves measured in response to different 3D directions (Figure 1a). Recent electrophysiological work identified that MT neurons exhibit selectivity for 3D motion, but the exact form of the tuning had not been examined (Sanada & DeAngelis, 2014; Czuba et al., 2014).

The proposed encoding model for 3D motion direction in MT neurons combines the projective geometry of 3D motion onto the retinae and the known monocular tuning of MT neurons to 2D velocities. This model is illustrated for a single model neuron in Figure 5.1b-c. When a particular 3D motion direction is presented, geometric projection renders it into (often distinct) left and right eye retinal velocities (Figure 5.1b). The model neuron's monocular velocity tuning curves are log-Gaussian and have similar tuning for the left and right eye (Figure 5.1c; Supplemental equations 5.4.1 and 5.4.2; Nover et al. 2005). Each eye's retinal velocity corresponds to a neural response on the monocular velocity tuning curve (the third panel of Figure 5.1b shows the monocular responses for the left and right eye as a function of 3D motion direction). The resulting predicted binocular response to 3D motion is taken simply from a linear combination of the monocular responses to the corresponding left and right eye retinal velocities (see Figure 5.1b, fourth panel; Supplemental equation 5.4.5).

Thus this model incorporates straightforward geometric projection from the environment to the two retinae, uses standard processing by the gaussian velocity tuning known to exist for frontoparallel stimulation, followed by simple linear combination of the monocular responses to the left and right eye retinal velocities. Despite the simplicity of these stages, the incorporation of projection geometry gives rise to a radically atypical tuning curve, characterized by distinct plateaus, separated by steep cliffs. This violates the empirical norm of bell-shaped tuning across virtually all sensory features and systems; from visual orientation in cat primary visual cortex to wind velocity in cricket cercal cells, neuronal tuning is almost always bell-shaped

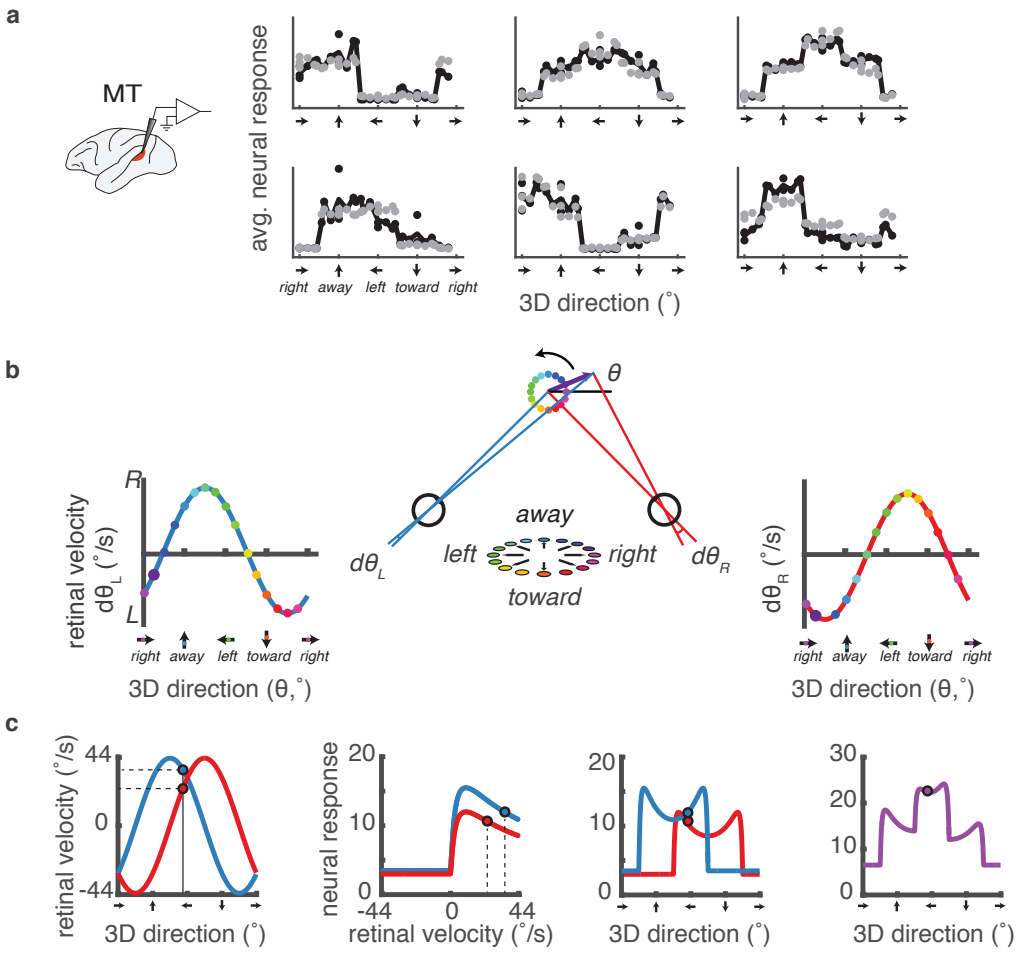

Figure 5.1: **A model that combines the geometry of 3D motion projected onto the retina with the monocular responses to retinal velocities accounts for the strange shapes of binocular 3D motion tuning curves in macaque Middle Temporal area. a.** The electrophysiological results of Czuba et al. (2014) compared to the predictions of our model. Each panel depicts the average response of a single neuron to the presentation of different 3D motion directions (black dots, solid line). The stimuli consisted of either binocular (fully crossed manipulation of retinal velocities in the two eyes: $-10°/s$, -2°/s, -1°/s, 1°/s, 2°/s, 10°/s) or monocular drifting gratings. Each stimulus was repeated 25 times. Across two anesthetized macaques, a total of 236 cells were recorded using extracellular tetrodes. Additional details can be found in the original paper (Czuba et al., 2014).
(continued on following page)

Figure 5.1: *(continued)* Each panel also depicts the prediction of the model given the neuron's monocular responses to the component drifting gratings (gray dots). Across neurons in this dataset, 47% of the variance in the binocular responses is explained by the linear combination of the corresponding monocular responses (using 2-fold cross-validation); 72% of neurons have a correlation value $r > .5$. **b.** Diagram of the projection of 3D motion (confined to the xz-plane; middle panel) onto the left eye (blue; left panel) and the right eye (red; right panel). For 3D motion confined the xz-plane, any resulting retinal velocity is moving left/right. For example, motion to the left results in equal rightward retinal velocities. Motion toward the observer results in retinal velocities of equal magnitude but in the opposite directions (leftward velocity in the left eye and rightward in the right eye). If the 3D motion happens to be directly along the line of sight for an eye, there is no motion in that eye. Retinal velocity as a function of 3D motion direction results in a sinusoid for each of the two eyes, which are phase-shifted due to the offset between the two eyes. **c.** A simple geometric model of 3D motion direction tuning. The input to the model is 3D motion direction which is first transformed to a left and right eye retinal velocities (Panel 1). Then monocular neural responses to these retinal velocity tuning curves are calculated from the monocular retinal velocity tuning curves (Panel 2). These two functions are then composed to give the monocular neural responses as a function of 3D motion direction (Panel 3). The prediction for the binocular response to 3D motion direction is given by a linear combination of the monocular responses, resulting in the binocular 3D motion direction tuning curve in Panel 4. The points called on the individual panels represent this transformation for a single motion direction.

(Hubel & Wiesel, 1959; Bacon & Murphey, 1984; Jacobs & Theunissen, 1996). However, a closer examination of empirical tuning curves for 3D motion in MT neurons (collected by Czuba et al. 2014) reveals tuning similar to that predicted by the geometric model. Figure 5.1a depicts the empirical tuning curves (the average neural response plotted as a function of 3D motion direction) for several example neurons (black dots; see Figure caption 5.1a for additional details). The model provided a good description of the shape of MT 3D tuning measured by Czuba et al. 2014 (47% of the variance in the binocular responses was explained by the linear combination of the corresponding monocular responses (using 2-fold cross-validation).

5.2 Decoding

Having found evidence in favor of incorporating projective geometry into an encoding model of MT neural responses to 3D motion, we then tested whether this encoding could also explain perceptual phenomena. We first confirmed that a simulated population of such neurons was in theory capable of being decoded to accurately estimate 3D direction. We used log-Gaussian functions (fitted to the monocular responses measured in MT experiments), combination coefficients (learned by minimizing the squared error between the binocular responses to 3D motion and the linear combination of the outputs of the log-Gaussian functions), and assumed Poisson noise (see Supplemental equations 5.4.1, 5.4.2, and 5.4.5). We stimulated that realistic simulated population with motion directions around the xz-plane, sampled at 1 degree intervals, and employed a standard maximum log-likelihood decoder to estimate 3D motion direction from the resulting population response (e.g. Graf et al. 2011; see also Supplemental equations 5.4.6-5.4.7).

The decoder successfully estimated 3D motion direction (Figure 5.2b; estimates fall on or near the unity line). However, the radically different tuning structure for 3D motion direction resulting from the projective geometry does have interesting ramifications. Decoding performance varies as a function of the true motion direction (see Figure 5.2c). The standard deviation of model estimation error shrinks considerably (i.e. there is higher sensitivity) for certain motion directions. Note that these motion directions correspond with the locations of the steep transitions on individual neural tuning curves (see Figure 5.1c, last panel). The steep transitions occur at the 3D motion directions where the retinal velocity in one eye changes direction, climbing or falling down the steep side of the underlying log-Gaussian monocular velocity tuning curve. All neurons in the population exhibit steep transitions in their tuning curves at these same locations. This is because the steep transitions are yoked to the angle formed by the locations of the eyes and the moving object (it is thus independent of fixation). Decoding along these steep parts of the tuning curves supports correspondingly higher sensitivity to changes in 3D motion direction. Those four locations along the x-z axis correspond to 3D directions whose retinal projection in one eye (or the other) changes direction. This pattern of results is very distinct from a decoder based on more canonical Gaussian tuning (see Figure 5.2f-h), which predicts consistent estimation error across all 3D

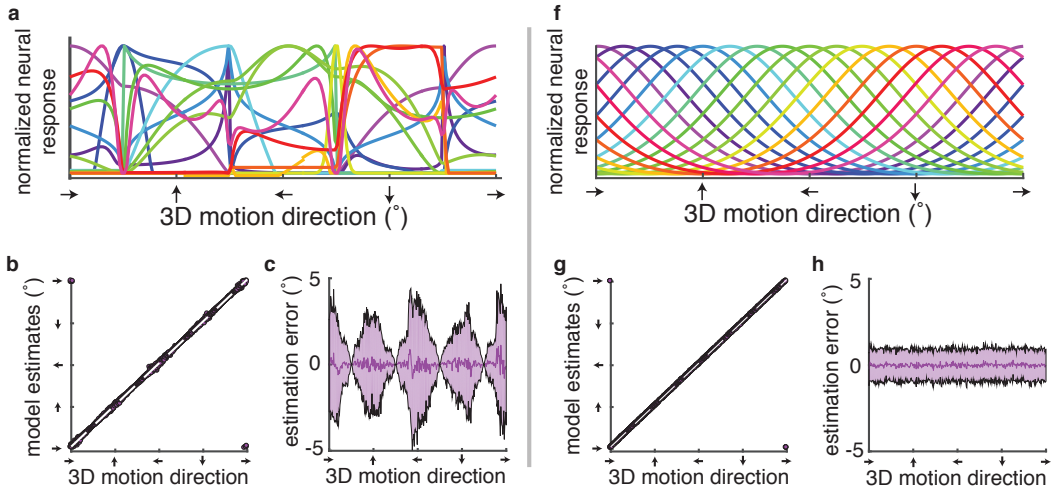

Figure 5.2: **A decoder based on the geometric model of 3D motion direction sensitivity can be used to estimate 3D motion direction and predicts a pattern of results that is distinct from the standard Gaussian model. a.** Binocular tuning curves from the geometric model for decoding 3D motion direction. These 16 example binocular tuning curves are taken from the geometric model population. Each was chosen because its preferred direction (as calculated by the vector average) was closest to one of 16 evenly spaced motion directions (0°, 22.5°, 45°, ... , 337.5°). **b.** Geometric model decoder successfully estimates 3D motion direction; estimates (dots) fall on the unity line (dashed white line). **c.** The standard deviation of estimates (purple cloud) varies cyclically as a function of the motion direction presented. **d** Hypothetical Gaussian tuning curves for decoding 3D motion direction. Here we show 16 evenly spaced Gaussian tuning curves (with preferred directions: 0°, 22.5°, 45°, ... , 337.5°). For the modeling based Gaussian tuning curves, 236 evenly spaced neurons were used. This matches the number of neurons in the population based on the geometric model. **e** Gaussian decoder successfully estimates 3D motion direction; estimates (purple dots) fall on the unity line (dashed white line). **f** The standard deviation of estimates (purple cloud) does *not* vary as a function of the motion direction presented.

motion directions.

The parameterization of the simulated population (see Supplemental eq. 5.4.1 - 5.4.5) also allows the examination of which aspects of neural tuning in MT neurons drive the successful estimation of 3D motion direction. A population with identical speed tuning parameters for input to the left and right eyes (i.e. the same speed preference, the same tuning bandwidth, the same response amplitude, and the same baseline firing rate) results in estimation performance that correctly identifies the x-component of the motion but, by virtue of having no differential binocular information, understandably chooses the wrong direction for the depth component half the time. However, merely incorporating differential tuning (at the levels measured in Czuba et al. 2014) for any of these parameters individually reveals that small (and seemingly trivial) differences in speed preference, tuning bandwidth, or response amplitude (i.e. ocular dominance) are each sufficient for estimating 3D motion direction without such depth sign errors. Previous work has suggested that differences in speed preference and tuning bandwidth could contribute to tuning for depth motion (Maunsell & Van Essen, 1983a); however the notion that the difference in response amplitude across the two eyes might contribute to tuning for depth motion was entirely unexpected and suggests that the brain may exploit ocular dominance in binocular computations of motion through depth.

Having established that the proposed encoding model supports the estimation of 3D motion direction, we incorporated a more realistic viewing distance (64 cm) to address whether decoding from this model could also explain puzzling aspects of perceptual behavior. As viewing distance increases, the angle formed by the eyes and the motion location is considerably smaller (e.g. Figure 5.3a). The retinal velocities at increased viewing distances have lower magnitudes and a decreased inter-ocular phase shift (compare Figure 5.3b at a 64 cm viewing distance to Figure 5.1b at 3.25cm viewing distance). At the farther viewing distance, the model makes systematic errors forming an X pattern of results in Figure 5.3d. These errors indicate that the model confuses the direction of the depth component, sometimes reporting a reflection across the x-axis from the true motion direction. There is however a simple explanation for this rather striking model error. Recall that as the viewing distance increases, the angle formed by the eyes and the motion location shrinks, resulting in a decreased phase shift in the retinal velocities between the two eyes. For a fixed environmental velocity, any single neuron's tuning curve is

dependent on those retinal velocities, and thus on viewing distance. For a larger (more realistic) viewing distance, the steep transitions of binocular tuning curves collapse towards each other (see Figure 5.3c). The result is a relatively symmetrical tuning curve, except for motion directions close to towards and away (see symmetry line Figure 5.3c). That symmetrical neural firing rate in the presence of noise, plus the coarse tiling of the neuronal tuning shapes (i.e., with the steep transitions geometrically yoked to the ocular axes), produces these model errors.

The geometric model decoder for 3D motion direction estimation provides a surprising set of predictions for perception. In particular, it is difficult to believe that human observers could confuse a motion having a substantial toward depth component with a motion having a substantial away depth component. However, there is existing psychophysical data consistent with this particular type of error (Landers & Cormack, 1997; Fulvio et al., 2015). Figure 5.3e, replotted from Fulvio et al. (2015) shows human performance in a 3D motion direction estimation task and demonstrates this relatively strange pattern of errors is actually observed in human perceptual behavior. Thus the predictions made by extending sensory encoding and decoding to incorporate the geometry of the spatiotemporal environment naturally account for what are at first glance strikingly odd aspects of neural tuning curves and human perception.

5.3 Conclusion

In summary, these successes emphasize the importance of recognizing the nervous system's ultimate goal of inferring the properties of the environment. Such inference is based on incoming sensory information that is fundamentally constrained by the geometric relationship between the environment and the initial sensors of the nervous system. We considered the case of 3D motion direction as an example, demonstrating that a geometrically constrained encoding model for 3D motion direction is consistent with electrophysiological recordings of neurons in MT and human perception as measured by direction estimation tasks. This framework can be extended to other visual domains (e.g. slant/tilt, 3D structure from motion) but should also be considered more generally as an example of how tuning for a higher order feature can be computed in the brain and that the brain might fundamentally not be charged with the simple task of decoding from Gaussian banks of sensory

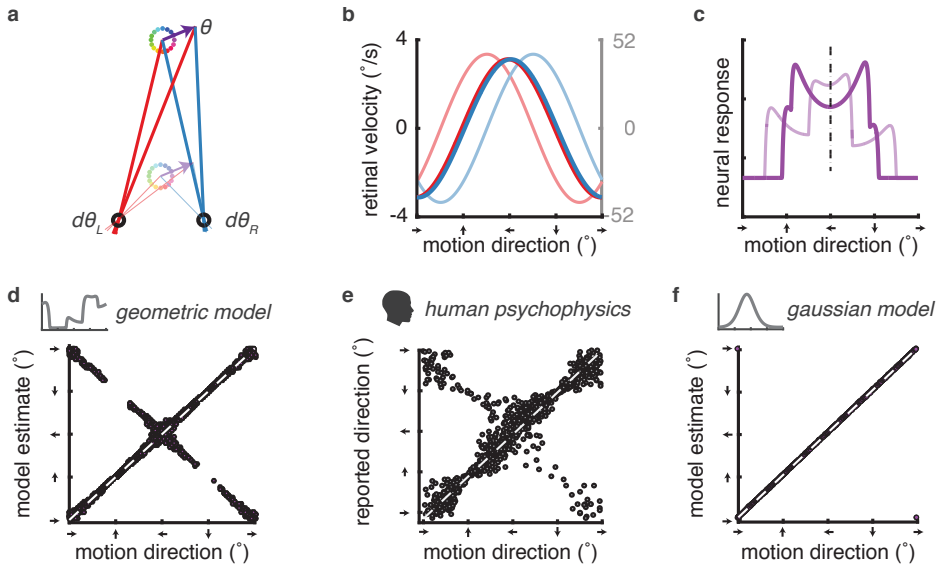

Figure 5.3: A geometric model decoder using realistic viewing distances makes a strange set of predictions that is surprisingly consistent with existing human psychophysical data. **a.** With increased viewing distance, the angle formed by the two eyes and the location of motion is compressed. **b.** For motions at large viewing distances, the retinal velocities are smaller in magnitude and the difference between the left and right eye retinal velocities is considerably less. **c.** The impact of an increased viewing distance on individual tuning curves is the convergence of steep transitions on the toward and away motion directions. This results in a relatively symmetrical neural firing rate across the left and right motion directions (i.e. the x-axis) except close to toward and away. This symmetry across the whole population leads to the unusual model errors evident in d. **d.** Geometric model decoder estimates 3D motion direction for a far viewing distance (64 cm). Many estimates (dots) still fall on the unity line (dashed white line), however a clear pattern of errors emerges result in a 'X' pattern, corresponding to both veridical estimates and sign errors. **e.** Motion directions reported by human psychophysical observers (Fulvio et al. 2015; Figure 3A) asked to estimate the motion direction of a 3D motion stimulus at viewing distance of 90 cm. Human observers make similar errors to the geometric model decoder, showing a confusion over the sign of the depth component of the motion. **f.** The Gaussian decoder is unaffected by viewing distance. It successfully estimates 3D motion direction; estimates (purple dots) fall on the unity line (dashed white line).

encoders.

5.4 Supplemental

Encoding. Monocular velocity tuning curves in MT are well-fit by log-Gaussian functions. In the case of purely direction selective MT neurons tuned to left or right, one motion direction (left or right) is largely unmodulated by changes in speed, but MT neurons are diverse in terms of their direction selectivity and thus the log-Gaussian function must be fit to both directions with coefficients to modulate the relative amplitude of the neural response:

$$f_L(\theta) = \begin{cases} \frac{a_{l+}}{L(\theta)\sigma_l} e^{\frac{-(log\, L(\theta)-\mu_l)^2}{2\sigma_l^2}} + b_L & L(\theta) \geq 0 \\ \frac{a_{l-}}{|L(\theta)|\sigma_l} e^{\frac{-(log\, |L(\theta)|-\mu_l)^2}{2\sigma_l^2}} + b_L & L(\theta) < 0 \end{cases} \tag{5.4.1}$$

$$f_R(\theta) = \begin{cases} \frac{a_{r+}}{R(\theta)\sigma_r} e^{\frac{-(log\, R(\theta)-\mu_r)^2}{2\sigma_r^2}} + b_R & R(\theta) \geq 0 \\ \frac{a_{r-}}{|R(\theta)|\sigma_r} e^{\frac{-(log\, |R(\theta)|-\mu_r)^2}{2\sigma_r^2}} + b_R & R(\theta) < 0 \end{cases} \tag{5.4.2}$$

where μ_L, σ_L, μ_R, and σ_R are the parameters of the log-Gaussian function; a_{l+}, a_{l-}, a_{r+}, and a_{r-} are the coefficients modulating the relative amplitude of the neural response; b_l and b_r are the baseline firing rates; $L(\theta)$ and $R(\theta)$ are functions that give the retinal velocities for the left and right eyes respectively (see below), given the xz motion direction θ.

$$L(\theta) = \frac{cos(\theta) * m * z - sin(\theta) * (x + \frac{ipd}{2})}{x^2 + z^2} \tag{5.4.3}$$

$$R(\theta) = \frac{cos(\theta) * m * z - sin(\theta) * (x - \frac{ipd}{2})}{x^2 + z^2} \tag{5.4.4}$$

The encoding model for 3D motion direction predicts that the binocular response ($f_B(\theta)$) is a linear combination of the monocular responses ($f_L(\theta)$, $f_R(\theta)$)) to the left and right eye retinal velocities resulting from that 3D motion direction:

$$f_B(\theta) = c_L * f_L(\theta) + c_R * f_R(\theta) \tag{5.4.5}$$

where $f_B(\theta)$ is the binocular response to 3D motion (purple trace, Figure 5.1b, fourth panel); $f_L(\theta)$ and $f_R(\theta)$ are the monocular responses to the corresponding left and right eye retinal velocities (red and blue traces, Figure 5.1b, third panel); c_L and c_R are the coefficients for linear combination (these allow for suppression or facilitation of the monocular responses).

Decoding. 3D motion direction estimation was performed by finding the peak of the log-likelihood function as a function of 3D motion direction given the assumption of independent Poisson noise:

$$logL(\theta) = log\left(\prod_{i=1}^{N} p(\mathbf{r}_i|\theta)\right) = \sum_{i=1}^{N} log\left(\frac{f_{B_i}(\theta)^{\mathbf{r}_i}}{\mathbf{r}_i!} e^{-f_{B_i}(\theta)}\right) \quad (5.4.6)$$

$$= \sum_{i=1}^{N} log(f_{B_i}(\theta))\mathbf{r}_i - \sum_{i=1}^{N} f_{B_i}(\theta) - \sum_{i=1}^{N} log(\mathbf{r}_i!) \quad (5.4.7)$$

where $\mathbf{r}$ is the population response, a vector composed of the spike count for N neurons; and f_B are the binocular tuning curves for 3D motion direction. Motion direction was estimated by finding the $argmax_\theta\, logL(\theta)$.

Chapter 6

3D motion direction estimation

This chapter expands upon Chapter 5 using the model developed in that chapter to make behavioral predictions for a variety of psychophysical experiments. Specifically this chapter examines model predictions for viewing distance manipulations and eccentricity manipulations. The second half of this chapter presents preliminary findings from a psychophysical experiment that manipulates viewing distance and compares those results to the model predictions.

6.1 Model Predictions

This modeling effort relies on a simulated neural population initially described in Chapter 5. This population is based on fits of an encoding model to electrophysiological recordings of neurons (n=236) collected by (Czuba et al., 2014). Log-Gaussian speed tuning curves were fit to the measured monocular responses and then combination coefficients were learned by minimizing the squared error between the binocular responses to 3D motion and the linear combination of the outputs of the log-Gaussian functions (see Supplemental section in chapter 5). Each neuron's simulated response to a particular motion direction is generated by calculating the left and right eye monocular velocities, computing the linear combination of the monocular response to those monocular velocities using the combination coefficients, and adding Poisson noise.

The model predictions outlined here all involve a 3D motion direction estimation task. In order to perform model estimation, first I simulated the population's neural response to a particular motion direction. To estimate 3D motion direction for that population response, I used a Poisson independent decoder which finds the peak of the log-likelihood given by equations 5.4.6 - 5.4.7. The two sections that follow describe how the model performance on the 3D motion direction estimation task changes as a function of viewing distance and eccentricity.

6.1.1 Viewing distance manipulations

For this simulation, I manipulated the viewing distance (20cm, 31cm, and 67cm), the motion speed (5 cm/s, 7.75 cm/s, and 16.75cm/s), and the direction of motion (0°to 359°at 1°intervals. The motion speeds were chosen so that the retinal speed for left/right motion directions were approximately equivalent for the following (viewing distance, environmental speed) pairs: (20cm, 5cm/s), (31cm, 7.75cm/s), (67cm,

16.75). The viewing distances and motion speeds were also chosen to match the psychophysical experiments presented at the end of this chapter. Estimation was repeated 100 times for each motion direction, viewing distance and speed. Figures 6.1 - 6.4 summarize the model predictions for manipulations of viewing distance and speed.

The panels in figure 6.1 show increasing speed along the columns (left to right) and increasing viewing distance down the rows (top to bottom). Each panel is a heatmap binned at 10°intervals where the lighter colors represent a higher density of estimates. There are two apparent effects of increasing viewing distance: 1) The number of depth-sign errors (described in chapter 5) increases as the angular locations where errors occur expands (aligned with the location of the ocular axes). 2) There is a small lateral bias of the responses near 0°(right) and 180°(left). Also notice the brighter flat line segment at motion directions near 180°at the farthest viewing distance (see the bottom row of figure 6.1). This density is much lighter for the close viewing distance (top panel). There are no immediately obvious effects of increasing speed for the model.

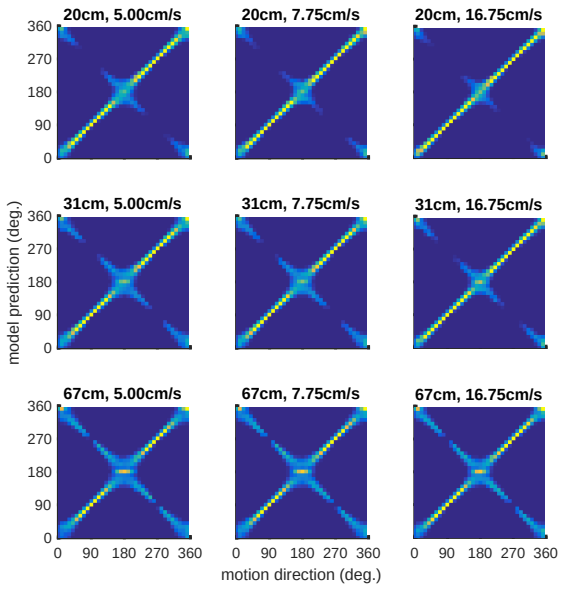

Figure 6.1: Model predictions as a function of viewing distance and speed

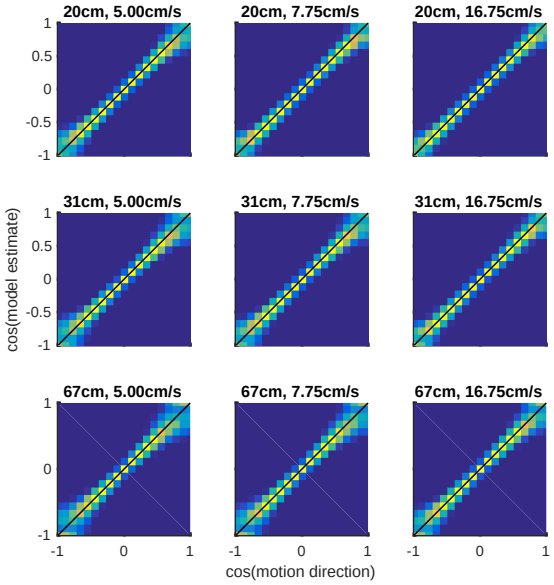

Figure 6.2: Cosine component of the model predictions as a function of viewing distance and speed.

The subsequent two figures (6.2 and 6.3) break down the estimates shown in figure 6.1 into their lateral (cosine) and depth components (sine), figures 6.2 and 6.3 respectively. Figure 6.2 plots the cosine of the model estimate as a function of the cosine of the presented motion direction. Similarly, figure 6.3 plots the sine of the model estimate as a function of the sine of the presented motion direction. Looking at the two components separately it is apparent that the changes in model performance as a function of viewing distance are primarily driven by changes in estimation of the depth component. Notice in figure 6.3 another 'X' pattern of results. Again the arms of the 'X' off the unity line represent the depth-sign errors. The emergence of the small lateral bias at far viewing distances is also more obvious in the sine plot (figure 6.3). The signature of this effect is the appearance of the flat disc of density about motion directions with a sine component near 0 (i.e. 0°and 180°, right and left).

Finally, separating the two components also reveals a small effect of the speed manipulation that is visible in both figures 6.2 and 6.3. There is a decrease in the variability of estimates at higher speeds particularly for motion directions where the

cosine is near 1 or -1, i.e. 0°or 180°(right or left). This effect is most apparent for the farthest viewing distance for the x component (figure 6.2, the closest viewing distance in the x component (figure 6.3), and the closest viewing distance in the original model estimate figure (6.1)

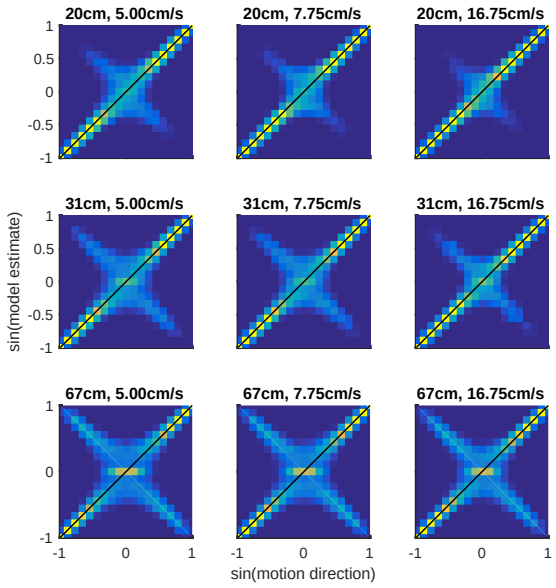

Figure 6.3: Sine component of the model predictions as a function of viewing distance and speed.

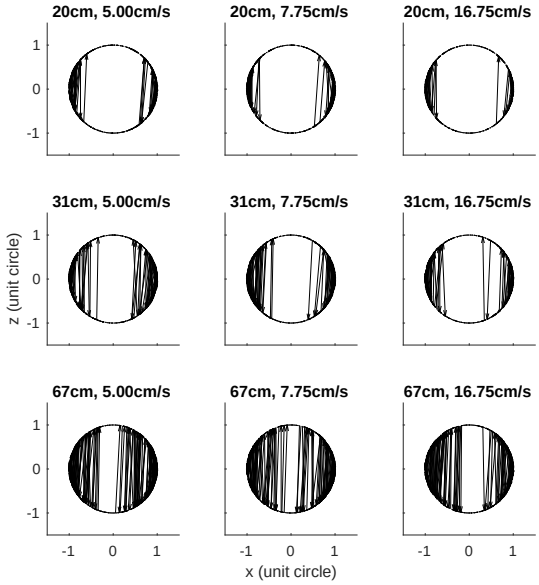

Figure 6.4: Visualization of errors for the model predictions as a function of viewing distance and speed.

Figure 6.4 illustrates the depth-sign errors that appear as a function of viewing distance. In this figure, every trial is represented by a single arrow from the presented motion direction to the model estimate on the unit circle. Small errors result in very small arrows and give the impression of a circle. Larger arrows form lines that cross the circle. Notice that all of these lines drawn across the circle are vertically oriented – indicating a depth-sign error.

In summary, the main predictions of the model for manipulations of viewing distance are an increase in depth-sign errors with a small lateral bias at far viewing distances. For the manipulation of environmental speed the main prediction is a decrease in variability as a function of the environmental speed.

6.1.2 Eccentricity manipulations

For this simulation, I held the viewing distance constant (67cm), while I manipulated eccentricity (0cm, 12.5cm, 22.5cm), the motion speed (5 cm/s, 7.75 cm/s, and 16.75cm/s), and the direction of motion (0°to 359°at 1°intervals. Estimation was

repeated 100 times for each motion direction, viewing distance and speed. Figures 6.5 and 6.6 summarize the model predictions for manipulations of viewing distance and speed.

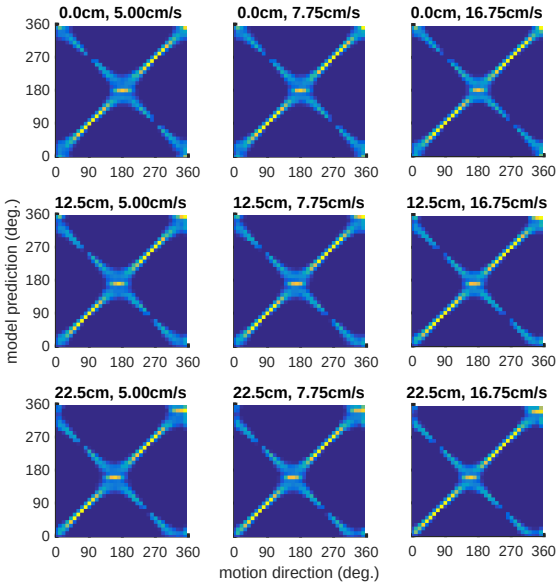

Figure 6.5: Model predictions as a function of eccentricity and speed

The panels in figure 6.5 show increasing speed along the columns (left to right) and increasing eccentricity down the rows (top to bottom). Each panel is a heatmap binned at 10°intervals where the lighter colors represent a higher density of estimates. The manipulation of eccentricity causes a shift in the pattern of errors. This is most apparent from the visualization of errors in figure 6.6. Notice that at the eccentric locations (rows 2-3) the error pattern is rotated. This rotation matches the shift eccentricity.

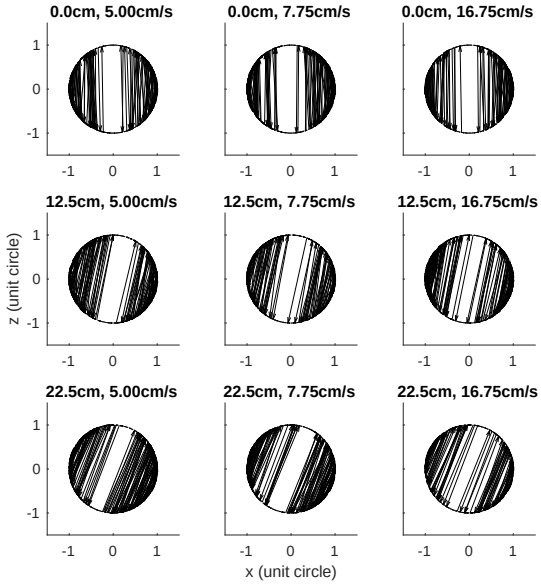

Figure 6.6: Visualization of errors for the model predictions as a function of eccentricity and speed.

6.2 Psychophysical Experiments

6.2.1 General Methods

Subjects

All subjects had normal or corrected-to-normal vision. All subjects participated with informed consent in accordance with the University of Texas at Austin Institutional Review Board. They were treated according to the principles set forth in the Declaration of Helsinki of the World Medical Association.

Apparatus

In order to manipulate viewing distance in a controlled manner we designed a rear-projection display system mounted on rails (ProPixx 3D projector; Screen Tech ST-PRO-DCF black acrylic glass. This system can be easily adjusted to present stimuli at viewing distances from 20cm to 120cm. For this set of experiments the

display was set at 20cm, 31cm or 67cm (near, middle or far, respectively). Although the viewing distance can be manipulated, the projector remains a fixed distance from the screen resulting in a ratio of 14 pixels per centimeter.

Subjects viewed the stimuli binocularly using a chin cup and forehead rest to maintain head position. Subjects wore passive circular filters to view the binocular stereo stimuli. During experiments involving eccentric fixation, subjects also used a bite bar to maintain head position. Subject responses were reported using a USB knob (Griffin Technology Powermate; Nashville, TN). All experiments and analyses were run using custom code written in MATLAB using the Psychophysics Toolbox (Pelli, 1997; Brainard, 1997; Kleiner et al., 2007).

Stimuli

The stimuli presented during motion epochs were spherical dot motion volumes, 5 degrees (frontoparallel) in diameter. The dots were .4 degrees (frontoparallel) in diameter. Dots within the spherical volume were at 5% contrast (half with luminance above the background luminance and half with luminance below) and rendered with looming and expansion cues. A static frame of this stimulus (both the left and right eye images) is depicted in Figure 6.7.

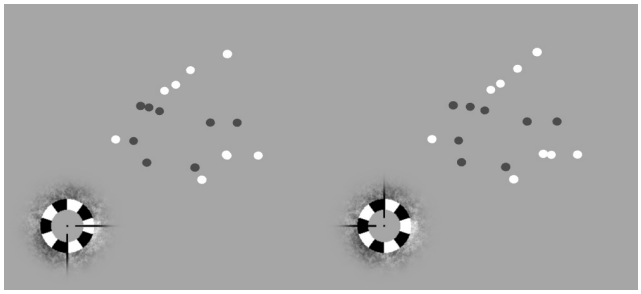

Figure 6.7: Still frame of 3D dot motion for the left and right eye. Free-fusing this image gives the percept of the dots relative depth.

Stereo-diagnostic procedure

Stereo-motion scotomas are relatively common (Barendregt et al., 2014). In order to avoid collecting data in a stereo-motion scotoma, we had subjects perform a stereo-motion diagnostic prior to data collection. Subjects viewed a cloud of dots moving

sinusoidally in depth. They were instructed to adjust the frequency of this sinusoidal motion until the point where it no longer appeared to move continuously through depth; reported percepts at/past this point typically involve flickering of dots across depth planes. This diagnostic was used to determine which motion locations were best for stimulus presentation (left vs. right, up vs. down).

6.2.2 Viewing Distance Manipulations

Methods

Each trial consisted of a single motion epoch that lasted 1 second. Subjects reported the motion direction of the dots using a knob to adjust the angle of an indicator that was also rendered in the virtual space. Each block consisted of 72 trials and was performed at a single viewing distance and location (5°up/down, or 5°left/right). Within the block the motion speed was varied (5cm/s, 7.75cm/s or 16.75cm/s) and motion direction was varied (between 0°and 360°at 5°intervals). The subject completed 70 blocks, resulting in 5 data points per location/direction/speed/viewing distance.

Results

Figure 6.8 shows the results from a single subject collapsed across motion locations. The panels show increasing speed along the columns (left to right) and increasing viewing distance down the rows (top to bottom). Each panel is a heatmap binned at 15°intervals where the lighter colors represent a higher density of estimates.

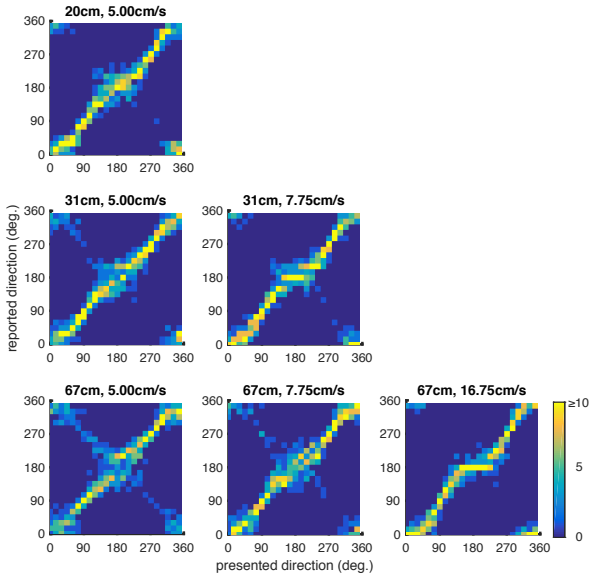

Figure 6.8: 3D motion direction estimation results for a single observer.

While depth-sign errors do emerge at the slowest speed and the farthest viewing distance, depth sign errors are not as prevalent at the fastest environmental speed. Moving along the diagonal, preserving retinal speed for frontoparallel motion shows basically no increase in depth sign errors. Instead there is an emergence of a lateral bias, a 'flattening' of responses to the frontoparallel plane. This bias is consistent with previous work in 3D motion direction estimation (Welchman et al., 2004, 2008), which predicts that this effect is due to a slow-speed bias. The tightening of the response around the frontoparallel plane due to this bias effectively prevents the depth sign errors observed at slower speeds for the farthest viewing distance.

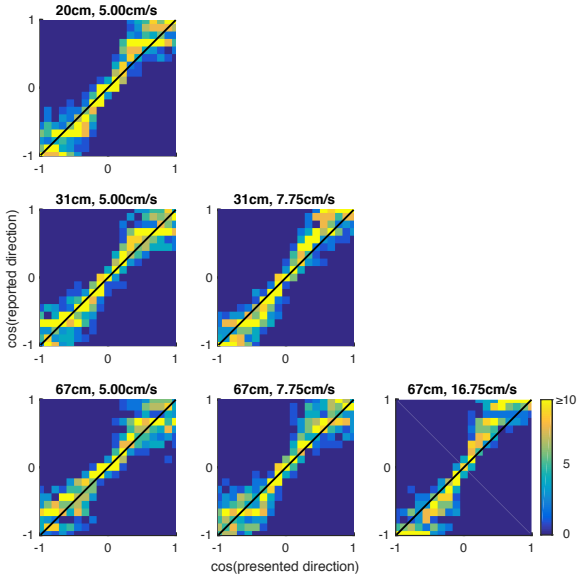

Figure 6.9: Cosine component of psychophysical 3D motion direction estimates.

The responses can be divided into their trigonometric components for a more clear picture of the performance. Figure 6.9 plots the cosine of the reported motion direction as a function of the cosine of the presented motion direction. From the panels in this figure, the emerging lateral bias at the farthest viewing distance and fastest speed is clear. Notice the rotation of the line of density slightly counter-clockwise from the unity line (thin black line). This rotation essentially means that the subject's response to motion directions with cosines close to 1 or -1 (i.e. right and left) are being reported as closer to right and left than they actually are. There is also a tightening of the variability of the responses as a function of increased environmental speed.

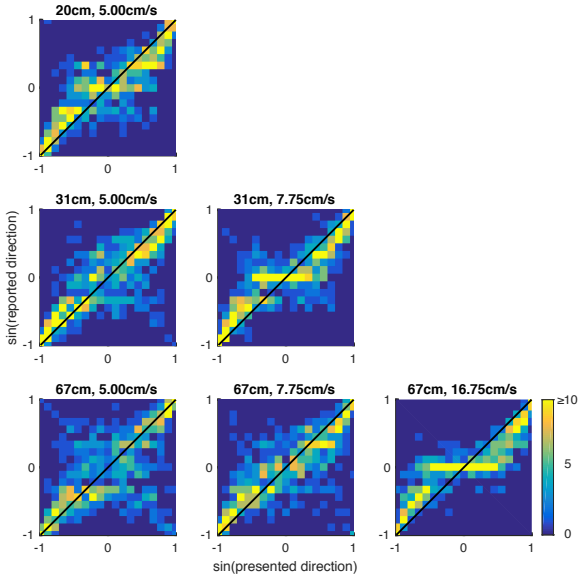

Figure 6.10: Sine component of psychophysical 3D motion direction estimates.

Figure 6.10 plots the sine of the reported motion direction as a function of the sine of the presented motion direction. Once again the lateral bias emerging at the farthest viewing distance and fastest speed is clear; the horizontal line of density is a signature of that lateral bias. Again with increased environmental speed we observe a decrease in the variability.

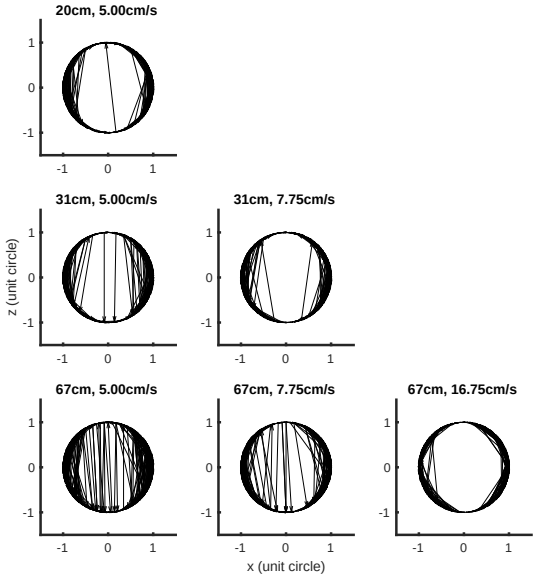

Figure 6.11: Visualization of errors for the 3D motion direction estimation results.

Figure 6.11 is a visualization of the errors. Each trial is represented by a signal arrow from the presented motion direction to the reported motion direction. The increase in large depth-sign errors (i.e. the long vertical arrows) requires both an increase in viewing distance and a relatively slow speed.

6.3 Discussion

In the first part of this chapter I laid out a set of predictions for human psychophysical performance in 3D motion direction estimation based on the model developed in chapter 5. The main predictions for manipulations of viewing distance and environmental speed were: 1) increased depth-sign errors with increased viewing distance, 2) an emerging lateral bias for far viewing distances, and 3) decreased variability with increased environmental speed.

The second half of this chapter presents preliminary data collected to examine human psychophysical performance in 3D motion direction estimation. These psychophysical data was collected in a single subject; more data from additional subjects are needed to make more definitive conclusions about human performance

and to make meaning comparisons to the model performance. Additionally, while the data from this sinlge subject represents a complete data set, there are known problems with how the motion direction was reported that may influence the results (particular those related to lateral biases) in ways that were not intended. Specifically there were markers every 22.5°that had a categorical effect on reported motion directions for some subjects' responses. Even placing markers at the cardinal directions is likely problematic because it provides additional information which may be combined with the motion cue differently for different motion directions. The markers have been removed and the subsequent data will not suffer from this issue. Though these data are preliminary for the reasons I have described, it is still interesting and perhaps informative to the ongoing experiments to begin the process of comparing model performance to human performance.

The results from the psychophysical experiment show: 1) an increase in depth-sign errors for the combination of increased viewing distance and decreased environmental speed, 2) emergence of a large lateral bias for large viewing distance and faster environmental speeds, 3) decreased variability at faster environmental speeds. These findings are only partially consistent with the main findings from the model. The primary difference is that the model predicted a far less drastic effect of environmental speed. In particular the depth-sign errors are basically non-existent at the far viewing distance for the fast environmental speed (see bottom right panel of 6.8), meaning that depth-sign errors in humans result from large viewing distances *combined with* slower environmental speeds modulate these errors. This suggests the the errors are driven primarily by slower retinal velocities rather than the pure manipulation of viewing distance.

The strength of the lateral bias at the farthest viewing distance and fastest environmental speed is striking. This effect has been reported in previous studies of 3D motion estimation (Welchman et al., 2004) and the accepted explanation involves the presence of a slow-speed bias (Welchman et al., 2008). This explanation motivated testing the decoder in a new way, decoding from the model assuming a slower speed (note: decoding up to this point assumed a correct estimate of environmental speed). Decoding under this assumption does result in a lateral bias.

These preliminary results are encouraging. Human subjects can estimate 3D motion direction in the psychophysical regime we have designed for testing manipulations of viewing distance and eccentricity. It appears from the preliminary

data that human performance may differ systematically from the model predictions. More data must be collected across a greater number of subjects, before a definitive set of conclusions can be drawn.

Chapter 7

Discussion

The study of perception in the primate visual system has yielded a large body of research that makes fundamental contributions to the understanding of neural information processing. This progress required the simplification of stimuli, and significant task constraints. The advantage of these simplifications was more precise control of stimuli and measurements, and well-defined models for behavior/model comparison. The downside of such simplifications is that it is often unclear how these findings generalize to the visual system in the natural environment outside of artificial laboratory tasks. However, emerging statistical and technological tools are allowing scientists to move beyond these constraints. Growing bodies of research are moving to more natural stimuli and introducing more complex/natural tasks (see Chapter 4 for a review of some of this work).

The work presented here joins that growing body of research. In that regard, there are two main contributions of this thesis. The first is the introduction of a novel psychophysical paradigm and analysis framework for the study of visual perception in the context of sensorimotor control loops, i.e. target tracking (Chapter 2). The second contribution is an elaboration of the neural coding approach that incorporates the environment-to-retina transformation, an extension which is central to answering questions about how the brain could encode and decode information about its natural 3D environment (Chapters 5 - 6). In this discussion I elaborate on these findings, their potential impact, and possible future directions of this work.

7.1 Target-tracking paradigms for examining visual perception

My work demonstrates that a target-tracking task combined with a Kalman filter analysis framework results in measures of visual sensitivity comparable to measures of visual sensitivity collected with more traditional psychophysical methods and analysis (i.e., forced choice tasks and signal detection theory). The two immediate advantages of the tracking paradigm over more traditional psychophysical methods are the decrease in the experimental time required to arrive at a measurement for an experimental condition (see figure 2.12), and the finer temporal scale of data collection. The latter of these two advantages is particular important in the context of typical measures of neural signals used to examine visual perception and decision-making in human and non-human primates (e.g., EEG, electrophysiology, etc.) Such

measures typically result in hundreds (or more) samples per second, whereas the typical psychophysical experiment gives approximately one bit of behavioral data per second. Relating data of such disparate temporal scales is incredible difficult and often requires model assumptions which cannot be tested in meaningful ways because of the lack of behavioral data. The first section of chapter 4 discusses these issues in much greater detail. One of the most important future developments of this work will be to extend the Kalman filter analysis framework so that it can be used to relate behavior and neural activity in the same way neuroscience has leveraged signal detection theory.

While the increased temporal resolution is particularly valuable in the context of high resolution measures of neural activity, it also critical for a more complete understanding of visual perception and human behavior. The 3D motion tracking experiment presented in chapter 3 provides a simple example of the advantages of the temporal resolution of tracking tasks for dissecting and analyzing behavioral/perceptual results. The main finding for 3D motion tracking was that the selective impairment of 3D motion perception during target-tracking was driven by two factors: the geometric constraint that motion-through-depth yields much smaller retinal projections than frontoparallel motion, and slower latencies associated with binocular disparity processing. Using the tracking paradigm, we were able to distinguish between the effects of the geometric constraint (primarily an effect on the signal-to-noise-ratio) and the effects of the limits of binocular disparity processing (primarily an increased latency). While it's certainly possible that similar results could have been acquired using more traditional psychophysical methods, the amount of experimental time required for the number of conditions tested would have been prohibitive (not to mention additional conditions tested that did not make it into the final manuscript). Furthermore, such experiments would have resulted in single values for performance and response time, instead of a full spatiotemporal response function. One of the most striking aspects of the final experiment of chapter 3 is the close match between the spatiotemporal response functions for frontoparallel and motion-through-depth tracking when frontoparallel motion is disparity-limited in the same manner as motion-through-depth (see figure 3.13, particularly subjects 1 and 2).

Another important opportunity that the tracking paradigm offers is the potential to work with a variety of 'non-traditional' vision science subjects: infants

and children, certain clinical patients, and other populations that may get impatient with traditional psychophysics. I address this point briefly at the beginning of chapter 2, but here I wish elaborate because of the successes I have observed in the research of colleagues and collaborators. Figure 7.1 shows data collected by Dr. Rowan Candy's lab at Indiana University. Dr. Candy studies the development of human vision in infants and young children. The figure shows the horizontal and vertical position of a target for a single trial and the eye tracking responses of 2 infants (77 days and 76 days old) and an adult. Infants as young as 11 weeks (and younger) will track targets with their eyes.

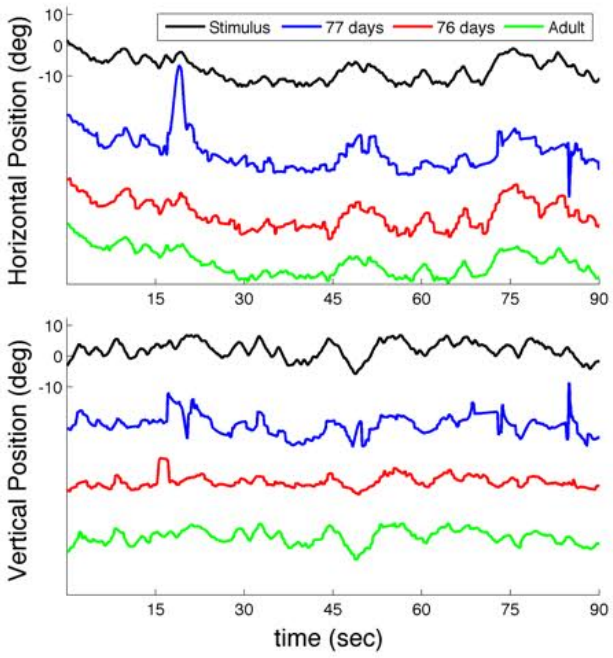

Figure 7.1: Example of the tracking paradigm for human infant subjects. Black traces represent the horizontal and vertical position of a stimulus. The colored traces represented the eye tracking traces from 2 infants (blue and red) and an adult (green).

It may be somewhat unsurprising that relatively young infants can track moving objects but the tracking paradigm and the kalman filter analysis provide a framework to harness this behavior to better understand the development of human vision across infancy and childhood. Given that the current gold standard for the study of perception in infants is the preferential looking paradigm (Fantz, 1963), this presents a huge opportunity for increasing the understanding of human visual development. More anecdotally, this technique is also proving useful in the study of non-human primates. Non-human primates (e.g. macaques, marmosets) will perform tracking tasks with minimal training compared to more traditional psychophysical tasks.

7.2 Incorporating the three-dimensional environment into neural coding models

The second half of this dissertation focuses on incorporating the three-dimensional environment into neural coding models of neurons in visual areas. With some notable exceptions (e.g. Baker & Bair 2016), there has been minimal effort to generalize neural coding models of visual stimuli in a manner that supports questions about how the brain could encode and decode information about its natural 3D environment. The model presented in chapter 5 builds a neural coding model for 3D motion stimuli and neural activity in primate middle temporal area (MT). The principle findings from this exercise in modelling were: 1) Neural representations are shaped by the environment-to-retina transformation resulting in encoding model tuning curves that are atypical but match existing electrophysiological measurements, 2) Relatively small differences in retinal velocity tuning across the two eyes can provide 3D motion sensitivity, 3) Decoding from neural representations shaped by the environment-to-retina transformations results in surprising predictions for perception; some of these predictions are confirmed by the existing psychophysical literature. Chapter 6 examines these predictions in greater detail and reports preliminary results from psychophysical experiments designed to test these predictions.

Continuing psychophysical work will compare the model predictions for 3D motion direction estimation with psychophysical performance across manipulations of viewing distance, speed and eccentricity. Motion adaptation will also be key to further psychophysical experiments in order to better explore of the structure of the

neural encoding of 3D motion. Historically, motion adaptation has proven to be a powerful tool for probing visual information processing in the human brain (Clifford et al., 2007). Here it will serve as method for examining to what degree the neural representation of 3D motion is integrated binocularly and globally versus simply inherited from local, monocular 2D signals.

Future electrophysiological work should record from awake-behaving non-human primates (the simulated neural population in chapters 5 and 6 are based on electrophysiological recordings from the MT of anesthetized macaques). While we have observed signatures of the proposed encoding model in human psychophysics it will be important to record from (and inactivate) neurons in awake-behaving primates across multiple areas in the dorsal stream (e.g. middle temporal area, medial superior temporal area, and ventral intraparietal area), while manipulating viewing distance and eccentricity. These recordings in conjunction with the motion adaptation experiments provide a much clearer picture of the neural information processing for 3D motion, particularly in regards to how motion is integrated binocularly and globally.

The extensions I have mentioned thus far have focused entirely on experiments to further understand 3D motion perception. However the neural coding approach presented in chapter 5 could applied to other domains (e.g. slant/tilt, structure from motion). The simplest of these extensions is to slant/tilt. Such a model would combine monocular orientation tuning curves, in place of the binocular combination of monocular velocity tuning curves. This type of model would complement existing normative models of slant/tilt perception that explain human perceptual behavior via the analysis of natural scene statistics (Burge et al., 2016; Kim & Burge, 2017).

7.3 Optimal filters for 3D motion perception

The encoding model presented in Chapter 5 proposes a functional form for a set of filters that encode 3D motion direction. The model provides a good fit to electrophysiological recordings of neurons in area MT, however it is a purely descriptive model and does not speak to whether such filters are optimal. Intuitions from other features encoded by the visual system (e.g. orientation, 2D motion direction, 2D motion speed) would suggest that 3D motion direction should be optimally encoded by

a set of filters which resemble evenly distributed Gaussian functions. The proposed set of filters are not consistent with such a representation. However, it remains an open question whether such filters are (or resemble) the optimal filters for encoding 3D motion given the biological constraints of the sensors in the visual system.

Efficient coding approaches have been used to derive filters and their properties for a variety of of sensory features (e.g.,Olshausen & Field 1996, 1997; Bell & Sejnowski 1997; van Hateren & van der Schaaf 1998; Lewicki 2002; Ganguli & Simoncelli 2016). These methods derive filters based on an efficient representation of the statistical properties of the stimuli. Accuracy Maximization Analysis (AMA) is a task-dependent framework which takes this approach farther, specifying which information should be represented given a specific task. It has been used to derive filters for a variety of tasks: image patch identification, foreground identification, defocus blur discrimination, disparity estimation, and speed estimation (Geisler et al., 2009; Burge & Geisler, 2011, 2014, 2015). AMA seems an appropriate choice for deriving optimal filters for 3D motion direction estimation. One potential challenge in implementation is that these approaches typically use databases of natural stimuli designed to effectively sample the feature in question. To my knowledge, such a database doesn't exist for 3D motion. However, recent advances in depth camera technologies make the collection of such a database significantly easier. Furthermore, early work using AMA to derive optimal 3D motion direction filters could be conducted on databases of simulated 3D motion prior to the collection of a database of natural stimuli.

7.4 Self-motion, optic flow, and binocular information

The work presented in this dissertation contributes to the effort to understand the relationship between neural activity and visual perception in the natural environment, focusing in particular on 3D motion. However the work here has focused on primarily on issues related 3D *object* motion, largely ignoring an critical component of motion in the natural environment: self-motion.

Appendix A examines questions relevant to self-motion, navigation and interocular velocity signals. The majority of studies of velocity-based 3D motion perception in primates have focused on interocular comparisons which occur in the binocular portion of the visual field. This study (inspired by studies of bee vision)

examines interocular velocity signals across the two monocular portions of visual field (i.e. the far periphery). The main finding of this work was that there was a privilege for speed discrimination across the two eyes compared to within a single eye's monocular visual field. Furthermore, this privilege was limited to cases consistent with ecologically valid self-motion.

Work on self-motion has largely focused on the combination of rich monocular visual signals (i.e. optic flow) and vestibular signals (see DeAngelis & Angelaki 2012 for review), largely ignoring the role of binocular information. However a recent review of the binocular mechanisms of 3D motion proposed the binoptic flow field (BFF), a stimulus which more completely encompasses the 3D information typically available in the natural environment, including the binocular information (Cormack et al., 2017). In examining the binoptic flow field, it becomes clear that the differential binocular signals are valuable sources of information and are especially pronounced in the binocular periphery. Future work leveraging such stimuli and examining the natural statistics of self-motion in the natural environment will be critical to understanding the neural activity that supports motion perception during self-motion and its relationship to the perception of objection motion, which is inherently a more local process.

7.5 Binocular cues for 3D motion at *far* viewing distances

The preliminary psychophysical results presented in chapter 6 are collected at three different distances: near (20cm), middle (31cm), and far (67cm). But even the "far" distance used is relatively close. These particular distances were chosen in order observe the most change in performance while also being certain that the viewing distance was within a range where the cues were still useful.

This raises important questions about what range of viewing distances lead to binocular 3D motion cues that are perceptible. Previous work examining natural static binocular disparity distributions determined that significant portions of the distributions of binocular disparities in the natural environment remain suprathreshold even at fixation distances greater than 15 meters (Liu et al., 2008). But it remains an open question whether binocular cues to 3D motion (i.e., changing disparities, CD; inter-ocular velocity differences, IOVD) are useful at that distance.

Early work on stereo-motion perception established that human subjects were less sensitive to the binocular percept than the monocular percept at the fovea (Tyler, 1971). The stimulus used in Tyler (1971) was an oscillating bar and the amplitude thresholds reported suggests sensitivity to binocular motion at speeds around 1 arcmin/sec or .017°/s. More recent work tested sensitivity in the periphery to random dot stereograms, also testing IOVD cues and CD cues separately. The lowest speed tested in that experiment was .3°/s, where they reported thresholds of 10% coherence at 3°-7°and 50% coherence at 11°-15°. For the CD cue the thresholds were 10% and 20% for 3°-7°and 11°-15°respectively. For the IOVD cue the thresholds were 12% and 50%.

In order to understand what this means for 3D motion in everyday life, let's consider a handful of idealized examples. For these examples I will render binoptic flow fields at different distances and speeds. These binoptic flow fields simply show the binocular information present for motion across the visual field (e.g., figure 7.2 and Cormack et al. 2017 for more details).

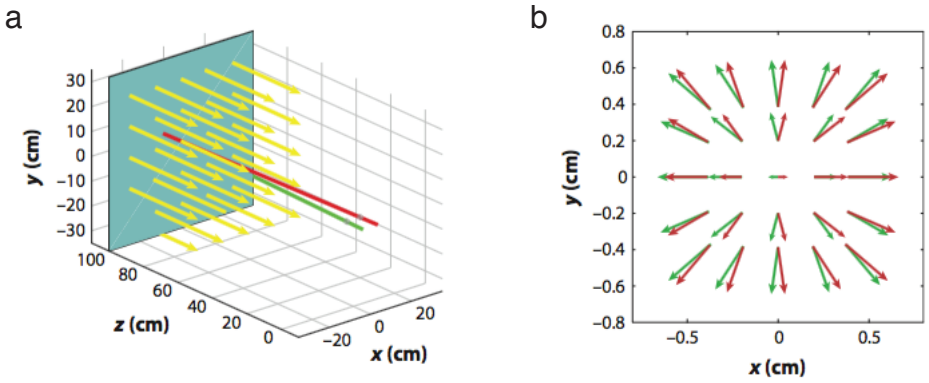

Figure 7.2: Binoptic flow fields. (Panel a) A 3D schematic of points in space moving forward, along the z-axis. (Panel b) The projection of the 3D schematic in panel a onto the left (green) and right (red) retinae.

In the first example (figure 7.3), the simulated viewing distance is 1 meter (still relatively close) with a motion speed of 1.3m/s directly towards the observer. This speed is consistent with an average walking speed. Here we observe that huge

swaths of the visual field have interocular velocity differences which are relatively large ($d\theta \geq 5°$ at speeds $\sim .5°$ in the near periphery).

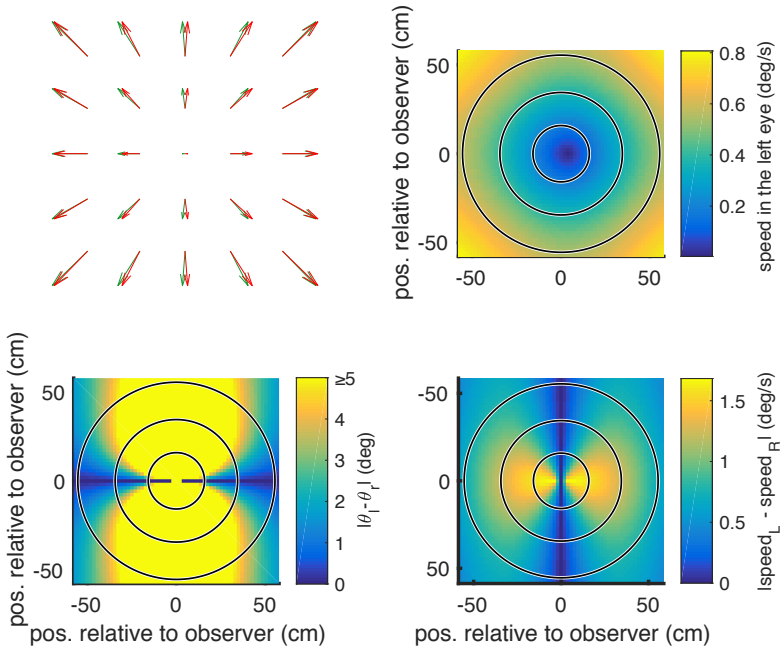

Figure 7.3: Binoptic flow field – viewing distance: 1m, motion towards at 1.3m/s. (Upper Left Panel) Binoptic flow field as projected onto the eyes. (Upper Right Panel) Magnitude of the motion in the left eye (deg/s). The black circles represent 10°, 20°, and 30°eccentricity. (Lower Left Panel) Angular difference between the motion signal in the left eye and the motion signal in the right eye (deg). (Lower Right Panel) Speed difference between the motion signal in the left eye and the motion signal in the right eye (deg/s).

At farther distances (e.g., 15 meters), the same environmental speed (1.3 m/s) results in interocular velocity differences which appear far less useful (see Figure 7.4). Here $d\theta \geq 2°$ within the central 20 °, but at these eccentricities speeds are $\leq .01°/s$ or 36 $arcsec/s$, which is certainly subthreshold. Tyler (1971) reported stereo-motion thresholds of $\sim 35 - 200$ $arcsec/s$ at fixation for targets which were moving sinusoidally directly towards and away (i.e., $d\theta = 180°$).

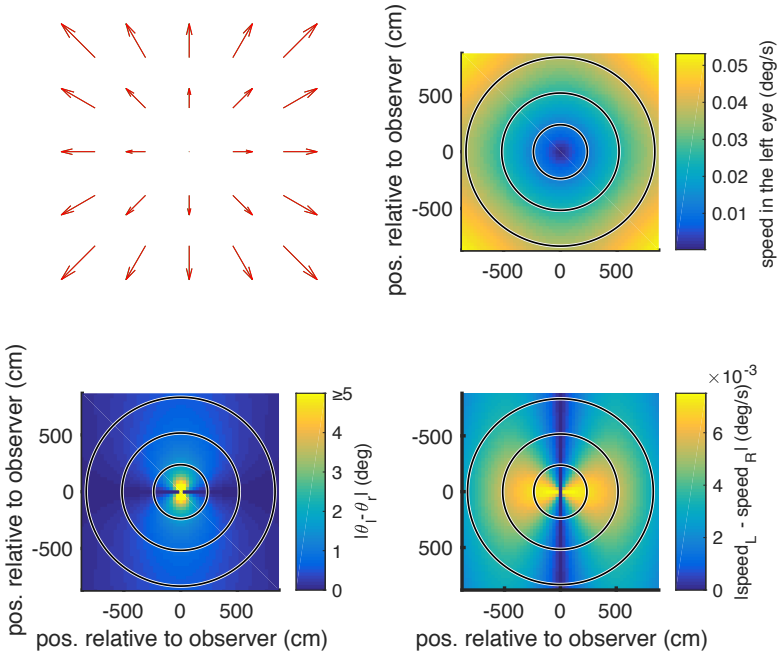

Figure 7.4: Binoptic flow field – viewing distance: 15m, motion towards at 1.3m/s. (Upper Left Panel) Binoptic flow field as projected onto the eyes. (Upper Right Panel) Magnitude of the motion in the left eye (deg/s). The black circles represent 10°, 20°, and 30°eccentricity. (Lower Left Panel) Angular difference between the motion signal in the left eye and the motion signal in the right eye (deg). (Lower Right Panel) Speed difference between the motion signal in the left eye and the motion signal in the right eye (deg/s).

Increasing the environmental motion to $13m/s$ (a speed consistent with a car moving at $29mph$) does increase the retinal speeds observed but doesn't have an effect on the distribution of $d\theta$ which is directly related to viewing distance. Though the speeds are perceptible, the $d\theta$ remains small. Thus, it is unlikely that there is a useful *stereo* signal.

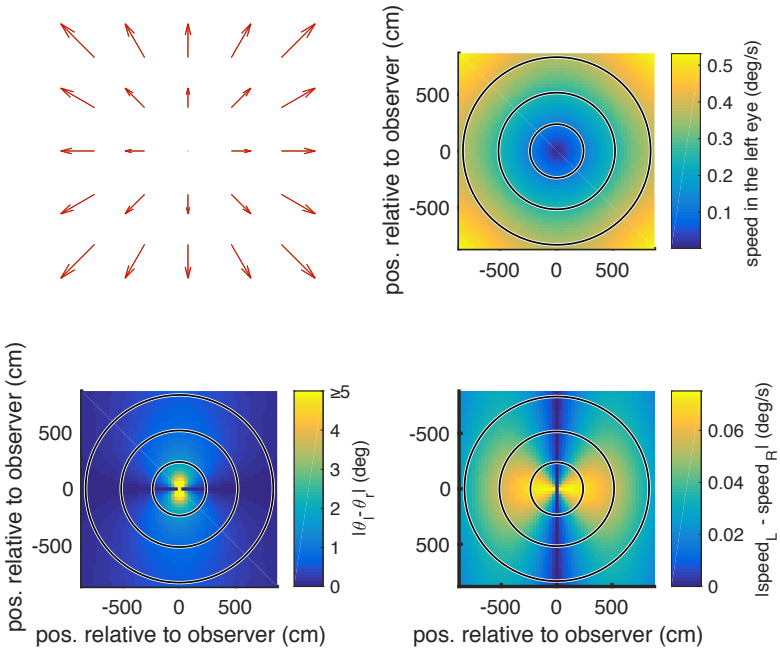

Figure 7.5: Binoptic flow field – viewing distance: 15m, motion towards at 13m/s. (Upper Left Panel) Binoptic flow field as projected onto the eyes. (Upper Right Panel) Magnitude of the motion in the left eye (deg/s). The black circles represent 10°, 20°, and 30°eccentricity. (Lower Left Panel) Angular difference between the motion signal in the left eye and the motion signal in the right eye (deg). (Lower Right Panel) Speed difference between the motion signal in the left eye and the motion signal in the right eye (deg/s).

While this exercise certainly demonstrates that stereomotion cues are not useful at extremely far distances (i.e., 15 meters), they may still be useful in intermediate ranges (e.g., 1-2 meters). This is beyond the reach of the arm and outside of the peripersonal space. It is in a range that might make it useful for choosing footholds while navigating difficult terrains (i.e. looking at the ground) or picking out a path through a dense forest.

7.6 Artificial vs. natural

This dissertation has emphasized the importance of extending laboratory tasks to more closely resemble the 'natural environment' through continuous tracking and tasks that include depth and motion-through-depth. These efforts do not rest on the assumption that tasks that more closely resemble the natural environment are inherently better for interrogation of neural systems and visual perception. There are tradeoffs between artificial tasks/stimuli vs. natural tasks/stimuli.

Rust & Movshon (2005) examines this issue with respect to natural versus artificial stimuli. They conclude that natural stimuli may be appropriate for exploratory experiments, particularly for interrogating neurons with complex properties in higher cortical areas, but that parameterized artificial stimuli are critical for the development of testable models of cortical neurons (and by extension visual perception). In their view, because natural images have complex statistics that are often poorly understood, it is incredibly difficult to rely on such images for the purposes of model building and (to some degree) model testing. Recent physiological and perceptual work has begun to challenge this conclusion (e.g. Coen-Cagli et al. 2015; Sebastian et al. 2017).

However, I believe their general point is useful – the real issue is whether stimuli (or behavior) can be used to build testable models and ask useful questions of the systems under examination. Does the stimulus/task help us to better understand the visual experience or human behavior and its underlying neural mechanisms? It is in this context that the contributions of this dissertation are offered. In fact one of the motivations for the development of the tracking paradigm and its analysis framework (chapter 2) was the failure of traditional forced choice tasks to provide the behavioral temporal resolution to meaningfully differentiate between models of perception/behavior (see chapter 4 for a discussion). The Kalman filter analysis developed alongside the tracking paradigm provides the formal framework for using tracking behavior to evaluate hypotheses and models of visual and sensorimotor function. Future work should continue to take advantage of advances in statistical methods, so that it is possible to leverage stimuli/tasks ranging across the spectrum from artificial to natural in the effort to understand neural systems, visual perception and human behavior.

Appendix A

Speed discrimination in the far monocular periphery: A relative advantage for interocular comparisons consistent with self-motion

This work was published in the Journal of Vision. Greer, D. A., Bonnen, K., Huk, A. C., & Cormack, L. K. (2016). Speed discrimination in the far monocular periphery: A relative advantage for interocular comparisons consistent with self-motion. Journal of Vision, 16(10), 7-7.

Author contributions: D.G. was the primary author for this paper. D.G., K.B. , A.H., and L.C. conceived and designed research; K.B. and D.G. performed experiments; D.G., K.B. , A.H., and L.C. analyzed data; D.G., K.B. , A.H., and L.C. interpreted results of experiments; D.G., K.B. , A.H., and L.C. prepared figures; D.G., K.B. , A.H., and L.C.. drafted manuscript; D.G., K.B. , A.H., and L.C. edited and revised manuscript; all authors approved final version of manuscript.

Some animals with lateral eyes (such as bees) control their navigation through the 3D world using velocity differences between the two eyes. Other animals with frontal eyes (such as primates, including humans) can perceive 3D motion based on the different velocities that a moving object projects upon the two retinae. Although one type of 3D motion perception involves a comparison between velocities from vastly different (monocular) portions of the visual field, and the other involves a comparison within overlapping (binocular) portions of the visual field, both compare velocities across the two eyes. Here we asked whether human interocular velocity comparisons, typically studied in the context of binocularly-overlapping vision, operate in the far lateral (and hence monocular) periphery and – if so – whether these comparisons were accordant with conventional interocular motion processing. We found that speed discrimination was indeed better between the two eyes' monocular visual fields, as compared to within a single eye's (monocular) visual field, but only when the velocities were consistent with self-motion. This intriguing finding suggests that mechanisms sensitive to relative motion information on opposite sides of an animal may have been retained, or at some point independently achieved, as the eyes became frontal in some animals.

KEYWORDS: binocular vision, interocular velocity difference, monocular vision, motion

A.1 Introduction

When a bee flies through the world, its (lateral) eyes each extract different velocities to gauge its 3D heading (Srinivasan et al., 2000). When a human views an object flying towards them, their (frontal) eyes are stimulated by different velocities, which are used to estimate a 3D direction (Harris et al., 2008; Regan & Gray, 2009). There are many differences between these two domains: insect versus primate, monocular visual fields versus binocular vision, and visually-guided navigation versus object perception. However, both fundamentally involve extracting eye-specific velocities and comparing them to estimate a 3D direction.

Humans and other primates are able to perceive the 3D direction of an object based on velocities within their central visual field. In the primate object motion literature, this differential velocity cue is called the interocular velocity difference (IOVD). Conventionally, this term refers to the dichoptic comparison of velocities

from overlapping portions of the left and right eyes' visual fields (Regan & Beverley, 1973a; Cumming & Parker, 1994; Shioiri et al., 2009; Fernandez & Farell, 2006; Czuba et al., 2010; Rokers et al., 2008). Alternatively, many animals have relatively little binocular overlap because of the lateral placement of their eyes. These animals, despite their lack of stereoscopic vision, are quite adept at navigating at high speeds through complex environments. A growing body of work shows that they accomplish this by comparing the velocities viewed separately in each eye to arrive at a 3D heading (Srinivasan et al., 1991; Götz, 1968; Clark et al., 2014; Martin & Shaw, 2010; Martin, 2009; Schiffner & Srinivasan, 2015; Bhagavatula et al., 2011; Eckmeier et al., 2008).

Considering both scenarios, interocular velocity differences per se may not be limited to encoding motion-through-depth of objects relative to the observer. The concept could be extended to describe the inter-monocular velocity comparisons used for navigation by animals with lateral eyes. Both processes involve differential velocity information between the eyes, which is used to encode a 3D motion direction. In fact, the only structural difference between these interocular velocity differences is the portion of the visual field which is being used. Put another way, there may be not only a system sensitive to central binocular IOVDs in primates, but also a system sensitive to peripheral monocular IOVDs (mIOVDs).

For these reasons, we sought to better understand whether the primate visual system processes IOVDs in the monocular and binocular fields similarly, or whether it can be said to process mIOVDs at all. To do so, we developed a paradigm that links conventional binocular motion perception studies with approaches from visually-guided bee navigation literature. This was accomplished by simultaneously presenting a pair of drifting gratings exclusively in the monocular visual fields of humans. Using a range of speeds that a walking observer would view in their peripheral vision (through a hallway or forest, for example), we compared speed discrimination performance between and within the monocular fields.

One might expect that, like for most visual functions, speed discrimination performance drops considerably as the speeds are viewed at greater eccentricity (McKee & Nakayama, 1984; Wright & Johnston, 1983). However we have found a scenario in which this decline in performance is remarkably spared. Human observers were substantially better at speed comparisons when speeds were compared across our vastly separate monocular fields, and the velocities encountered by the

right eye and left eye monocular views were consistent with forward or backward self-motion, than when the same moving stimuli were presented within the same monocular field. We suggest that this robustness of inter-monocular velocity comparisons demonstrates that humans are indeed also sensitive to mIOVDs.

A.2 General Methods

Observers

Data were collected from 3 observers (aged 25-26, 1 naive, plus 2 of the authors), all with normal or corrected-to-normal visual acuity. Observers needing correction wore contact lenses rather than glasses to insure unobstructed peripheral vision. Two subjects (authors) were experienced psychophysical observers, while the naive subject had no previous psychophysical experience. All observers completed every experiment. Each observer gave written consent, and procedures were approved by The University of Texas at Austin Institutional Review Board. All data were collected at UT Austin, and all observers were recruited from the UT Austin community.

Apparatus and setup

Stimuli were generated using MATLAB (The MathWorks, Inc., Natick, MA, USA) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) on a Quad-Core Intel (Intel Corporation, Santa Clara, CA, USA) Mac Pro computer (Apple, Inc., Cupertino, CA, USA) with an ATI Radeon HD 4870 graphics card (Advanced Micro Devices, Inc., Sunnyvale, CA, USA), and displayed on three 23" monitors (NEC MultiSync PA231W LCD displays, NEC Display Solutions, Ltd., Minato-ku, Tokyo, JP). The luminance functions of the 3 displays were linearized using standard gamma-correction procedures. The displays were connected via a multi-display adaptor (Matrox TripleHead2Go, Matrox Graphics, Inc., Dorval, Quebec, CA), creating a merged display of 1920 x 480 at 60 Hz resolution.

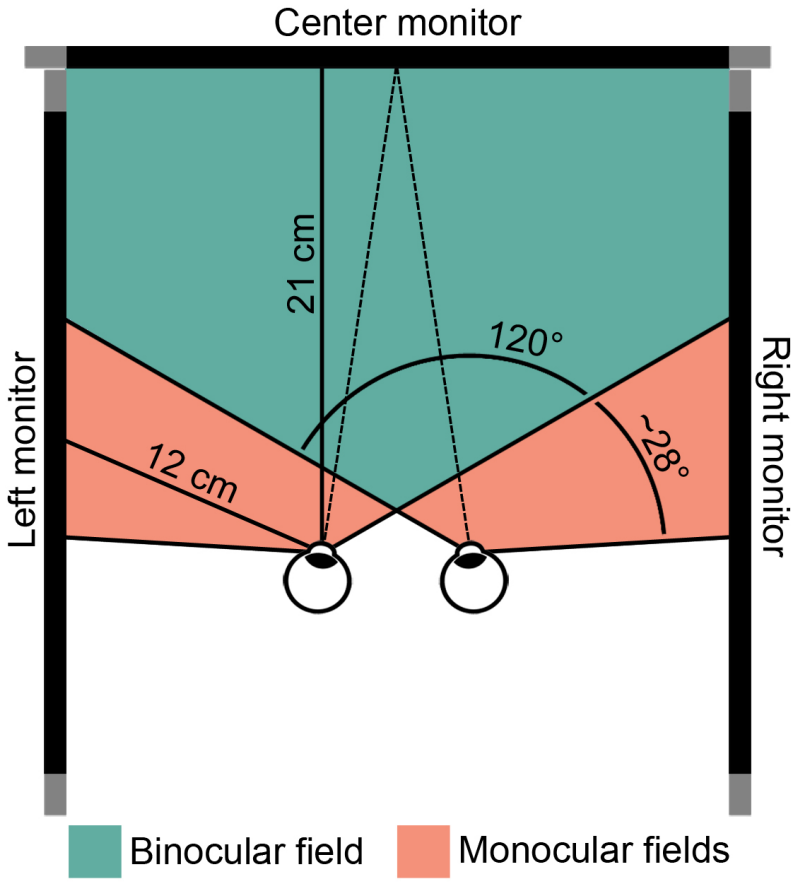

Figure A.1: Top view schematic of the 3 monitor setup, observer viewing location, and visual field locations. The observer simultaneously viewed 3 monitors: one at 21 cm directly in front, and one each to the left and the right, positioned such that stimuli could be presented in the center of the monocular fields (orange) at roughly 12 cm from the nearest eye. The typical observer's binocular field (teal) in this setup was measured to span approximately 120 visual degrees and monocular fields spanned approximately 28 visual degrees on each side.

As shown in Figure A.1, we used a triptych stimulus display comprising 3 monitors in portrait orientation (i.e., longer dimension vertical). The center monitor occupied the majority of the observer's binocular field, and the lateral monitors, each perpendicular to the center one, almost entirely filled the observer's monocular fields. Note that the nose obstructed the left display from the right eye, and vice

versa. A chin cup and forehead rest minimized observer's head movements. The forehead rest was constructed so not to occlude any peripheral vision. Because of the heat generated by the monitors, a small USB fan was used to circulate air through the interior of the monitor setup during the experiments.

Task

On each trial, observers viewed two simultaneously presented drifting gratings and indicated which of the two appeared to have moved faster. Locations of the stimulus elements (grating patches) varied, depending upon the condition (described below). Observers were instructed to maintain gaze on a fixation cross that remained in the middle of the center monitor. The stimuli were presented for 750 ms and, following a 200 ms blank period, the observer had a 2 s interval in which to respond with a button press. Auditory feedback indicated if the observer was correct, incorrect, or did not respond, and this was followed by a 300 ms delay before presentation of the next trial.

Stimuli

The stimuli were drifting compound gratings consisting of 3 superposed sinusoids with spatial frequencies of $\frac{1}{4}$, $\frac{1}{3}$, and $\frac{1}{2}$ cyclesdeg-1. The starting phase of each component grating was randomized from trial to trial so that a trivial spatial changing-phase cue could not be used to do the task. The contrast of the components was scaled to yield a maximum Michelson contrast of 50% for the compound grating on each trial. The gratings were windowed with a spatial Gaussian function with a space constant (σ) of 3 degrees and truncated at $\pm 3.5\sigma$. Due to the unconventional geometry of the display, the spatial numbers (and the speeds to follow) are slight approximations, but this does not affect the experimental comparisons of interest.

Baseline velocities (5, 10, 20, 30, 40 degs-1) were determined roughly by the range of speeds seen in the peripheral vision by a person, walking (1.4 ms-1) down a hallway (Browning & Kram, 2005; Mohler et al., 2007) or, by extension, a moderately dense wooded area. An individual walking at a normal pace through a 6 foot (182.88 cm) wide hallway, no closer than 1.5 feet from a wall, would experience speeds ranging from 25-55 degs-1. Considering that walking through many environments (such as wooded areas with an average tree spacing greater than 6 feet) would

generate slower velocities, we included 5 and 10 degs-1 baseline velocities. These additional velocities also allowed for comparison with other studies in the literature.

The velocities shown involved temporal frequencies that were within hardware refresh rate limits. If, due to the staircase, the maximal velocity (60 degs-1) was reached more than 5 times in a run, that run was discarded (however this only occurred in initial practice sessions).

Procedure

Before performing the main experiments, we mapped the visual fields of our observers using the same apparatus and monitor configuration described above. This was necessary to ensure that our "monocular" stimuli were placed exclusively in the monocular visual fields, including conditions requiring two stimuli to fit in the same monocular field. Each eye was tested in a separate perimetry session. Observers were instructed to respond if they saw the stimulus (a white circle 15 pixels in diameter) by pushing a button; no response correspondingly indicated that the stimulus was not visible. Figure A.2 shows the resulting visual field of one observer, with the stimulus locations shown by the black circles. For two of the observers, the "monocular" stimuli were in the monocular fields when centered on the lateral monitors (as shown). For the third observer, the stimuli were displayed 5 degrees lower.

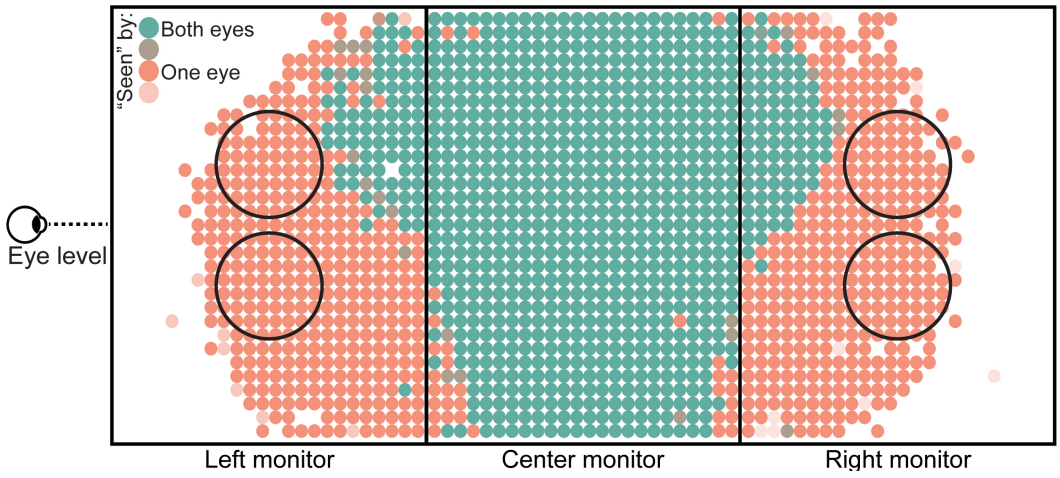

Figure A.2: Perimetry visualization from the naive observer confirmed stimuli placement exclusively in the monocular fields. Observer visual field was estimated by superimposing the left and right eye perimetry results. Teal areas show locations in which the subject reported seeing the stimulus in both left and right eye ("Both eyes"). Areas of blended teal and orange specify locations in which the observer provided mixed responses as to if the stimulus was seen with one eye or both. Orange areas show locations in which the observer reported seeing the stimulus in only one eye ("One eye"). Lighter orange areas mark locations that the observer reported seeing the stimulus at that location for 50% of the trials in one eye. Blank areas were not reported as seen. The black circles show the locations of one of the peripheral stimulus configurations (left and right conditions), demonstrating that our "monocular" stimuli did indeed fall only in the truly monocular fields. Dashed line and eye show the elevation of eye level, approximately at the midline of the monitors.

All observers completed between 2 and 12 full length practice sessions to become familiarized with the task prior to participating in the main experiment. Practice sessions continued until performance stabilized. The observer with little psychophysics experience was monitored during practice sessions in order to confirm correct eye/head position. These practice sessions were identical to the experiment sessions and averaged to 720 trials per session.

There were 3 basic experimental conditions, in which the 2 gratings were

either: 1) both within the central binocular visual field (separated either vertically or horizontally, Figure A.3C,D); 2) both within the same monocular visual field (either to the far right or far left, separated vertically, Figure A.3A); or 3) distributed across the monocular visual fields (separated horizontally, Figure A.3B). We used both vertically and horizontally separated stimuli in the central field so that each peripheral monocular condition could be paired with a central condition for which the grating patches themselves differed only in eccentricity. Figure A.3 shows the directions tested specifically in Experiment 1, however the locations shown describe the experimental conditions tested for all experiments.

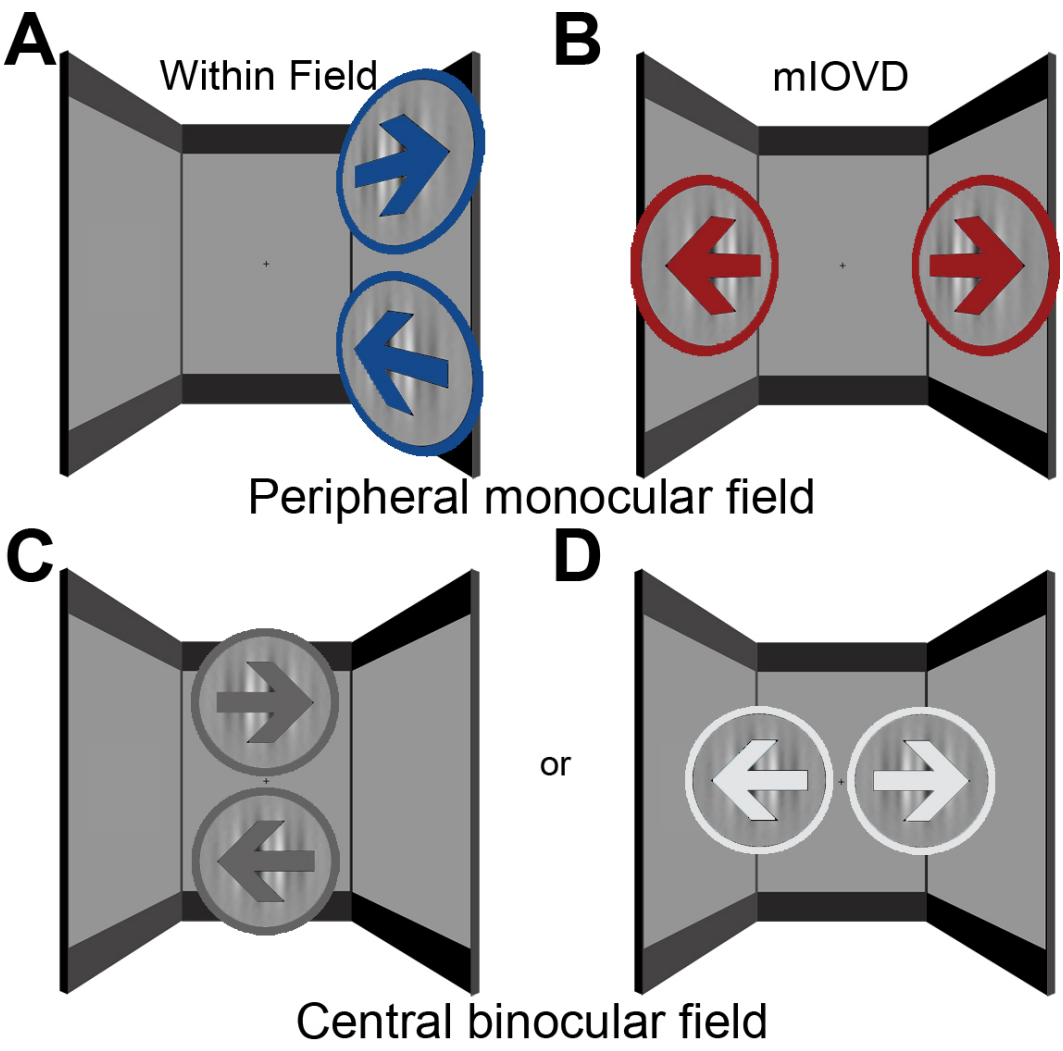

Figure A.3: Speed discrimination conditions in Experiment 1 consisted of spatially separate, oppositely drifting gratings presented simultaneously. Arrows show the drift direction of the gratings. In the monocular "Within Field" condition (A), stimuli were presented exclusively within a single monocular field of the observer. The "mIOVD" condition stimuli (B) were presented separately in each monocular field. The "Central Binocular Field" condition stimuli (C,D) were presented in the central area of the observer's binocular field. The stimuli in this condition are vertically (C) or horizontally (D) offset across the fixation point. Note that Figure A.3 and the other similar figures are not to scale; in the horizontally offset "Central Binocular Field" condition (D), for example, the stimuli were entirely on the center monitor.

Given that the gratings within the same monocular field could be either on the left or the right, and the central binocular stimuli could be separated either horizontally or vertically, the 3 basic conditions actually yielded 5 total stimulus location combinations. These location combinations were tested in separate sessions. Observers ran at least 2 sessions for each stimulus location combination. Thus each observer completed a total of 10 or more sessions for each of the experiments described below.

Performance for baseline velocities was determined from 4-8 staircases for each observer. No more than 2 staircases for a baseline velocity were tested in a single session. A staircase terminated when either 6 reversals or 100 trials were collected. With each reversal, the step size of the staircase decreased slightly. Observers took breaks during sessions as needed.

The velocity difference yielding 79% correct performance was estimated with a 3 down, 1 up staircase. Threshold was defined as the average of the velocity differences for the last five trials of each staircase. We describe performance using Weber fractions (speed discrimination threshold divided by baseline speed). The results shown were determined by averaging the Weber fractions calculated for every staircase for all observers. Uncertainty was estimated using bootstrapping methods (resampling 10,000 times with replacement); as the performance across conditions was very similar for all observers, thresholds were resampled for each condition without regard to observer identity. Unless otherwise indicated, error bands indicate 1 standard error of the mean (i.e. the central 68% of the sampling distribution).

A.3 Results

Experiment 1

In this experiment, we compared speed discrimination performance within a single eye's monocular field versus across both eye's monocular fields. We reasoned that if mIOVDs are processed in a privileged fashion, observers should be better at speed discrimination when the two moving patches were separated across the left and right eyes, as compared to within a single eye.

Methods

In the monocular "Within Field" condition, the 2 gratings were both presented within the same monocular field (left or right). Gratings were placed vertically relative to each other to allow constant stimulus size while maximizing use of the monocular field (see Figure A.3A; also Figure 2). The gratings were vertical (i.e. horizontal contrast energy) and were presented simultaneously, drifting horizontally in opposite directions. The left and right monocular fields were tested in separate sessions.

In the "mIOVD" condition, the 2 gratings were located in separate monocular fields (Figure A.3B). The centers of the gratings were located at an eccentricity of 70 degrees. The gratings had an approximate radius of 10.5 degrees. The 2 gratings were presented simultaneously, and drifted horizontally in opposite directions within their Gaussian spatial envelopes.

Finally, two additional "Central Binocular Field" conditions (Figure A.3C,D) tested gratings in the central area of the binocular field, offset either vertically or horizontally across the fixation point. In one condition, the 2 stimuli were placed side by side about the fixation point in the center display. In the other, the stimuli were placed vertically about the fixation point. These central binocular conditions provided straightforward baselines for comparison to the peripheral monocular field conditions in that each monocular stimulus condition differed from its binocular counterpart only in eccentricity. There was no reason to suspect, however, that performance in the two binocular conditions would differ greatly, and this proved to be the case in Experiment 1 (a slight difference was seen in Experiment 2, discussed later).

Results

If interocular velocity differences extracted between the two monocular portions of the visual field are processed in some privileged way, observers should have greater sensitivity to speed differences (i.e. lower Weber fractions) in the "mIOVD" condition compared to the monocular "Within Field" condition. In Figure A.4A, the gray curve shows the thresholds measured centrally. The blue curve shows the thresholds that result when we increased the stimulus eccentricity, either to the left or to the right, moving the stimuli into an exclusively monocular portion of the visual field

(monocular "Within Field" condition). Not surprisingly, sensitivity gets worse at every baseline speed. The bands on the plots show bootstrapped 68% confidence intervals. The red curve shows the thresholds we obtained when we again increased the stimulus eccentricity, but this time in opposite directions such that the two gratings occupied opposite monocular fields on either side of the head ("mIOVD" condition). This increase in eccentricity also yielded an increase in thresholds but, crucially, the increase was much less than for the within-field data.

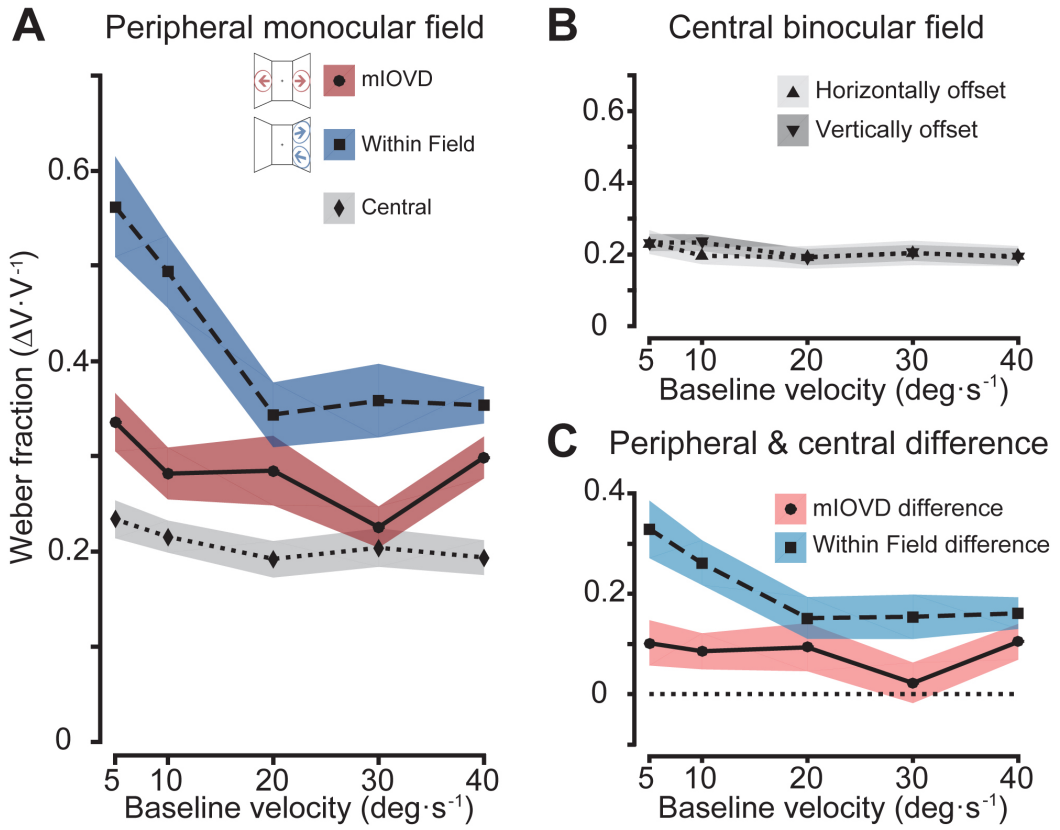

Figure A.4: Speed discrimination thresholds for Experiment 1. A: Sensitivity for speed differences within a single monocular field ("Within Field", blue) and for the "mIOVD" condition (red). Average sensitivity for the "Central Binocular Field" is shown in gray. Icons illustrate stimulus configurations. B: Sensitivity to speed differences in the "Central Binocular Field" condition with vertically offset gratings (dark gray) and with horizontally offset gratings (light gray). C: Difference in average thresholds for the peripheral monocular field and the central binocular field conditions. Monocular "Within Field" sensitivity difference is shown in light blue and "mIOVD" sensitivity difference is shown in pink. The colored regions indicate the bootstrapped 68% confidence intervals.

The dark gray curve in Figure A.4B shows the discrimination thresholds for two grating patches separated vertically in the central binocular visual field as a function of pedestal speed. The light gray curve in Figure A.4B shows the

central binocular data when the two gratings were separated horizontally instead of vertically; this change in configuration had no discernible effect on threshold (the two gray curves are the same within measurement error). Put another way, when the stimuli moved from the central visual field to the far monocular periphery, relative discrimination performance was actually better when the two gratings to be discriminated were on opposite sides of the head than when they were in the same visual field.

To quantify the drop in performance when the stimulus eccentricity increased, we calculated the difference between the peripheral monocular field sensitivity and the central binocular field sensitivity. Specifically, we took the difference in sensitivity for the monocular "Within Field" condition and compared it to the central binocular (vertically offset) condition sensitivity (Figure A.4C, light blue). For the "mIOVD" condition, a similar difference in sensitivity was estimated by comparing to the horizontally offset central binocular condition (Figure 4C, pink). As there were negligible differences between the two central binocular field conditions, these differences simply recapitulate the differences between "mIOVD" and monocular "Within Field" conditions seen in Figure A.4A. But these central binocular conditions are important for testing whether the relative directions of the gratings affect sensitivity, independent of their locations in various monocular field locations (an issue that becomes more important in Experiment 2).

Discussion

Observers were better at monocular field speed discrimination when the speeds were presented in separate eyes rather than the same eye. We note that this effect was observed using stimuli that were consistent with local velocity vectors seen by an observer walking forward or backward either at different average distances from two surfaces, or while turning slightly while walking between the two surfaces. Moreover, these same basic stimuli are known to cause bees to change their flight to null the velocity difference between the lateral visual fields (Srinivasan et al., 1991).

Experiment 2 and 3 are designed to further test the theory that speed discrimination is enhanced for self-motion, and also serves to rule out other effects like simple crowding.

Experiment 2

Experiment 2 examined speed discrimination for drifting gratings that move in the same direction, as schematized in Figure A.5. Given the unique viewing geometry of these experiments, a brief aside on terminology is warranted here. By "move in the same direction", we mean that both gratings drifted to the left or both to the right when viewed in the central binocular field (Figure A.5C,D). Note, however, that when the eccentricity of the components was increased, one to the left and one to the right as in the "mIOVD" condition (Figure A.5A), one component ended up drifting forward and the other drifting backward in head-centered coordinates. Nonetheless, they both still drifted in the same relative direction – left-to-right – in each eye's visual field. A comparison of Figure A.3 ("opposite direction") and Figure A.5 ("same direction") should make this point clear.

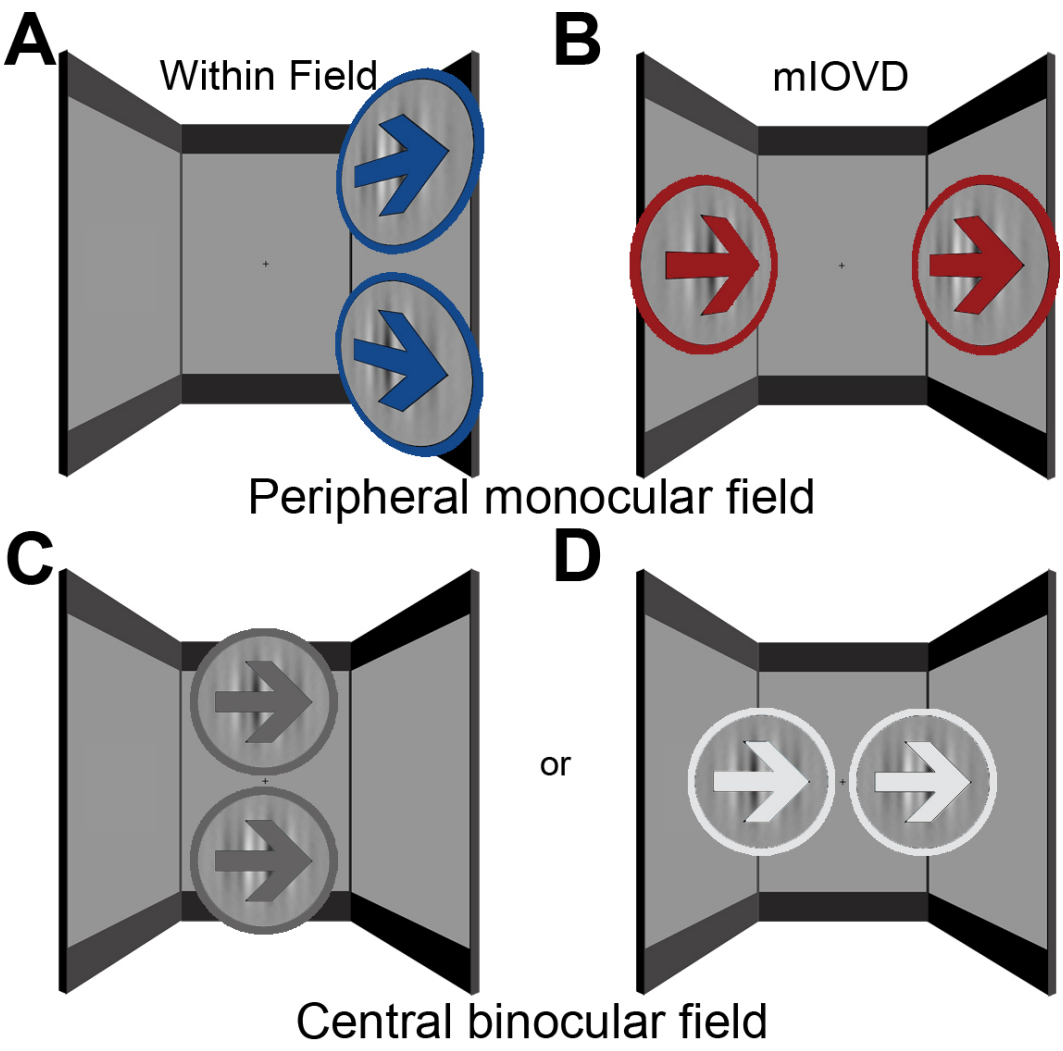

Figure A.5: Experiment 2 speed discrimination conditions in which simultaneously presented gratings drifted in the same relative direction. Same general format as described for Figure A.3. Arrows indicate the relative drift directions of the gratings (same horizontal drift direction for Experiment 2's vertically oriented gratings).

The "same direction" motion used in this experiment is not really consistent with any common ecologically valid situation for primates with mobile eyes. It could, however, easily occur in insects (or any animal with fixed eyes) that were turning in place but were slightly closer to one of two parallel walls. This experiment will thus help test whether mIOVD in humans is confined to velocity differences commonly

encountered by animals with mobile frontal eyes, or whether it generalizes to other kinds of velocity differences.

Methods

The stimuli and procedures were identical to those in Experiment 1, with one exception: rather than the stimuli moving in opposite horizontal directions (relative to each eyes field of view), stimuli moved in the same direction (Figure 5). Like Experiment 1, the observer was instructed to indicate which stimulus was moving the fastest.

Results

Just as in Experiment 1, when the gratings were moved from the central binocular visual field into the same peripheral monocular field, thresholds were elevated (Figure A.6A). For the slowest speeds, the peripheral thresholds were lower for these same-direction stimuli than they were for the opposite-direction stimuli of Experiment 1.

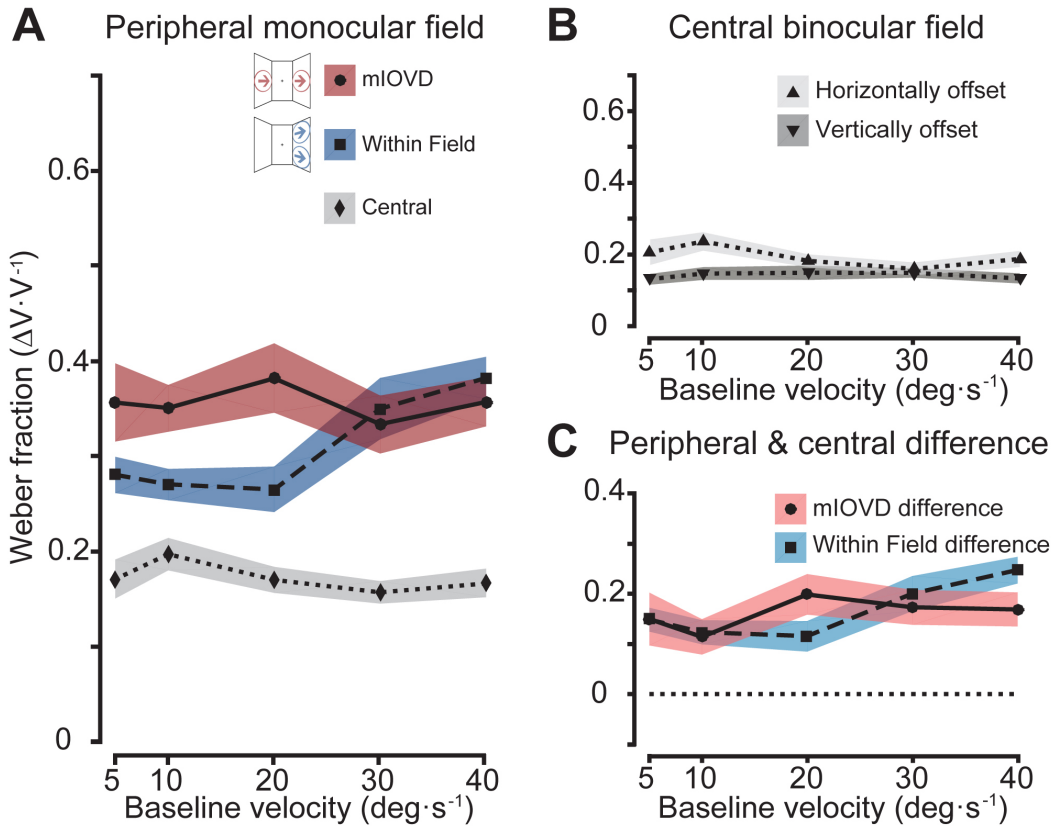

Figure A.6: Speed discrimination thresholds for Experiment 2. A: Sensitivity for speed differences within either the left or right monocular field (blue) and for the "mIOVD" stimulus (red). Averaged central binocular sensitivity (gray) is shown for reference. B: Speed discrimination sensitivity for the central binocular field with vertically offset gratings (dark gray) and for the central binocular field with horizontally offset gratings (light gray). C: The difference in average thresholds of the peripheral monocular field conditions (monocular "Within Field" in light blue, "mIOVD" in pink) and the equivalent central binocular field condition. Plotting conventions as in Figure A.4.

Unlike Experiment 1, however, the "mIOVD" thresholds (Figure A.6A, red) were elevated by an amount that surpassed the monocular "Within Field" thresholds (Figure 6A, blue) at lower speeds. The peripheral monocular field conditions were compared to the equivalent central binocular field condition (to quantify the

drop in sensitivity as stimulus eccentricity increased). As shown in Figure A.6C, both peripheral monocular field conditions actually showed a similar drop in performance. Thus, in Experiment 2, when the gratings were moved into the far periphery, velocity discrimination suffered by a similar amount, regardless of whether the grating patches were moved to opposite monocular fields ("mIOVD"), or to the same peripheral field.

Another way to look at these data is to compare the horizontally-separated conditions from Experiment 1 (Figure A.4) with those from Experiment 2 (Figure A.6). When horizontally-separated gratings were viewed in the central binocular field, velocity discrimination was the same regardless of whether the gratings were moving in the same or opposite directions. When these gratings moved into the far periphery on opposite sides of the head, however, velocity discrimination was better when the gratings drifted in opposite directions (when gratings were consistent with forward or backward observer motion).

Discussion

In Experiment 1, observers were better at speed discrimination in the far periphery when the stimuli to be discriminated were actually presented in separate eyes on opposite sides of the head rather than to the same eye in relatively close proximity to one another. In Experiment 2, when the relative directions of gratings were changed, there was no longer any advantage to comparing speeds across the monocular fields as opposed to comparing speeds within a monocular field.

We also observed higher sensitivity to speed differences in the vertical central binocular condition (compared to the horizontal central binocular condition). This is possibly due to the stimulus configuration in this condition in which observers could use changing relative phase information when comparing speeds. As humans are thought to be largely phase blind in the periphery (Bennett & Banks, 1987; Rentschler & Treutwein, 1985; Stephenson & Braddick, 1983), this information is unlikely to have been used in the monocular "Within Field" condition. If this is the case, the threshold difference between these conditions may actually be smaller than our results show.

It is perhaps worth briefly pausing to consider the difference between Experiment 1 and Experiment 2, both in terms of the stimuli per se as well as their

ecological validity. With respect to the stimuli themselves, the difference between the two experiments seems trivial; the only difference is a reversal of the velocity sign of one of the two gratings. Ecologically, however, this reversal makes a profound difference when the stimuli are placed in opposite monocular fields in the far periphery. For a primate (having mobile frontally-located eyes), the stimuli from Experiment 1 are rough approximations of what is experienced when walking along a path in a forest or along a hallway. By flipping the direction of one grating, however, one renders the stimuli consistent with what would be experienced by rotating the head in place with the eyes fixed in the head. But because optokinetic nystagmus (OKN) is reflexive and compulsory, this motion would be present only during the fast phase of the OKN, during which saccadic suppression would presumably be inhibiting visual processing. It is true that if an observer fixated an object that was anchored with respect to the head and then rotated in place, the stimulation would be not unlike ours in the present experiment, but this seems a rather contrived situation (e.g., holding a pen at arms length in front of the face, fixating it, and then spinning about). Veering trajectories could also, in principle, create the same sort motion under consideration. If an observer walked down a hall veering to and fro in an "S" shaped trajectory, then if (and when) the radius of a turn was shorter than the distance to a surface patch on one wall but longer than that to the other, such motion would be generated but only if the eyes were fixed in the head and the head always pointed in the immediate direction of travel. A brief walk down the hall should convince most readers that is not what occurs. Of course, many scenarios in addition to these can be considered, but we think it is reasonable to claim that stimulation similar to the mIOVD stimulus in Experiment 1 (Figure A.3B) is rather commonly encountered by humans, whereas stimulation similar to the mIOVD stimulus in Experiment 2 (Figure A.5B) is not.

To further this line of argument – that mIOVD is used in humans only for stimuli that are ecologically valid – we tested whether the effect (compared to velocity discrimination within a monocular field) was seen in speed discrimination for directions that were completely inconsistent with any type of self-motion.

Experiment 3

To remove any possible ecologically-valid motion pattern from stimuli, we used gratings that drifted in orthogonal directions. In no (survivable) situation would an observer view these velocity vectors during self-motion in a roughly rigid environment. We anticipated that mIOVD performance would be no better than speed discrimination within a monocular field in this scenario.

Methods

The stimuli in this experiment were identical to those used in the previous two experiments, except that one of the two gratings drifted vertically as shown in Figure A.7. The only way a human observer could experience this type of motion naturally would be during a fleeting moment of consciousness while the structural integrity of the head was being severely compromised. We speculate that psychophysical reports from such an observer would be difficult if not impossible to obtain. In this experiment, we tested only 5, 10, and 20 degs-1 baseline velocities as these were the most diagnostic speeds in the first two experiments and because the faster speeds resulted in unmeasurably high thresholds.

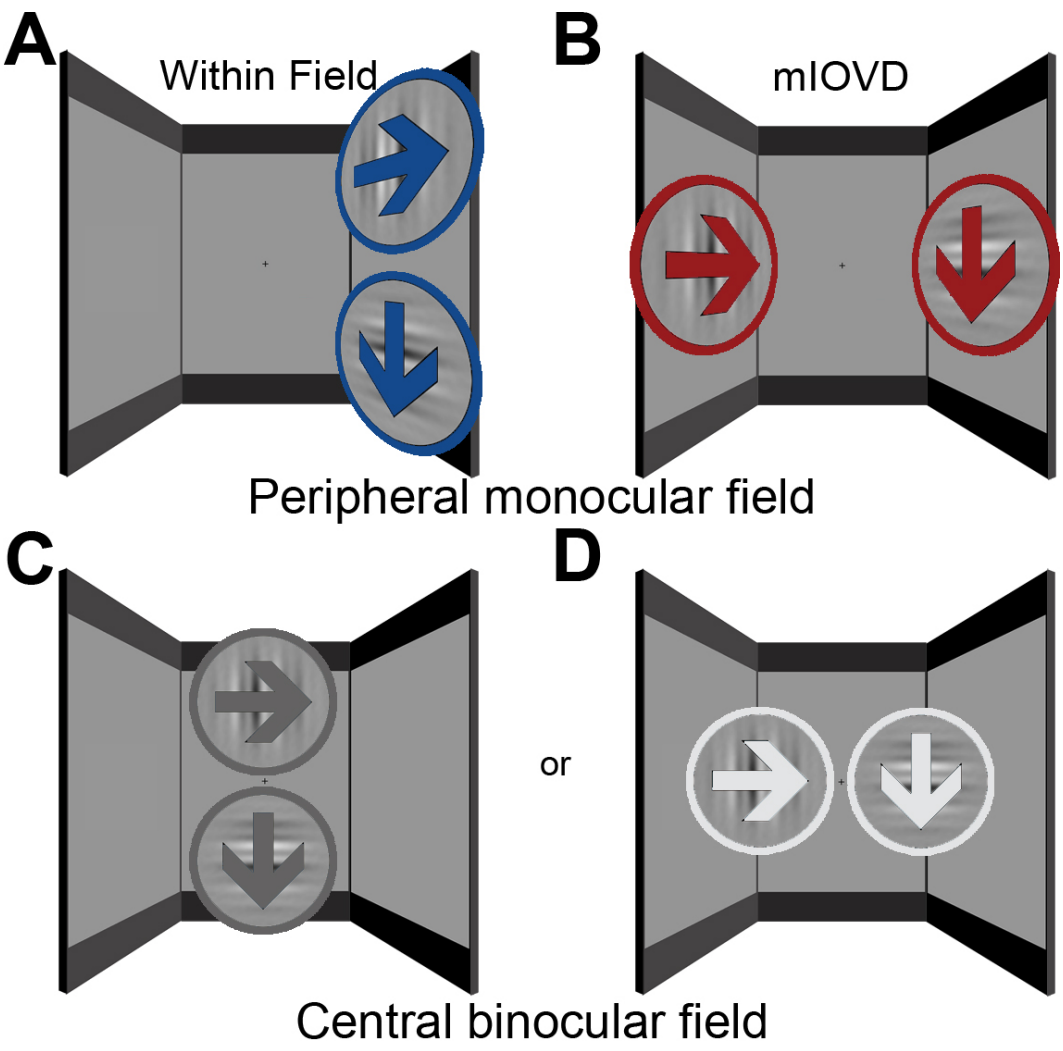

Figure A.7: Speed discrimination conditions in Experiment 3, in which gratings drifted orthogonally. Conventions as in Figures A.3 and A.5. For each condition, one grating was oriented vertically and drifted horizontally, the second grating was oriented horizontally and drifted vertically. See text for details.

Results

In Experiment 3, we found little difference between the central binocular field conditions (Figure A.8B, gray) so, as in Experiment 1 it did not matter if the patches flanked the fixation point vertically (dark gray) or horizontally (light gray). Thresh-

olds were elevated when we moved the stimuli into the periphery, either into the same monocular field (Figure A.8A, blue) or into monocular fields on the opposite sides of the head ("mIOVD"; Figure A.8A, red). Crucially, however, this threshold elevation was the same for the two conditions. Thus, like in Experiment 2, but unlike in Experiment 1, there was no less of a drop in performance for mIOVD compared to within a monocular field speed comparisons.

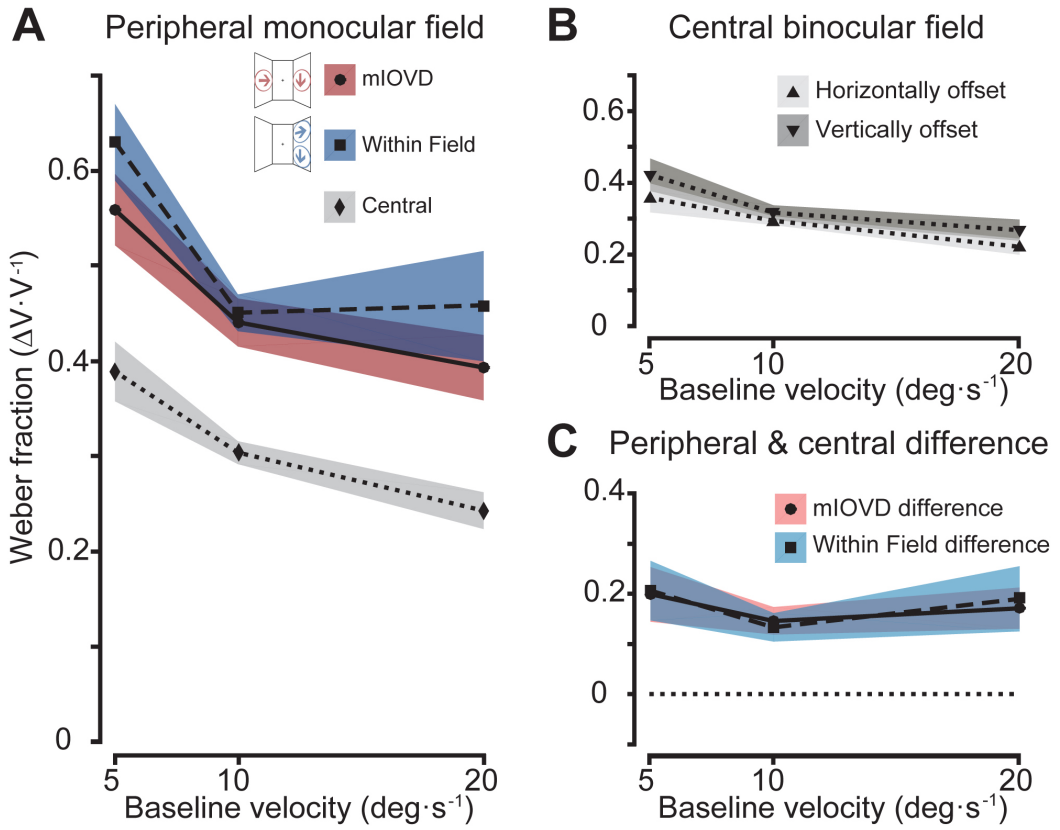

Figure A.8: Speed discrimination thresholds for orthogonal motion. A: Sensitivity for speed differences within a single monocular field ("Within Field", blue) and for between both monocular fields ("mIOVD", red). Averaged central binocular sensitivity (gray) is shown for comparison. B: Speed difference sensitivity in the central binocular field with vertically offset gratings (dark gray) and for the central binocular field with horizontally offset gratings (light gray). C: The difference in average thresholds measured the drop in performance of the peripheral monocular field conditions (monocular "Within Field" in light blue, "mIOVD" in pink) after the equivalent central binocular field condition was deducted. Plotting conventions as in Figures A.3 and A.6.

Discussion

As predicted, without any pattern of ecologically-valid self-motion, velocity discrimination between the monocular fields ("mIOVD") and within a single monocular field ("Within Field") both have an equal drop in performance with stimulus eccentricity. This result is consistent with our expectation that – for velocity vectors inconsistent with ego-motion – mIOVDs would have no advantage in speed difference sensitivity, compared to those made in a single monocular field.

A.4 General Discussion

Our results support the hypothesis that inter-monocular speed comparisons are processed in a privileged fashion. The sensitivity of velocity discrimination of our observers was highly dependent upon the relative directions of motions to be compared, and discrimination was best when the velocities seen in the two monocular visual fields were consistent with forward or backward self-motion. Figure 9 summarizes the results for all 3 experiments. The figure shows the difference in threshold between the central binocular and peripheral monocular conditions for both the case in which 1) the gratings were offset vertically when viewed centrally, and then were both moved into the same eye's monocular field when viewed peripherally (monocular "Within Field difference"; light blue) and 2) the gratings were offset horizontally when viewed centrally, and then moved into separate monocular fields on opposite sides of the head when viewed peripherally ("mIOVD difference"; pink). For the two experiments in which the peripheral motion was not ecologically valid or common – Experiments 2 and 3 (Figure A.9B,C) – thresholds were not statistically different. For Experiment 1 however (Figure A.9A), in which the peripheral stimulus was consistent with self-motion, relative thresholds for stimuli on opposite sides of the head were actually much better than for stimuli immediately adjacent to one another.

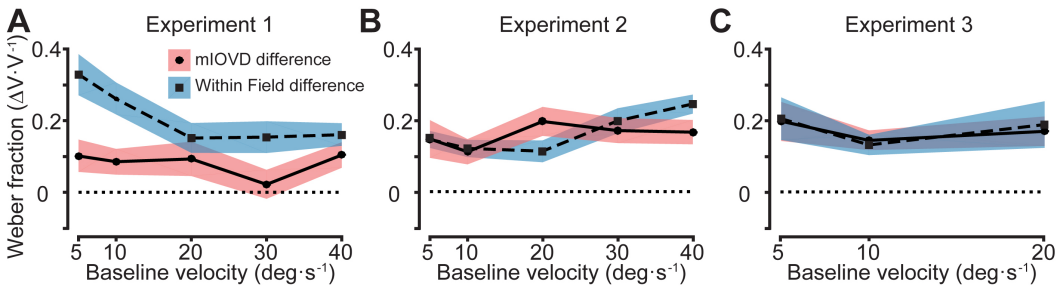

Figure A.9: Speed discrimination threshold differences between peripheral and central vision more pronounced in Experiment 1, compared to Experiments 2 & 3. Each point shows the difference between the peripheral monocular and central binocular thresholds for equivalent stimulus configurations for all 3 experiments. Note that only in Experiment 1 (A) are the relative "mIOVD difference" thresholds much lower than their "Within Field difference" counterparts. Error bars show bootstrapped 68% CIs.

It is interesting to contemplate why the inter-monocular advantage was only evident for monocular motion consistent with forward/backward self-motion (Experiment 1, Figure A.9A), and not self-motion due to spinning or sharp veering (Experiment 2, Figure A.9B). This effect was seen for all observers. Note that eye movements would have been the most useful in the "Within Field" monocular condition, where observers could have simply reduced the eccentricity of both stimuli at once since eye movements were not tracked. Eye movements therefore do not explain why performance was lower in this condition. As mentioned above, the optic flow produced by rotating in place or making sharp turns while walking is disrupted by head and eye movements (Lappe et al., 1999). Thus, despite the seemingly-trivial difference between the stimuli in Experiments 1 and 2, the former approximates a common situation in the environment (walking between two things) while the latter represents a situation that is uncommon at best (e.g., spinning in place with the gaze direction fixed relative to the head). Or consider moving in a "zig zag" trajectory along a hallway (avoiding obstacles on the ground, say). Even if gaze could be held straight-ahead with respect to the head and the head could be held straight-ahead with respect to the body, the retinal projections of texture on either wall would undergo transformations over and above changes in velocity as both the angle to and the distance from the walls continuously changed. Again,

one could explore various environmental scenarios, but the bottom line is that the stimuli in Experiment 1 roughly approximate something seen daily, whereas those of Experiment 2 do not. That having been said, we did do one further test – Experiment 3 (Figure A.9C) – in which we used stimuli that could not be created solely by self-motion, and these results were in accord of those from Experiment 2. Our main empirical conclusion is thus that human observers are more sensitive to patterns of peripheral motion across the eyes that could potential help in guiding self-motion. What follows hereafter is even more speculative, but which we hope frame hypotheses for future investigation.

We propose that interocular velocity differences are used for processing both object and ego motion. When used for estimating the 3D direction of objects, the requisite eye-specific velocity signals come from corresponding locations within the overlapping binocular field from both eyes (the conventional "IOVD"). When used for estimating the direction an observer is moving, these IOVD signals come from the far peripheral (temporal) portions of each eye's view, including large portions of completely monocular visual fields (which we have termed here the "mIOVD"). The relevance of IOVDs to navigation is perhaps best understood in the ecological context of optic flow. If one considers an observer that fixates straight ahead while moving forward, the resulting radial flow field would contain velocities which emanate from a common central point, the focus of expansion (FOE). Although classical conceptions of optic flow (Gibson, 1950; Koenderink, 1986) are effectively cyclopean (i.e. a single optic flow field is considered), our results suggest that it is important to appreciate that both eyes receive optic flow, and that the spatial patterns of velocities differ in lawful ways between the two eyes due to the relative positions of the eyes in the head (as well as the occlusions that features of the head and face pose to each eye's view).

In animals with lateral eyes, scene structure to the sides of these animals projects primarily to one eye or the other, and in the simple case of an animal moving forward, these velocities both "point" backwards, which means that the left eye receives leftward oculocentric velocities and the right eye receives rightward oculocentric velocities. The differential directions and speeds are directly indicative of the animal's motion relative to the scene. Although it is straightforward to think about this pattern as involving a comparison between the far lateral portions of the visual field, it may be more appropriate to consider these comparisons as going on

between the two eyes.

One possible explanation for these results is that some of the increase in thresholds seen when both gratings are placed in the same monocular visual field is due to crowding or some other form of spatial interference. To confirm crowding effects did not produce these results, a control experiment we piloted included additional flanker gratings directly above the "mIOVD" stimuli that were irrelevant to the speed discrimination task. These flankers had the same spatial properties as the test stimuli, but their speeds ranged between the test and reference speeds in a given trial, and did not give the observer any extra information to perform the task (i.e. they were completely task-irrelevent). Performance was unaffected, mitigating our initial concern of crowding being an issue as we began the experiments. More direct evidence against crowding comes from considering the different stimulus directions examined throughout Experiments 1 through 3 (see e.g., Levi et al. 2002; Bex et al. 2003). Specifically, if crowding was a key factor in Experiment 1 (in which sensitivity was higher for motions in different eyes compared to motions in the same monocular field), a similar difference would have been present in the results of the other experiments, but it was not (although the exact amounts of interaction between target and flankers might be tuned for direction and speed, Bex & Dakin 2005).

In summary, comparisons of velocities between the monocular fields might be supported by a mechanism related to the IOVD computations currently studied in the context of 3D motion perception (Shioiri et al., 2009; Brooks, 2002; Fernandez & Farell, 2006; Czuba et al., 2010; Rokers et al., 2008). It is tempting to speculate that this mechanism, which computes IOVDs for a single object, and thus operates in the same portion of the visual field in the two eyes, is perhaps derived from an older mechanism that compares velocities between opposite sides of the head, and that this mechanism indeed developed as the eyes migrated forward in the head. These findings reinforce the importance of eye-specific motion signals and suggest it may be possible to integrate interocular computations across multiple visual domains and species. More generally, we have established that, in some cases, the brain is better at comparing stimuli presented to different eyes on opposite sides of the head than it is to comparing adjacent stimuli in the same eye. The fact that these cases correspond to ecologically valid motion, whereas other cases we explored do not, is certainly intriguing and begs further investigation.

ACKNOWLEDGMENT

This work was supported by NIH NEI R01-EY020592 to A. C. Huk, L. K. Cormack & A. Kohn (Albert Einstein College of Medicine), NSF GRFP DGE-1110007 to K. Bonnen, and Harrington Fellowship to K. Bonnen. We also thank Sung Jun Joo for constructive feedback and discussion. The authors declare no competing financial interests.

References

Ackermann, J. F., & Landy, M. S. (2010). Suboptimal Choice of Saccade Endpoint in Search with Unequal Payoffs. *Journal of Vision*, *10*(7), 530–530.

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

Albrecht, D. G., & Geisler, W. S. (1991). Motion Selectivity and the Contrast-Response Function of Simple Cells in the Visual-Cortex. *Visual neuroscience*, *7*(6), 531–546.

Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, *52*(6), 1106–1130.

Albright, T. D., Desimone, R., & Gross, C. G. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. *Journal of Neurophysiology*, *51*(1), 16–31.

Allen, B., Haun, A. M., Hanley, T., Green, C. S., & Rokers, B. (2015). Optimal Combination of the Binocular Cues to 3D MotionBinocular Cues to 3D Motion. *Investigative ophthalmology & visual science*, *56*(12), 7589–7596.

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, *7*(5), 358–366.

Bacon, J. P., & Murphey, R. K. (1984). Receptive fields of cricket giant interneurones are related to their dendritic structure. *The Journal of physiology*, *352*, 601–623.

Badcock, D. R., & Schor, C. M. (1985). Depth-increment detection function for individual spatial channels. *Journal of the Optical Society of America. A, Optics and image science*, *2*(7), 1211–1216.

Baddeley, R. J., Ingram, H. A., & Miall, R. C. (2003). System identification applied to a visuomotor task: near-optimal human performance in a noisy changing task. *The Journal of Neuroscience*, *23*(7), 3066–3075.

Baker, P. M., & Bair, W. (2016). A Model of Binocular Motion Integration in MT Neurons. *Journal of Neuroscience*, *36*(24), 6563–6582.

Banks, W. (1970). Signal detection theory and human memory. *Psychological Bulletin*, *74*(2), 81–99.

Barendregt, M., Dumoulin, S. O., & Rokers, B. (2014). Stereomotion scotomas occur after binocular combination. *Vision research*.

Baria, A. T., Maniscalco, B., & He, B. J. (2017). Initial-state-dependent, robust, transient neural dynamics encode conscious visual perception. *PLoS Computational Biology*, *13*(11), e1005806.

Barlow, H. B. (1957). Increment thresholds at low intensities considered as signal/noise discriminations. *The Journal of physiology*, *136*(3), 469–488.

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(5), 1672–1677.

Bell, A. J., & Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision research*, *37*(23), 3327–3338.

Bennett, P. J., & Banks, M. S. (1987). Sensitivity loss in odd-symmetric mechanisms and phase anomalies in peripheral vision. *Nature*, *326*(6116), 873–876.

Berniker, M., & Kording, K. (2008). Estimating the sources of motor errors for adaptation and generalization. *Nature Neuroscience*, *11*(12), 1454–1461.

Bex, P. J., & Dakin, S. C. (2005). Spatial interference among moving targets. *Vision research*, *45*(11), 1385–1398.

Bex, P. J., Dakin, S. C., & Simmers, A. J. (2003). The shape and size of crowding for moving targets. *Vision research*, *43*(27), 2895–2904.

Bhagavatula, P. S., Claudianos, C., Ibbotson, M. R., & Srinivasan, M. V. (2011). Optic flow cues guide flight in birds. *Current biology : CB*, *21*(21), 1794–1799.

Bolkan, S. S., Stujenske, J. M., Parnaudeau, S., Spellman, T. J., Rauffenbart, C., Abbas, A. I., ... Kellendonk, C. (2017). Thalamic projections sustain prefrontal activity during working memory maintenance. *Nature Neuroscience*, *20*(7), 987–996.

Boly, M., Balteau, E., Schnakers, C., Degueldre, C., Moonen, G., Luxen, A., ... Laureys, S. (2007). Baseline brain activity fluctuations predict somatosensory perception in humans. *Proceedings of the National Academy of Sciences*, *104*(29), 12187–12192.

Bonnen, K., Burge, J., Yates, J., Pillow, J., & Cormack, L. K. (2015). Continuous psychophysics: Target-tracking to measure visual sensitivity. *Journal of Vision*, *15*(3), 14–14.

Bonnen, K., Huk, A. C., & Cormack, L. K. (2017). Dynamic mechanisms of visually guided 3D motion tracking. *Journal of Neurophysiology*, *118*(3), 1515–1531.

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience*, *28*(1), 157–189.

Braddick, O. (1974). A short-range process in apparent motion. *Vision research*, *14*(7), 519–527.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial vision*(4), 433–436.

Bredfeldt, C. E., & Ringach, D. L. (2002). Dynamics of spatial frequency tuning in macaque V1. *The Journal of Neuroscience*, *22*(5), 1976–1984.

Brett, J. (1987). Goldilocks and the three bears (Retold and illustrated). *New York: Dodd Mead*.

Brooks, K. R. (2002). Interocular velocity difference contributes to stereomotion speed perception. *Journal of Vision*, *2*(3), 218–231.

Brooks, K. R., & Stone, L. S. (2006). Stereomotion suppression and the perception of speed: accuracy and precision as a function of 3D trajectory. *Journal of Vision*, *6*(11), 1214–1223.

Browning, R. C., & Kram, R. (2005). Energetic cost and preferred speed of walking in obese vs. normal weight women. *Obesity Research*, *13*(5), 891–899.

Brueggemann, J. (2007). The hand is NOT quicker than the eye. *Journal of Vision*, *7*(15), 54–54.

Buckner, R. L., Krienen, F. M., & Yeo, B. T. T. (2013). Opportunities and limitations of intrinsic functional connectivity MRI. *Nature Neuroscience*, *16*(7), 832–837.

Bullmore, E., Long, C., Suckling, J., Fadili, J., Calvert, G., Zelaya, F., ... Brammer, M. (2001). Colored noise and computational inference in neurophysiological (fMRI) time series analysis: Resampling methods in time and wavelet domains. *Human Brain Mapping*, *12*(2), 61–78.

Burbeck, S. L., & Luce, R. D. (1982). Evidence from auditory simple reaction times for both change and level detectors. *Perception & Psychophysics*, *32*(2), 117–133.

Burge, J., Ernst, M. O., & Banks, M. S. (2008a). The statistical determinants of adaptation rate in human reaching. *Journal of Vision*, *8*(4), 20–19.

Burge, J., Ernst, M. O., & Banks, M. S. (2008b). The statistical determinants of adaptation rate in human reaching. *Journal of Vision*, *8*(4), 20–20.

Burge, J., & Geisler, W. S. (2011). Optimal defocus estimation in individual natural images. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(40), 16849–16854.

Burge, J., & Geisler, W. S. (2014). Optimal disparity estimation in natural stereo images. *Journal of Vision*, *14*(2), 1–1.

Burge, J., & Geisler, W. S. (2015). Optimal speed estimation in natural image movies predicts human performance. *Nature communications*, *6*(1), 7900.

Burge, J., Girshick, A. R., & Banks, M. S. (2010). Visual-haptic adaptation is determined by relative reliability. *The Journal of Neuroscience*, *30*(22), 7714–7721.

Burge, J., McCann, B. C., & Geisler, W. S. (2016). Estimating 3D tilt from local image cues in natural scenes. *Journal of Vision*, *16*(13), 2–2.

Carmena, J. M., Lebedev, M. A., Crist, R. E., O'Doherty, J. E., Santucci, D. M., Dimitrov, D. F., ... Nicolelis, M. A. L. (2003). Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS biology*, *1*(2), E42.

Chopin, A., & Mamassian, P. (2012). Predictive properties of visual adaptation. *Current biology : CB*, *22*(7), 622–626.

Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., ... Shenoy, K. V. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature Neuroscience*, *13*(3), 369–378.

Clark, D. A., Fitzgerald, J. E., Ales, J. M., Gohl, D. M., Silies, M. A., Norcia, A. M., & Clandinin, T. R. (2014). Flies and humans share a motion estimation strategy that exploits natural scene statistics. *Nature Neuroscience*, *17*(2), 296–303.

Clayton, K., & Frey, B. B. (1997). Studies of Mental "Noise". *Nonlinear Dynamics, Psychology and Life Sciences*, *1*(3), 173–180.

Clifford, C. W. G., Webster, M. A., Stanley, G. B., Stocker, A. A., Kohn, A., Sharpee, T. O., & Schwartz, O. (2007). Visual adaptation: Neural, psychological and computational aspects. *Vision research*, *47*(25), 3125–3131.

Coen-Cagli, R., Kohn, A., & Schwartz, O. (2015). Flexible gating of contextual influences in natural vision. *Nature Publishing Group*, *18*(11), 1648–1655.

Cohn, T. E., & Lasley, D. J. (1974). Detectability of a luminance increment: effect of spatial uncertainty. *Journal of the Optical Society of America*, *64*(12), 1715–1719.

Cooper, E. A., van Ginkel, M., & Rokers, B. (2016). Sensitivity and bias in the discrimination of two-dimensional and three-dimensional motion direction. *Journal of Vision*, *16*(10), 5–11.

Cormack, L. K., Czuba, T. B., Knöll, J., & Huk, A. C. (2017). Binocular Mechanisms of 3D Motion Processing. *doi.org*, *3*(1), 297–318.

Craik, K. J. W. (1947). Theory of the human operator in control systems; the operator as an engineering system. *The British journal of psychology. General section*, *38*(Pt 2), 56–61.

Craik, K. J. W. (1948). Theory of the human operator in control systems; man as an element in a control system. *The British journal of psychology. General section*, *38*(Pt 3), 142–148.

Cumming, B. G., & Parker, A. J. (1994). Binocular mechanisms for detecting motion-in-depth. *Vision research*, *34*(4), 483–495.

Czuba, T. B., Huk, A. C., Cormack, L. K., & Kohn, A. (2014). Area MT encodes three-dimensional motion. *The Journal of Neuroscience*, *34*(47), 15522–15533.

Czuba, T. B., Rokers, B., Guillet, K., Huk, A. C., & Cormack, L. K. (2011). Three-dimensional motion aftereffects reveal distinct direction-selective mechanisms for binocular processing of motion through depth. *Journal of Vision*, *11*(10), 18–18.

Czuba, T. B., Rokers, B., Huk, A. C., & Cormack, L. K. (2010). Speed and eccentricity tuning reveal a central role for the velocity-based cue to 3D visual motion. *Journal of Neurophysiology*, *104*(5), 2886–2899.

DeAngelis, G. C., & Angelaki, D. E. (2012). Visual–Vestibular Integration for Self-Motion Perception.

De Coensel, B., Botteldooren, D., & De Muer, T. (2003). 1/f Noise in Rural and Urban Soundscapes. *Acta Acustica united with acustica*, *89*(2), 287–295.

Dehghani, N., Bédard, C., Cash, S. S., Halgren, E., & Destexhe, A. (2010). Comparative power spectral analysis of simultaneous electroencephalographic and magnetoencephalographic recordings in humans suggests non-resistive extracellular media : EEG and MEG power spectra. *Journal of computational neuroscience*, *4*, 167.

Denève, S., Duhamel, J.-R., & Pouget, A. (2007). Optimal sensorimotor integration in recurrent cortical networks: a neural implementation of Kalman filters. *The Journal of Neuroscience*, *27*(21), 5744–5756.

de Vries, H. L. (1943). The quantum character of light and its bearing upon threshold of vision, the differential sensitivity and visual acuity of the eye. *Physica*, *10*(7), 553–564.

Ditterich, J. (2006). Stochastic models of decisions about motion direction: Behavior and physiology. *Neural Networks*, *19*(8), 981–1012.

Dixon, W. J., & Mood, A. M. (1948). A Method for Obtaining and Analyzing Sensitivity Data. *Journal of the American Statistical Association*, *43*(241), 109–126.

Dong, D., & Atick, J. (1995). Statistics of natural time-varying images. *Network: Computation in Neural Systems*, *6*(3), 345–358.

Eckmeier, D., Geurten, B. R. H., Kress, D., Mertes, M., Kern, R., Egelhaaf, M., & Bischof, H.-J. (2008). Gaze strategy in the free flying zebra finch (Taeniopygia guttata). *PloS one*, *3*(12), e3956.

Fantz, R. L. (1963). Pattern Vision in Newborn Infants. *Science*, *140*(3564), 296–297.

Fechner, G. T. (1860). *Elemente Der Psychophysik*. Leipzig: Breitkopf und Härtel.

Fernandez, J. M., & Farell, B. (2006). A reversed structure-from-motion effect for simultaneously viewed stereo-surfaces. *Vision research*, *46*(8-9), 1230–1241.

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5), 738–743.

Fox, M. D., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2007). Intrinsic fluctuations within cortical systems account for intertrial variability in human behavior. *NEURON*, *56*(1), 171–184.

Freedman, D. J., & Assad, J. A. (2016). Neuronal Mechanisms of Visual Categorization: An Abstract View on Decision Making. *Annual Review of Neuroscience*, *39*(1), 129–147.

Fritsche, M., Mostert, P., & de Lange, F. P. (2017). Opposite Effects of Recent History on Perception and Decision. *Current biology : CB*, *27*(4), 590–595.

Fulvio, J. M., Rosen, M. L., & Rokers, B. (2015). Sensory uncertainty leads to systematic misperception of the direction of motion in depth. *Attention Perception & Psychophysics*, *77*(5), 1685–1696.

Ganguli, D., & Simoncelli, E. P. (2016). Neural and perceptual signatures of efficient sensory coding. *arXiv.org*.

Gavornik, J. P., & Bear, M. F. (2014). Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nature Neuroscience*, *17*(5), 732–737.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological review*, *96*(2), 267–314.

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision research*, *51*(7), 771–781.

Geisler, W. S., & Albrecht, D. G. (1995). Bayesian analysis of identification performance in monkey visual cortex: nonlinear mechanisms and stimulus certainty. *Vision research*, *35*(19), 2723–2730.

Geisler, W. S., Najemnik, J., & Ing, A. D. (2009). Optimal stimulus encoders for natural tasks. *Journal of Vision*, *9*(13), 17.1–16.

Gibson, J. J. (1950). *The Perception of the Visual World.*

Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological review*, *108*(1), 33–56.

Gilden, D. L., Thornton, T., & Mallon, M. W. (1995). 1/f noise in human cognition. *Science*, *267*(5205), 1837–1839.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*(1), 535–574.

Goris, R. L. T., Movshon, J. A., & Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nature Neuroscience*, *17*(6), 858–865.

Götz, K. G. (1968). Flight control in Drosophila by visual perception of motion. *Kybernetik*, *4*(6), 199–208.

Graf, A. B. A., Kohn, A., Jazayeri, M., & Movshon, J. A. (2011). Decoding the activity of neuronal populations in macaque primary visual cortex. In *Nature neuroscience* (pp. 239–247).

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.

Harris, J. M., Nefs, H. T., & Grafton, C. E. (2008). Binocular vision and motion-in-depth. *Spatial vision*, *21*(6), 531–547.

Harvey, L. O. (1986). Efficient estimation of sensory thresholds. *Behavior Research Methods, Instruments, & Computers*, *18*(6), 623–632.

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in cognitive sciences*, *14*(1), 40–48.

Hawken, M. J., Shapley, R. M., & Grosof, D. H. (1996). Temporal-frequency selectivity in monkey visual cortex. *Visual neuroscience*, *13*(3), 477–492.

He, B. J. (2011). Scale-free properties of the functional magnetic resonance imaging signal during rest and task. *The Journal of Neuroscience*, *31*(39), 13786–13795.

He, B. J. (2014). Scale-free brain activity: past, present, and future. *Trends in cognitive sciences*, *18*(9), 480–487.

He, B. J., Snyder, A. Z., Zempel, J. M., Smyth, M. D., & Raichle, M. E. (2008). Electrophysiological correlates of the brain's intrinsic large-scale functional architecture. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(41), 16039–16044.

He, B. J., Zempel, J. M., Snyder, A. Z., & Raichle, M. E. (2010). The temporal structures and functional significance of scale-free brain activity. *NEURON*, *66*(3), 353–369.

Hesselmann, G., Kell, C. A., Eger, E., & Kleinschmidt, A. (2008). Spontaneous local variations in ongoing neural activity bias perceptual decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(31), 10984–10989.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, *148*, 574–591.

Huk, A. C., Katz, L. N., & Yates, J. L. (2017). The Role of the Lateral Intraparietal Area in (the Study of) Decision Making. *Annual Review of Neuroscience*, *40*(1), 349–372.

Jacobs, G. A., & Theunissen, F. E. (1996). Functional organization of a neural map in the cricket cercal sensory system. *Journal of Neuroscience*, *16*(2), 769–784.

Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological review*, *121*(1), 1–32.

Julesz, B. (1971). *Foundations of cyclopean vision*. Chicago: University of Chicago.

Julesz, B., & Bosche, C. (1966). *Studies of visual texture and binocular depth perception. A computer-generated movie series containing monocular and binocular movies.*

Julesz, B., & Payne, R. A. (1968). Differences between monocular and binocular stroboscopic movement perception. *Vision research*, *8*(4), 433–444.

Julier, S. J., & Uhlmann, J. K. (1997). A new extension of the Kalman filter to nonlinear systems. In *Int. symp. aerospace/defense sensing, simul. and controls* (pp. 182–193).

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Fluids Engineering*, *82*(1), 35–45.

Katz, L. N., Hennig, J. A., Cormack, L. K., & Huk, A. C. (2015). A Distinct Mechanism of Temporal Integration for Motion through Depth. *The Journal of Neuroscience*, *35*(28), 10212–10216.

Kelly, D. H. (1971). Theory of Flicker and Transient Responses,* II. Counterphase Gratings. *JOSA*, *61*(5), 632–640.

Kelly, D. H. (1976). Pattern detection and the two-dimensional Fourier transform: Flickering checkerboards and chromatic mechanisms. *Vision research*, *16*(3), 277–287.

Kim, S., & Burge, J. (2017). The lawful imprecision of human surface tilt estimation in natural scenes. *bioRxiv*, 180984.

King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., & Supowit, A. (1993). Efficient and Unbiased Modifications of the QUEST Threshold Method: Theory, Simulations, Experimental Evaluatiun, and Practical Implementation. *Vision research*, *34*(7), 885–912.

Kleiner, M., Brainard, D., Pelli, D., & Ingling, A. (2007). What's new in Psychtoolbox-3. *Perception*, *36*.

Koenderink, J. J. (1986). Optic flow. *Vision research*, *26*(1), 161–179.

Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in cognitive sciences*, *10*(7), 319–326.

Landers, D. D., & Cormack, L. K. (1997). Asymmetries and errors in perception of depth from disparity suggest a multicomponent model of disparity processing. *Perception & Psychophysics*, *59*(2), 219–231.

Lappe, M., Bremmer, F., & van den Berg AV. (1999). Perception of self-motion from visual flow. *Trends in cognitive sciences*, *3*(9), 329–336.

Levi, D. M., Hariharan, S., & Klein, S. A. (2002). Suppressive and facilitatory spatial interactions in peripheral vision: peripheral crowding is neither size invariant nor simple contrast masking. *Journal of Vision*, *2*(2), 167–177.

Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nature Neuroscience*, *5*(4), 356–363.

Li, Q., Hill, Z., & He, B. J. (2014). Spatiotemporal dissociation of brain activity underlying subjective awareness, objective performance and confidence. *The Journal of Neuroscience*, *34*(12), 4382–4395.

Liberman, A., Fischer, J., & Whitney, D. (2014). Serial dependence in the perception of faces. *Current biology : CB*, *24*(21), 2569–2574.

Lin, A., Maniscalco, B., & He, B. J. (2016). Scale-Free Neural and Physiological Dynamics in Naturalistic Stimuli Processing. *eNeuro*, *3*(5), ENEURO.0191–16.2016.

Liu, Y., Bovik, A. C., & Cormack, L. K. (2008). Disparity statistics in natural scenes. *Journal of Vision*, *8*(11), 19–19.

Lu, Z.-L., & Sperling, G. (1995). The functional architecture of human visual motion perception. *Vision research*, *35*(19), 2697–2722.

Lundqvist, M., Rose, J., Herman, P., Brincat, S. L., Buschman, T. J., & Miller, E. K. (2016). Gamma and Beta Bursts Underlie Working Memory. *NEURON*, *90*(1), 152–164.

Makin, J. G., Dichter, B. K., & Sabes, P. N. (2015). Learning to Estimate Dynamical State with Probabilistic Population Codes. *PLoS Computational Biology*, *11*(11), e1004554.

Maniscalco, B., Lee, J. L., Abry, P., Lin, A., Holroyd, T., & He, B. J. (2018). Neural integration of stimulus history underlies prediction for naturalistically evolving sequences. *The Journal of Neuroscience*, 1779–17.

Manning, J. R., Jacobs, J., Fried, I., & Kahana, M. J. (2009). Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. *The Journal of Neuroscience*, *29*(43), 13613–13620.

Martin, G. R. (2009). What is binocular vision for? A birds' eye view. *Journal of Vision*, *9*(11), 14.1–19.

Martin, G. R., & Shaw, J. M. (2010). Bird collisions with power lines: Failing to see the way ahead? *Biological Conservation*, *143*(11), 2695–2702.

Maunsell, J. H., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, *10*(1), 363–401.

Maunsell, J. H., & Van Essen, D. C. (1983a). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology*, *49*(5), 1148–1167.

Maunsell, J. H., & Van Essen, D. C. (1983b). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology*, *49*(5), 1127–1147.

Maus, G. W., Chaney, W., Liberman, A., & Whitney, D. (2013). The challenge of measuring long-term positive aftereffects. *Current biology : CB*, *23*(10), R438–9.

Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cerebral cortex (New York, N.Y. : 1991)*, *13*(11), 1257–1269.

McKee, S. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision research*, *30*(11), 1763–1779.

McKee, S. P., & Nakayama, K. (1984). The detection of motion in the peripheral visual field. *Vision research*, *24*(1), 25–32.

Menz, M. D., & Freeman, R. D. (2003). Stereoscopic depth processing in the visual cortex: a coarse-to-fine mechanism. *Nature Neuroscience*, *6*(1), 59–65.

Mikami, A., Newsome, W. T., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. II. Spatiotemporal range of directional interactions in MT and V1. *Journal of Neurophysiology*, *55*(6), 1328–1339.

Miller, J. P., Jacobs, G. A., & Theunissen, F. E. (1991). Representation of sensory information in the cricket cercal sensory system. I. Response properties of the primary interneurons. *Journal of Neurophysiology*.

Miller, K. J., Sorensen, L. B., Ojemann, J. G., & den Nijs, M. (2009). Power-law scaling in the brain surface electric potential. *PLoS Computational Biology*, *5*(12), e1000609.

Milstein, J., Mormann, F., Fried, I., & Koch, C. (2009). Neuronal shot noise and Brownian 1/f2 behavior in the local field potential. *PloS one*, *4*(2), e4338.

Mohler, B. J., Thompson, W. B., Creem-Regehr, S. H., Pick, H. L., & Warren, W. H. (2007). Visual flow influences gait transition speed and preferred walking speed. *Experimental brain research*, *181*(2), 221–228.

Monto, S., Palva, S., Voipio, J., & Palva, J. M. (2008). Very slow EEG fluctuations predict the dynamics of stimulus detection and oscillation amplitudes in humans. *The Journal of Neuroscience*, *28*(33), 8268–8272.

Movshon, J. A., & Newsome, W. T. (1996). Visual Response Properties of Striate Cortical Neurons Projecting to Area MT in Macaque Monkeys. *Journal of Neuroscience*, *16*(23), 7733–7741.

Mulligan, J. B., Stevenson, S. B., & Cormack, L. K. (2013). Reflexive and voluntary control of smooth eye movements. *SPIE Proceedings*, *8651*(Human Vision and Electronic Imaging XVIII).

Nachmias, J. (1981). On the psychometric function for contrast detection. *Vision research*, *21*(2), 215–223.

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *NeuroImage*, *56*(2), 400–410.

Neri, P. (2011). Coarse to fine dynamics of monocular and binocular processing in human pattern vision. In *Proceedings of the national academy of sciences* (pp. 10726–10731).

Nevin, J. A. (1969). Signal Detection Theory and Operant Behavior: A Review of David M. Green and John A. Swets: Signal Detection Theory and Psychophysics. *Journal of the Experimental Analysis of Behavior*, *12*(3), 475–480.

Newsome, W. T., & Pare, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience*, *8*(6), 2201–2211.

Nienborg, H., Bridge, H., Parker, A. J., & Cumming, B. G. (2005). Neuronal computation of disparity in V1 limits temporal resolution for detecting disparity modulation. *The Journal of Neuroscience*, *25*(44), 10207–10219.

Norcia, A. M., Sutter, E. E., & Tyler, C. W. (1985). Electrophysiological evidence for the existence of coarse and fine disparity mechanisms in human. *Vision research*, *25*(11), 1603–1611.

Norcia, A. M., & Tyler, C. W. (1984). Temporal frequency limits for stereoscopic apparent motion processes. *Vision research*, *24*(5), 395–401.

Nover, H., Anderson, C. H., & DeAngelis, G. C. (2005). A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for

speed discrimination performance. *The Journal of Neuroscience*, *25*(43), 10049–10060.

Ohzawa, I. (1998). Mechanisms of stereoscopic vision: the disparity energy model. *Current opinion in neurobiology*, *8*(4), 509–515.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607–609.

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision research*, *37*(23), 3311–3325.

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*(5), 376–404.

Palmer, S. E., Marre, O., Berry, M. J., & Bialek, W. (2015). Predictive information in a sensory population. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(22), 6908–6913.

Pan, W.-J., Thompson, G. J., Magnuson, M. E., Jaeger, D., & Keilholz, S. (2013). Infraslow LFP correlates to resting-state fMRI BOLD signals. *NeuroImage*, *74*, 288–297.

Paradiso, M. A. (1988). A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biological cybernetics*, *58*(1), 35–49.

Pelli, D. G. (1990). The quantum efficiency of vision. In C. Blakemore (Ed.), *Vision coding and efficiency* (pp. 3–24). Cambridge.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial vision*, *10*(4), 437–442.

Perrett, D. I., Smith, P. A., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., & Jeeves, M. A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London. Series B, Biological sciences*, *223*(1232), 293–317.

Petersen, S. E., & Sporns, O. (2015). Brain Networks and Cognitive Architectures. *NEURON*, *88*(1), 207–219.

Peterson, W., Birdsall, T., & Fox, W. (1954). The theory of signal detectability. *Transactions of the IRE Professional Group on Information Theory*, *4*(4), 171–212.

Pouget, A., Dayan, P., & Zemel, R. S. (2003). Inference and computation with population codes. *Annual Review of Neuroscience*, *26*(1), 381–410.

Raichle, M. E. (2015). The brain's default mode network. *Annual Review of Neuroscience*, *38*(1), 433–447.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, *85*(2), 59–108.

Ratcliff, R. (2002). A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin & Review*, *9*(2), 278–291.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, *20*(4), 873–922.

Ratcliff, R., & Rouder, J. N. (1998). Modeling Response Times for Two-Choice Decisions. *Psychological Science*, *9*(5), 347–356.

Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 127–140.

Reddi, B. A. J., Asrress, K. N., & Carpenter, R. H. S. (2003). Accuracy, information, and response time in a saccadic decision task. *Journal of Neurophysiology*, *90*(5), 3538–3546.

Regan, D., & Beverley, K. I. (1973a). The dissociation of sideways movements from movements in depth: Psychophysics. *Vision research*, *13*, 2403–2415.

Regan, D., & Beverley, K. I. (1973b). Some dynamic features of depth perception. *Vision research*, *13*(12), 2369–2379.

Regan, D., & Beverley, K. I. (1979). Binocular and monocular stimuli for motion in depth: Changing-disparity and changing-size feed the same motion-in-depth stage. *Vision research*, *19*(12), 1331–1342.

Regan, D., & Gray, R. (2009). Binocular processing of motion: some unresolved questions. *Spatial vision*, *22*(1), 1–43.

Rentschler, I., & Treutwein, B. (1985). Loss of spatial phase relationships in extrafoveal vision. *Nature*, *313*(6000), 308–310.

Ringach, D. L., Hawken, M. J., & Shapley, R. (2003). Dynamics of orientation tuning in macaque V1: the role of global and tuned suppression. *Journal of Neurophysiology*, *90*(1), 342–352.

Rokers, B., Cormack, L. K., & Huk, A. C. (2008). Strong percepts of motion through depth without strong percepts of position in depth. *Journal of Vision*, *8*(4), 6.1–10.

Rokers, B., Cormack, L. K., & Huk, A. C. (2009). Disparity- and velocity-based signals for three-dimensional motion perception in human MT+. *Nature Neuroscience*, *12*(8), 1050–1055.

Rose, A. (1948). The sensitivity performance of the human eye on an absolute scale. *Journal of the Optical Society of America*, *38*(2), 196–208.

Rust, N. C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nature Neuroscience*, *9*(11), 1421–1431.

Rust, N. C., & Movshon, J. A. (2005). In praise of artifice. *Nature Neuroscience*, *8*(12), 1647–1650.

Samonds, J. M., Potetz, B. R., & Lee, T. S. (2009). Cooperative and Competitive Interactions Facilitate Stereo Computations in Macaque Primary Visual Cortex. *Journal of Neuroscience*, *29*(50), 15780–15795.

Sanada, T. M., & DeAngelis, G. C. (2014). Neural representation of motion-in-depth in area MT. *The Journal of Neuroscience*, *34*(47), 15508–15521.

Scharf, L. L., & Demeure, C. (1991). *Statistical Signal Processing*. Prentice Hall.

Schiffner, I., & Srinivasan, M. V. (2015). Direct Evidence for Vision-based Control of Flight Speed in Budgerigars. *Scientific reports*, *5*(1), 10992.

Schmitt, L. I., Wimmer, R. D., Nakajima, M., Happ, M., Mofakham, S., & Halassa, M. M. (2017). Thalamic amplification of cortical connectivity sustains attentional control. *Nature Publishing Group*, *545*(7653), 219–223.

Sebastian, S., Abrams, J., & Geisler, W. S. (2017). Constrained sampling experiments reveal principles of detection in natural scenes. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(28), E5731–E5740.

Shadlen, M. N., & Newsome, W. T. (2001). Neural Basis of a Perceptual Decision in the Parietal Cortex (Area LIP) of the Rhesus Monkey. *Journal of Neurophysiology*, *86*(4), 1916–1936.

Shioiri, S., Kakehi, D., Tashiro, T., & Yaguchi, H. (2009). Integration of monocular motion signals and the analysis of interocular velocity differences for the perception of motion-in-depth. *Journal of Vision*, *9*(13), 10.1–17.

Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision research*, *38*(5), 743–761.

Simpson, W. A., Falkenberg, H. K., & Manahilov, V. (2003). Sampling efficiency and internal noise for motion detection, discrimination, and summation. *Vision research*, *43*(20), 2125–2132.

Smith, P. L. (1995). Psychophysically Principled Models of Visual Simple Reaction-Time. *Psychological review*, *102*(3), 567–593.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*(3), 161–168.

Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, *32*(2), 135–168.

Sorenson, H. W. (1985). *Kalman Filtering*. IEEE.

Srinivasan, M. V., Lehrer, M., Kirchner, W. H., & Zhang, S. W. (1991). Range perception through apparent image speed in freely flying honeybees. *Visual neuroscience*, *6*(5), 519–535.

Srinivasan, M. V., Zhang, S., Altwein, M., & Tautz, J. (2000). Honeybee navigation: nature and calibration of the "odometer". *Science*, *287*(5454), 851–853.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149.

Stephenson, C., & Braddick, O. (1983). Discrimination of relative spatial phase in fovea and periphery. *Investigative ophthalmology & visual science*, *24*, 146.

St John-Saaltink, E., Kok, P., Lau, H. C., & de Lange, F. P. (2016). Serial Dependence in Perceptual Decisions Is Reflected in Activity Patterns in Primary Visual Cortex. *The Journal of Neuroscience*, *36*(23), 6186–6192.

Swets, J. A. (2014). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers.*

Theunissen, F. E., & Miller, J. P. (1991). Representation of sensory information in the cricket cercal sensory system. II. Information theoretic calculation of system accuracy and optimal tuning-curve widths of four primary interneurons. *Journal of Neurophysiology*.

Todorov, E. (2005). Stochastic Optimal Control and Estimation Methods Adapted to the Noise Characteristics of the Sensorimotor System. *Neural computation*, *17*(5), 1084–1108.

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision research*, *35*(17), 2503–2522.

Tyler, C. W. (1971). Stereoscopic depth movement: two eyes less sensitive than one. *Science*, *174*(4012), 958–961.

Uhlmann, J. K. (1992). Algorithms for Multiple-Target Tracking. *American Scientist*, *80*(2), 128–141.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological review*, *108*(3), 550–592.

van Dam, L. C. J., & Ernst, M. O. (2013). Knowing each random error of our ways, but hardly correcting for it: an instance of optimal performance. *PloS one*, *8*(10), e78757.

van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings. Biological sciences*, *265*(1394), 359–366.

von Helmholtz, H. (1867). *Treatise on Physiological Optics.*

Voss, R. F., & Clarke, J. (1975). '1/f noise' in music and speech. *Nature*, *258*(5533), 317–318.

Wald, A. (1947). *Sequential Analysis.* New York: WIley.

Walsh, G., & Charman, W. N. (1988). Visual sensitivity to temporal change in focus and its relevance to the accommodation response. *Vision research*, *28*(11), 1207–1221.

Wang, B., & Ciuffreda, K. J. (2006). Depth-of-focus of the human eye: theory and clinical implications. *Survey of ophthalmology*, *51*(1), 75–85.

Watson, A. B. (1979). Probability summation over time. *Vision research*, *19*(5), 515–522.

Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120.

Weber, E. H. (1834). *De Pulsu, resorptione, auditu et tactu.*

Welchman, A. E., Lam, J. M., & Bülthoff, H. H. (2008). Bayesian motion estimation accounts for a surprising bias in 3D vision. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(33), 12087–12092.

Welchman, A. E., Tuck, V. L., & Harris, J. M. (2004). Human observers are biased in judging the angular approach of a projectile. *Vision research*, *44*(17), 2027–2042.

Wheatstone, C. (1838). Contributions to the physiology of vision.–Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical transactions of the Royal Society of London.*

Williams, P. E., Mechler, F., Gordon, J., Shapley, R., & Hawken, M. J. (2004). Entrainment to video displays in primary visual cortex of macaque and humans. *The Journal of Neuroscience*, *24*(38), 8278–8288.

Wolpert, D. M., & Ghahramani, Z. (1995). An internal model for sensorimotor integration. *Science*, *269*, 1880–1883.

Wozniak, R. H. (1999). *Classics in Psychology, 1855-1914*.

Wright, M. J., & Johnston, A. (1983). Spatiotemporal contrast sensitivity and visual field locus. *Vision research*, *23*(10), 983–989.

Wu, W., Black, M. J., Mumford, D., Gao, Y., Bienenstock, E., & Donoghue, J. P. (2004). Modeling and Decoding Motor Cortical Activity Using a Switching Kalman Filter. *IEEE Transactions on Biomedical Engineering*, *51*(6), 933–942.

Yaron, A., Hershenhoren, I., & Nelken, I. (2012). Sensitivity to Complex Statistical Regularities in Rat Auditory Cortex. *NEURON*, *76*(3), 603–615.

Yates, J. L., Park, I. M., Katz, L. N., Pillow, J. W., & Huk, A. C. (2017). Functional dissection of signal and noise in MT and LIP during decision-making. *Nature Neuroscience*, *20*(9), 1285–1292.

Zeki, S. M. (1974). Cells responding to changing image size and disparity in the cortex of the rhesus monkey. *The Journal of physiology*, *242*(3), 827–841.