NON-RETINOTOPIC REFERENCE FRAMES FOR DYNAMIC FORM AND MOTION PERCEPTION

A Dissertation
Presented to
The Faculty of the Department of Electrical and Computer Engineering
University of Houston

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
In Electrical and Computer Engineering

by
Mehmet Naci Agaoglu

May 2015

NON-RETINOTOPIC REFERENCE FRAMES FOR DYNAMIC FORM AND MOTION PERCEPTION

Mehmet Naci Agaoglu

Approved:

Chair of the Committee,
Haluk Ogmen, Professor,
Electrical and Computer Engineering

Committee Members:

Jose Luis Contreras-Vidal, Professor,
Electrical and Computer Engineering

Vallabh E. Das, Professor,
College of Optometry

Michael H. Herzog, Professor,
Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL)

Bhavin Sheth, Associate Professor,
Electrical and Computer Engineering

Suresh K. Khator, Associate Dean,
Cullen College of Engineering

Badri Roysam, Professor, Chair,
Electrical and Computer Engineering

Acknowledgements

> *"…Initially the hypotheses, ideas can be "too simple", "unrealistic" etc.; but that should not discourage you, that's how we learn (to learn to walk, a baby starts by awkward gestures, lots of falls etc. – but that's normal!). So don't worry; just try… I am here to work with you every step of the way. Haluk"*

I would like to express my deepest gratitude to Dr. Haluk Ogmen for being such a great mentor and for teaching me *how to walk*. Without his constant support, encouragement, constructive criticism, time, and funding, this dissertation would not be possible. I especially thank him for the countless discussions we had about research and many other things.

I wish to extend my sincere gratitude to Dr. Vallabh Das, who, in addition to serving in my doctoral committee, has been an exceptional advisor as well. I thank him especially for opening my eyes to the field of neurophysiology, and teaching me an invaluable set of skills over the years. I am grateful for his confidence in me; without his time, efforts, and funding, I would not be able to complete my dissertation.

I also thank other distinguished members of my doctoral committee, Dr. Sheth, Dr. Contreras-Vidal, and Dr. Herzog, for their feedback and encouragements.

I wish to thank the members of our lab, my friends, especially those who participated in my experiments, for their help in my research. I wish to extend a special thanks to my best friend, colleague, lab-mate, and wife, Sevda, for her unconditional support.

And last but not least, I am indebted to my parents for their unparalleled support.

Bismihi Sübhanehû…

NON-RETINOTOPIC REFERENCE FRAMES FOR DYNAMIC FORM AND MOTION PERCEPTION

An Abstract
of a
Dissertation
Presented to
The Faculty of the Department of Electrical and Computer Engineering
University of Houston

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
In Electrical and Computer Engineering

by
Mehmet Naci Agaoglu

May 2015

Abstract

The spatial representation of a visual scene in the early visual system is well known. The optics of the eye map the three-dimensional environment onto two-dimensional images on the retina. These retinotopic representations are preserved in the early visual system. Retinotopic representations and processing are among the most prevalent concepts in visual neuroscience. However, it has long been known that a retinotopic representation of the stimulus is neither sufficient nor necessary for perception. Many visual processes (form and motion perception, visual search, attention, and perceptual learning) that have been thought to occur in retinotopic coordinates, have been found to operate in non-retinotopic coordinates. Based on these findings, our goal was to characterize the non-retinotopic representations and their underlying reference frames. We proposed that each retinotopic motion vector creates a perceptual reference-frame field in the retinotopic space (like an electromagnetic field), and interactions between these fields determine the selection of the effective reference frame. To test this theory, we performed a series of psychophysical experiments. We first used the slit-viewing paradigm to investigate how features of a moving object are attributed. Our results support the predictions of the non-retinotopic feature-processing hypothesis and demonstrate the ability of the visual system to operate non-retinotopically at a fine feature processing level. We then used a variant of the induced motion paradigm to investigate non-retinotopic reference frames for motion perception. We found that the effective reference frame for motion perception is non-retinotopic, and emerges from an amalgamation of motion-based, retinotopic and spatiotopic reference frames. In determining the percept, the influence of relative motion, defined by a motion-based reference frame, dominates those of retinotopic and spatiotopic motions within a finite region. Moreover, we found that different reference fields interact nonlinearly, and the way they interact depends on how motion vectors are grouped. Finally, we investigated how various spatiotemporal factors influence reference frame selection for motion perception. In line with our theory, we found that the motion-based-nearest-vector metric can fully account for all the data reported here. Taken together, these findings suggest that the brain actively constructs perceptual space by using motion-based reference frames.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1. Background and Rationale

The optics of the eye projects the three-dimensional environment onto two-dimensional images on the retina such that neighboring points in the environment activate neighboring receptors in the retina. This topographical structure, called retinotopy, is preserved in the early visual cortex (Sereno et al., 1995; Tootell, Silverman, Switkes, & Valois, 1982). Retinotopic images undergo frequent changes because of the dynamic nature of ecological viewing conditions. These changes occur either due to observer's motion (endogenous) or motions of objects in the environment that are external to the observer (exogenous). Endogenous motion creates global changes in the retinotopic image, whereas retinotopic motion due to exogenous motion is typically confined to a local region, which represents the motion trajectory of a moving object. Global changes in the retinotopic images due to endogenous motion are not uniform throughout the entire visual scene. The projections of closer objects on the retina move faster than more distant objects, creating motion parallax. These distance dependent differences are used to infer depth from two dimensional retinotopic images (Rogers & Graham, 1979). On the other hand, retinotopic motions due to exogenous motion may or may not be correlated. For instance, a bird colony flying as a group can generate correlated retinotopic motions, while the motion of each bird can be independent from each other. Moreover, since endogenous motions originate from within the observer, neural signals related to motor planning, execution, and feedback are available before and during the motion of the observer. These signals can be used to stabilize the perception before and during the motion. On the other hand, the visual system has no information about any exogenous motion in the visual field *in advance*; it should be calculated *during* the motion.

These fundamental differences between endogenous and exogenous motions imply different mechanisms for processing changes in retinal stimulation. Under normal viewing conditions though, these mechanisms can work in harmony to give coherent, clear and stable percepts. For instance, eyes can be moved to track a moving object, keeping its percept sharp and clear. But the presence of other objects (moving or stationary) in the environment limits the cooperation between the two mechanisms; tracking a moving object will generate additional and spurious retinal motions for other objects and background. Therefore, the cooperation between different mechanisms does not solve the problem; it translates it to the background or to other objects.

1.1.1. Motion Blur and Moving Ghosts

Retinotopic images are not point-samples of the environment at a given time instant. They are formed by stimulation of the receptors in the retina for a finite period of time, analogous to how a camera works. Furthermore, after the offset of a light source, the retinotopic image persists for a finite amount of time, called *visible persistence* (Coltheart, 1980; Haber & Standing, 1970). If a camera with a long exposure duration is moved (or an object moves in its field of view) at speeds that do not allow enough time to fully expose or discharge the camera sensor, the resultant images will have extensive smear. Likewise, any type of retinal motion mentioned so far should generate blurry percepts. However, in normal viewing, humans do not experience such motion blur (Bex, Edgar, & Smith, 1995; Hammett, 1997; Ramachandran, Rao, & Vidyasagar, 1974). This is a significant difference between a camera and the human visual system, suggesting the presence of brain mechanisms that are developed to actively deal with the motion blur problem.

Previous studies showed that the perceived motion blur depends on exposure duration in a non-monotonic way: It increases up to 30 ms exposure durations and decreases for longer exposure durations (Burr & Morgan, 1997; Burr, 1980). This decrease in the perceived motion blur is called *motion deblurring*. Moreover, many studies showed that motion blur is also modulated by the density of objects in the visual field: the more objects presented in a given area the less motion blur is perceived (Chen, Bedell, & Ogmen, 1995; Di Lollo & Hogben, 1985; Dixon & Hammond, 1972; Farrel, 1984). More specifically, the presence of spatiotemporally proximal stimuli can reduce the spatial extent of perceived motion blur. These findings suggest that there are stimulus-driven inhibitory mechanisms in the human visual system. Burr and his colleagues proposed that spatio-temporally oriented receptive fields can account for motion deblurring (Burr & Morgan, 1997; Burr, Ross, & Morrone, 1986). However, this model cannot explain the blur that would be generated by receptive fields with different spatio-temporal orientations because all mechanism will be activated to varying degrees by the stimulus. In addition, motion trajectories can be arbitrarily complex which require infinitely many possibilities of receptive field profiles. In fact, Chen, Bedell, and Ogmen (1995) directly tested whether motion deblurring results from a motion compensation mechanism (e.g., integration within the spatiotemporally oriented receptive fields) or from inhibition by spatiotemporally adjacent stimuli. They found that for motion deblurring, the activation of the motion mechanism is not sufficient, and that it primarily results from masking exerted by spatiotemporally adjacent stimuli (Chen et al., 1995).

As an alternative, metacontrast masking, defined as a reduction of visibility of a 'target' stimulus by a spatiotemporally non-overlapping 'mask' stimulus, is proposed to reconcile contradictory findings on visible persistence and motion deblurring (Chen et al., 1995; Ogmen, 2007; Purushothaman, Patel, Bedell, & Ogmen, 1998). Although it can explain the observations made in different studies, metacontrast mechanisms provide only a partial solution to the motion blur problem. Metacontrast mechanisms would only decrease the length of motion streaks so that the amount of motion blur is reduced. But the ghost-like, i.e. without significant form information, appearances will still be present (see Fig.1.1). There should be an alternative way of representing the visual information about moving objects other than solely by their retinotopic representations to establish clear, sharp and stable percepts. The required representations should be independent of occlusions and dynamic changes in the retinotopic representations.

Figure 1.1. Motion blur and moving ghost problems. Note that the vehicles closer to the camera have ghost-like appearances and the spatial extents of their motion blur are larger compared to those far away from the camera. Reproduced from Free-photo.com.

Our fundamental hypothesis is that the retinotopic information about the form of moving stimuli is transferred to non-retinotopic representations through motion-

based reference frames, and these representations allow enough time for the neural circuitry to synthesize the shape information from accumulated information over time. There have been several studies that support our hypothesis. In the following section, we will explain some of them and how they lead to the rationale of the proposed experiments.

1.1.2. Dynamic Occlusions and Feature Attribution

When an object moves in front of another object, a retinotopically organized receptive field will receive a mixture of features of the two objects over time. This imposes a problem for the visual system. As features are being processed in time, how can the human visual system correctly attribute the features to their corresponding targets? This question can be addressed by pitting retinotopic and non-retinotopic processes against each other. The traditional experimental method for distinguishing between retinotopic and non-retinotopic processes is the Saccadic Stimulus Presentation Paradigm (SSPP). In this paradigm, observers are asked to make a saccade and two stimuli are presented, one before the saccade and the second after the saccade. Due to the intervening saccade, relative alignment of the stimuli (e.g., the letter array and the ring in Fig. 1.2) differs according to different reference frames. According to the retinotopic representation of the stimulus, the ring surrounds letter C as shown in Fig. 1.2, while according to the spatiotopic representation of the stimulus (e.g., a non-retinotopic coordinate system based at the center of the display), the ring surrounds the letter B. Because of the self-generated eye movement, the SSPP cannot be applied to the cases where eyes are stationary. Various processes related to saccadic eye movements, such as saccadic suppression and corollary discharge signaling come into play, and any potential non-retinotopic processes can be overshadowed by the eye-movement related processes. In order to investigate non-retinotopic processes in the absence of eye movements, an alternative paradigm has been proposed (Ogmen, Otto, & Herzog, 2006) based on the Ternus-Pikler display (Pikler, 1917; Ternus, 1926).

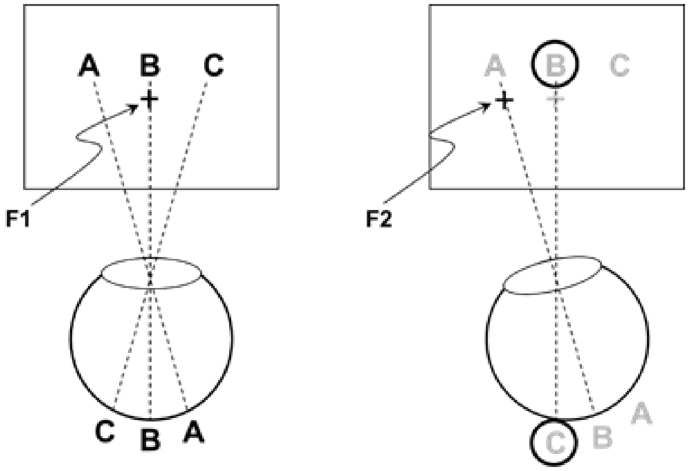

Figure 1.2. Saccadic Stimulus Presentation Paradigm (SSPP). (Adapted from Boi et al., 2009).

Fig. 1.3A shows a basic Ternus-Pikler (T-P) display. The first frame contains three elements each consisting of two abutting lines with a small vertical gap in between. After an inter-stimulus interval (ISI), these elements are shifted to the right by one inter-element distance so that two of the elements retinotopically overlap across frames. For short durations of the blank interval, e.g., ISI = 0 ms, observers report that the leftmost element in Frame 1 jumps to the rightmost element’s location in Frame 2, while the other two elements appears stationary (Fig. 1.3B). This percept is called as *element motion*. For longer ISIs, e.g., ISI = 100 ms, all three elements appear to be moving to the right by one inter-element distance as a group. This percept is called as *group motion*. Ogmen, Otto, and Herzog (2006) inserted a small horizontal offset, called the *probe Vernier*, to the central element in Frame 1 as shown in Fig. 1.3A to study retinotopic versus non-retinotopic processes. In their experiments, observers' task was to attend a specific element and to report the perceived direction of probe Vernier at this element (direction of horizontal offset of the lower line segment with respect to the upper segment). None of the elements in Frame 2 contained an offset, and observers did not know the location of the probe Vernier.

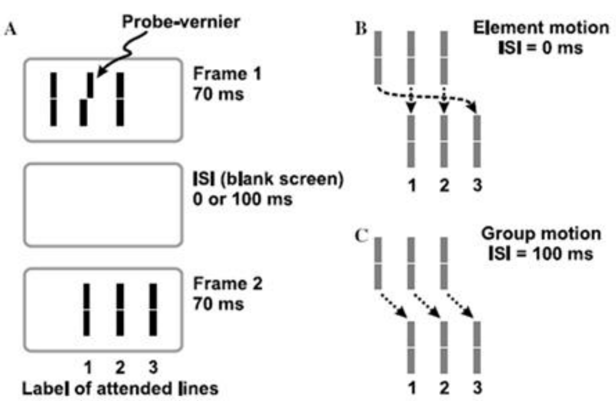

Figure 1.3. The Ternus-Pikler Display (A) and the associated percepts of "Element Motion" (B) and "Group Motion" (C). From (Ogmen, Otto, and Herzog, 2006).

The retinotopic hypothesis predicts that the probe Vernier should be integrated with Element 1 as long as the ISI is within the temporal integration window (~120 ms as indicated by visible persistence studies, e.g., Coltheart, 1980). On the other hand, if feature processing and integration take place according to non-retinotopic relations, motion grouping in this case (Fig. 1.3B and 1.3C), one would expect the probe Vernier to be integrated with Element 1 when the percept is element motion, and with Element 2 in the case of group motion. Their experimental results supported the predictions of non-retinotopic hypothesis as shown in Fig. 1.4A. The percentage of responses in agreement with the probe Vernier depends on ISI, in contrary to the predictions of retinotopic hypothesis. Replication of the same experiment with a larger inter-element separation also yielded a similar pattern (Fig. 1.4B). Finally, in a control experiment where the leftmost element of Frame 1 and the rightmost element of Frame 2 are removed, different ISI values did not affect the observers' responses, ruling out any contribution of temporal separation of two stimulus frames.

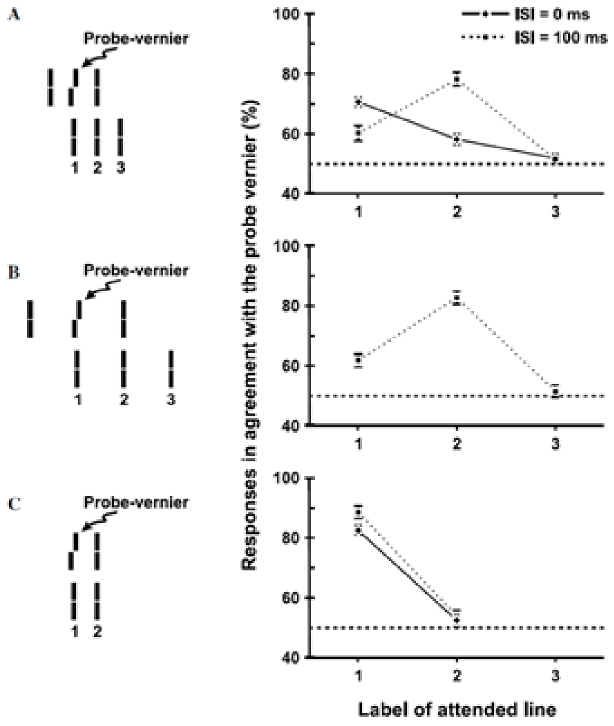

Figure 1.4. The percentage of responses in agreement with the probe Vernier as a function of the attended line for the Ternus-Pikler display shown in Fig. 3. From (Ogmen, Otto, and Herzog, 2006).

It has also been demonstrated that the Ternus-Pikler paradigm can be used as a litmus test to investigate whether a given function of the human visual system is implemented via retinotopic or non-retinotopic processes (Boi, Ogmen, & Herzog, 2011; Boi, Ogmen, Krummenacher, Otto, & Herzog, 2009). In a series of experiments, visual search, motion perception, form perception, exogenous attention, and binocular rivalry have been shown to be operating on non-retinotopic representations (Boi et al., 2009; Boi, Vergeer, Ogmen, & Herzog, 2011; Ogmen et al., 2006). However, in all of the studies involving Ternus-Pikler displays or SSPP paradigm, the complete retinotopic images of the stimuli were available to the observer. Therefore, the potential contributions of retinotopic processes were not completely eliminated in these methods.

1.1.3. Anorthoscopic Perception: A Tool to Investigate the Non-retinotopic Processes

Anorthoscopic perception occurs when an object moves behind a narrow slit. Observers often report seeing the object under motion as a spatially extended whole although each fragment of the object falls on the same retinotopic strip over time. This fact shows that a retinotopic representation is not necessary for the perception of form, and indicates that the visual system somehow constructs the spatially extended form of the objects that are moving behind a slit. A temporary storage medium and an algorithm to integrate different fragments of the objects are required to accomplish this feat. There have been many hypotheses to explain anorthoscopic perception. One of the first hypotheses stated that the anorthoscopic percepts are results of retinal painting of successive parts of a stimulus, assuming that the observers unconsciously pursue the moving object (von Helmholtz, 1867). However, Parks (1965) revisited the same problem and found that objects appear as a whole in the vicinity of the slit even in the absence of any eye movements. He also proposed an alternative hypothesis that each successive part of an object is temporarily stored in a post-retinal storage and integration into a whole is done by using a time-of-arrival code. Fig. 1.5A shows a stimulus used to test these two hypotheses (McCloskey & Watkins, 1978; Sohmiya & Sohmiya, 1992, 1994). Two triangular shaped arrays of dots move in opposite directions behind a narrow slit. The tips of the triangles become visible through the slit simultaneously, followed by the middle rows and finally the longest segments. Since eyes cannot track both objects at the same time due to the push-pull mechanism of the ocular muscles, observers can move both eyes in either to the left or to the right. Assuming a leftward eye movement, retinal painting hypothesis predicts two triangles whose tips are at the leftmost locations as shown in Fig. 1.5B. Similarly, since the time of arrival of the tips, the middle segments, and the longest segments are equal for both triangles, Parks' hypothesis makes the same predictions as retinal painting hypothesis (Fig. 1.5C). On the contrary, observers report seeing two mirrored image triangles, i.e. the actual physical configuration of the stimuli (McCloskey & Watkins, 1978; Sohmiya & Sohmiya, 1992, 1994). Therefore, the spatial layout is constructed neither through eye movements nor through a transformation based on time dimension. Interestingly, the direction of motion is very informative about the veridical shape information while the retinal information available through the slit is ambiguous. Ogmen (2007) suggested that the space dimension is created from motion dimension through the use of motion-based reference frames and this proposal has been supported by a recent study (Aydin, Herzog, & Ogmen, 2008).

In Chapter 2, we will investigate how non-retinotopic representations are constructed under occlusions, i.e. when complete retinotopic images are not available (Specific Aim 1). We will bring the type of stimuli used in Ternus-Pikler displays and anorthoscopic perception together to directly pit retinotopic and non-retinotopic feature processing against each other. We will also test the hypothesis that motion groupings generate local reference frames, and these reference frames are used to attribute features in amodal completion. Anorthoscopic perception will provide us a very useful tool to investigate non-retinotopic feature processing because in contrast to Ternus-Pikler and SSPP paradigms, it minimizes retinotopic representations.

1.1.4. Phenomenal Identity

The motion of retinotopic images can be interpreted as either the motion of the observer or the motion(s) in the environment as we mentioned before. In order to conclude that an object is moving, correspondence operations should be carried out (Ternus, 1926). Motion groupings of different features can help to amalgamate retinotopic activations into a coherent whole. Kahneman and colleagues adapted two concepts from computer science, namely addresses and files, to explain how the phenomenal identity of an object could be established and maintained despite drastic changes of its retinotopic image (Kahneman, Treisman, & Gibbs, 1992). Object files, each containing information about an object, form the basis of their theory. Object files are addressed by the location of the object at a given time instant. Therefore, any change in features of an object will not lead to a perception of a separate object; the phenomenal identity of that object will be preserved.

The object file theory emphasizes the reference frames (moving with the object) and has been successful in explaining a wide range of data in the literature. But it has several important shortcomings. First, in the object file theory, features are obtained instantly and placed in appropriate files, while in reality feature processing takes time and comes with the motion blur problem (see Section 1.1.1). Second, an object is defined by its features, and a requirement for opening a new object file is having at least some features of the new object already processed. It becomes a chicken-egg problem; which one of these processes, i.e., feature extraction and object file opening, takes place first is vague. Third, how the location index is calculated is another problem of the object files theory. For instance, if the center of mass of the object is used as the location index, then the retinotopic extent, i.e., the geometry, of the object must be known. However, the geometry of objects is not much of a concern in the object file theory. Fourth, as we mentioned before, objects occlude each other, and as a result, location indices of different objects may occupy the same retinotopic locus. Indexing different objects with the same location index induces ambiguity in attribution of features. Therefore, location information alone is not sufficient; spatial extent and the occlusion information should also be represented for insertion of correct features to correct object files. Fifth, the neural processes that lead to object individuation as suggested by the object file theory has never been specified. Sixth, the object file theory falls short regarding the form-from-motion displays. For instance, adult humans can recognize point-light walkers as coherently moving objects in less than 100 ms (Johansson, 1973; Pavlova, Birbaumer, & Sokolov, 2006). Accounting for this ability requires creation of object files for each point-light and organizing them hierarchically to produce the “multi-object” percept, i.e., biological motion, in less than half the time normally it would take for unimodal feature binding (Fields, 2011). In addition, newborns show a predisposition to biological motion displays (Bardi, Regolin, & Simion, 2011) suggesting that whatever the underlying mechanism of object individuation and grouping, its major constituent is very primitive. It is not clear how object files can be created for disconnected, boundary-less sets of objects such as point-light walkers. The object file theory does not address these issues.

1.1.5. Our Theoretical Approach

Ternus, Kahneman and colleagues stress the role of motion based reference frames and grouping. But in order to account for the problems of motion blur, moving ghosts, feature attribution, dynamic occlusions, and phenomenal identity, the dynamics and the geometry of visual processing should also be considered. Here, we propose a unified solution to these problems. Fig. 1.6 illustrates our approach conceptually. According to our approach, moving objects generate first retinotopic motion vectors. From these motion vectors, common motion vectors can be grouped to generate local reference frames. Each local reference frame creates a field (like an electromagnetic field) around itself and different fields generated by different reference frames interact with each other to reach an equilibrium. This equilibrium state determines the reference frames for the perception of form and motion.

According to our theory, which will be referred to as a Reference-Frame Metric Field (RFMF) theory from now on, non-retinotopic representations are constructed from local reference frames that are generated by local motion vectors, and features are attributed based on these local reference frames. Since early vision is retinotopic and perception is non-retinotopic, a reference-frame transform should take place to construct non-retinotopic representations. We cannot arbitrarily choose a certain reference frame over another for perception. In other words, we cannot use at will any reference frame; the choice and the transformations occur automatically in perception although they can be manipulated to some extent with attention. The exact nature of this coordinate transform can only be established by first determining the non-retinotopic reference frames and how they are constructed. In this dissertation, we will characterize the non-retinotopic reference

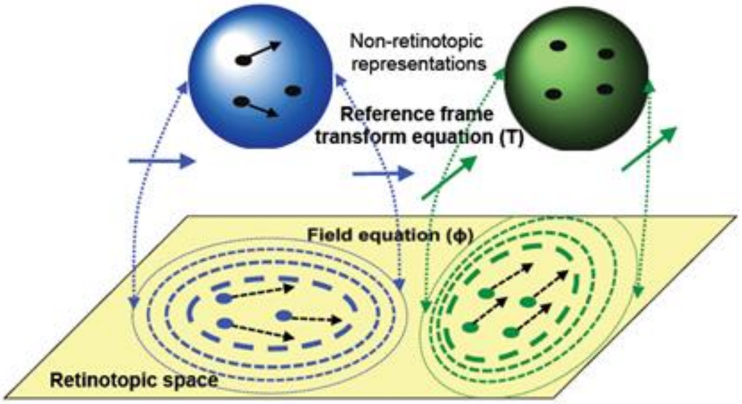

Figure 1.5. Our approach provides a unified solution to problems emanating from dynamic changes in the retinotopic representations.

frames for the perception of form (Chapter 2) and motion (Chapters 3, 4, and 5). In Chapter 2, we will investigate whether and how non-retinotopic representations of form can be constructed in the brain by use of perceptual dimension of motion as reference frame for feature processing and attribution (***Specific Aim 1***). In Chapter 3, we will investigate what non-retinotopic reference frames are used for motion perception, and how they are selected in the transformations without occlusions (***Specific Aim 2***). In Chapter 4, we will also test the hypothesis that the motion groupings generate field effects that interact with each other to determine local reference frames (***Specific Aim 2***). Finally, in Chapter 5, we will investigate the metric of these non-retinotopic reference frames (***Specific Aim 3***). More specifically, we will compare and contrast several form-based and motion-based metrics for non-retinotopic reference frames for motion perception to test our theoretical approach.

1.2. Specific Aims

Perceptual and cognitive abilities are determined by how the nervous system registers, processes, and transforms information obtained by sensory organs. These functions must be carried out efficiently for an organism to sense and act upon whatever information it can get. The invention of object oriented computing opened a new era in computer science; information is represented by objects or classes which allowed efficient and very fast algorithms. Likewise, the nervous system may build such representations to implement its functions. It has long been known that most of the primary sensory and motor representations are dictated by the physics of their corresponding organs and are relatively well understood (see reviews: Crawford, Henriques, & Medendorp, 2011; Galati, Pelle, Berthoz, & Committeri, 2010). However, the understanding of intermediate representations between the primary sensory and motor areas has been a fundamental yet unresolved question in brain science. The broad goals of this study are to characterize these intermediate representations, their underlying reference frames, and to discover their implications for perception and cognition.

The human visual system is one of the primary sensory systems in which the physics of the sensory organ shapes the corresponding representations (Sereno et al., 1995; Tootell et al., 1982). Based on retinotopic representations and visible persistence, moving objects should appear highly blurred (Burr, 1980; Coltheart, 1980; Hammett, 1997; Ogmen, 2007). Moreover, objects in the environment frequently undergo occlusions. As a consequence, features of different objects, such as color and texture, should blend into each other and to those of the background. But under normal viewing conditions, moving objects appear relatively sharp and clear, and features are attributed correctly (Burr & Morgan, 1997; Hammett, 1997; Marinovic & Arnold, 2013; Ogmen, 2007). Objects also maintain their identity despite dynamic changes in their retinal projections. How are features attributed to correct objects, and how is phenomenal identity established and maintained?

Is what we perceive dictated by what falls on the retina? Does the visual system construct percepts of the distal stimuli by using representations that are different than retinotopic representations?

Most theories of vision make use of purely retinotopic representations and/or are based on features extracted by retinotopically organized receptive fields. However, it has been shown that retinotopic representations are neither sufficient nor necessary for visual perception (Agaoglu, Herzog, & Ogmen, 2012; Breitmeyer & Ogmen, 2000; Ogmen & Herzog, 2010). Moreover, many visual processes (form perception (Ogmen et al., 2006), motion perception (Boi et al., 2009), visual search and attention (Boi, Vergeer, et al., 2011; Otto, Ogmen, & Herzog, 2010a), and processing of visual attributes such as size (Kawabe, 2008), luminance (Shimozaki, Eckstein, & Thomas, 1999), and color (Nishida, Watanabe, Kuriki, & Tokimoto, 2007)) that have been thought to occur in retinotopic coordinates, have been found to be operating in non-retinotopic coordinates. Thus, based on these findings, our broad goal is to characterize the non-retinotopic (intermediate) representations used in the visual system.

A representation can be conceptualized by its coordinate system (reference frame), which is used to localize the stimulus, and by its metric (scale), which defines distance upon which properties such as size, speed, volume, etc. can be defined. What are reference frames for non-retinotopic representations? How are stimuli quantified in non-retinotopic representations? We propose a set of experiments to address the questions outlined so far through three specific aims:

Specific Aim 1: How are non-retinotopic representations constructed under occlusions, a ubiquitous scenario in everyday viewing? What are non-retinotopic reference frames for form perception? Test the hypothesis that local motion groupings serve as non-retinotopic reference frames and these reference frames are used to attribute features in anorthoscopic perception.

Specific Aim 2: How are non-retinotopic representations constructed in the absence of occlusions, a simpler case which allows us to tease apart contributions of various factors? What are non-retinotopic reference frames for motion perception? Test the hypothesis that motion groupings generate field effects in space and interactions among fields determine local reference vectors in space-time.

Specific Aim 3: What is the metric of non-retinotopic reference frames? Compare and contrast the form-based and motion-based metrics for non-retinotopic reference frames for motion perception.

1.3. Significance

One of the starting points of the scientific revolution was the paradigm shift away from the geocentric model, which posits the Earth at the center of the galaxy, towards the heliocentric model with the Sun at the center of our solar system. This shift simplified the expressions for planets' motions and led to other significant discoveries. Similar transforms of reference frames in perception and cognition are claimed to occur during early infancy in order to construct an understanding of continuous existence of some aspects of the environment whether or not these aspects reach senses. Piaget interprets the development of phenomenal identity and object permanence during the early infancy as "Copernican revolution" in the brain at the end of the sensory-motor stage of cognitive development

(Piaget, 1969). A shift analogous to the one Copernicus[1] postulated takes place in order to construct an understanding of an objective world independent from the subject. Much of the sensory and motor representations of the information about perception and action are dictated by their corresponding sensory or motor organs and well understood. A fundamental question in brain science has been the understanding the intermediate representations between the primary sensory and motor areas.

Intermediate representations can be categorized in two groups. The first type guides the organism's actions with respect to the environment, such as navigation and manipulation. Since the first type of representations is tightly related to self-movements of the organism, their main reference frame is ego-centric, i.e., based on organism (eye-based, body-based, limb-based etc.) resembling the aforementioned geocentric model. The second type of intermediate representations construct the understanding of the self as an object among other objects in the environment. The conceptualization of the environment as an entity independent from the subject's perception necessitates exo-centric reference frames, just like the need for the heliocentric model to understand the motion of other planets that are independent from Earth's motion. While most research focused on ego-centric intermediate representations, very little is known about exo-centric reference frames. The outcome of the proposed study will be the characterization of exo-centric representations, which will advance our understanding of how the visual system operates under normal viewing conditions.

---

[1] In order to give proper credit to those who had done the same or similar work independently before Copernicus, we cite two other scientists Ibn Shatir and Nasir al-Din al-Tusi. Interested readers can refer to discussions on the topic, such as Roberts (1957) and Veselovsky (1973).

To address the issue outlined above, we will focus on the human visual system where neighboring points in the environment are mapped on to neighboring receptors in the sensory organ, creating retinotopic (sensory) representations. Most theories of vision have been based on these retinotopic representations and retinotopically organized receptive fields. In other words, the intermediate representations have not been utilized much in explaining the visual phenomena. However, visual phenomena such as anorthoscopic perception (perception of an object moving behind a narrow slit) and para-meta-contrast masking (reduction of visibility of a stimulus by a spatially non-overlapping mask stimulus) show that a retinotopic image (a sensory representation alone) is neither necessary nor sufficient for perception (Ogmen, 2007). The limitations of retinotopic representations stem, on the one hand, from the movements of the organism (eye, head and body movements cause rapid changes in the retinotopic image) and, on the other hand, from the movements of the objects, i.e., movements that are external and independent from the organism. According to our approach, exo-centric reference frames that avoid ambiguities in the retinotopic representations are required to overcome these problems. The proposed study aims at investigating the second type of intermediate representations to provide a unified solution to a broad range of fundamental questions in perception and cognition. These include the problems of motion blur, moving ghosts, feature attribution, dynamic occlusions, and the phenomenal identity described before (see Background and Rationale). We propose that non-retinotopic representations can bring a unified solution.

1.4. Innovation

Each organism needs to sense and act upon its surrounding environment to survive. How the information acquired by the sensory organs is represented has been the main goal of many physiological and behavioral studies. In this regard, the human visual system is one of the most investigated sensory systems. In addition to retinotopic representations (eye-centered), in which neighborhood relationships are preserved, neurons with other ego-centric (e.g., head-centered, limb-based), and exo-centric (e.g., spatiotopic, object-centered) tuning curves have been discovered (see reviews: Crawford, Henriques, & Medendorp, 2011; Galati, Pelle, Berthoz, & Committeri, 2010). However, most studies in this area focused on the limitations of the retinotopic representations and two broad types of problems associated with processes based on purely retinotopic representations. The first type of problems is how the brain compensates for the changes in retinotopic images due to subject's movements. The second type of problems pertains to how movements of external objects and their forms are processed in the brain so that the resultant percepts are sharp, clear, and free of identity binding issues (see Background and Rationale for problems of moving ghosts, motion blur, phenomenal identity, and dynamic occlusions and feature attribution).

The first type of problems is relatively simpler than the second type of problems for two reasons: First, since the brain generates the commands for the subject's movements, a copy of the command signal (known as *efference copy* or *corollary discharge*) can be used to compensate for the changes in sensory input due to subject's movements. In this sense, imminent changes in sensory representations due to subject's movements are predictable. Second, the changes due to self-motion, is global on the sensory representations whereas the second type of problems can occur within a local region in the visual field. Consider a man standing at a bus station and looking at a bus approaching him. The projections of everything (except the bus) in the visual field, such as roads, people, trees, buildings, on the retina will remain stationary. As soon as the man starts walking towards the bus, the entire retinal image will undergo global dynamic changes. Understanding the visual stability, perceiving a stable world despite fast and drastic changes in the retinotopic representations due to eye, head, and body movements, has been a quest since the investigations of the Persian scientist Alhazen (965 – 1040 AD) (see reviews: Bridgeman, Van der Heijden, & Velichkovsky, 1994; Cavanagh, Hunt, Afraz, & Rolfs, 2010).

The attempts to solve the first type of problems can be categorized in two groups: A single global compensation mechanism that applies a shift in the opposite direction of the motion vector to the entire retinotopic representation (Bischof & Kramer, 1968; Melcher & Colby, 2008; Sperry, 1950; von Holst & Mittelstaedt, 1971; von Holst, 1950; Wurtz, 2008), and local reference-object based mechanisms (Currie, McConkie, Carlson-Radvansky, & Irwin, 2000; Deubel, 2004) which generalize the stability of a few items whose post-movement locations are predictable, to the entire visual field. Note that even with a complete information about an ongoing or pending eye movement, such compensation mechanisms also require depth information since due to projective geometry, changes in retinal projections of objects in the environment depend on distance to the observer. Nevertheless, accumulating evidence shows that significant global changes in retinotopic receptive field structures of the neurons in many cortical and subcortical areas take place indicating the presence of some sort of compensation mechanism (Churan, Guitton, & Pack, 2011, 2012; Hamker, Zirnsak, Calow, & Lappe, 2008; Joiner, Cavanaugh, & Wurtz, 2011; Sommer & Wurtz, 2004; Zirnsak, Gerhards, Kiani, Lappe, & Hamker, 2011; Zirnsak, Lappe, & Hamker, 2010; Zirnsak, Steinmetz, Noudoost, Xu, & Moore, 2014).

The second type of problems is neither predictable nor global. The brain does not have any information *in advance* about the movements of external objects. The movements of external objects should be calculated on the fly during objects' motion and motion trajectories can be arbitrarily complex. Moreover, different objects can move simultaneously in different trajectories. Therefore, a single global compensation mechanism cannot account for all the changes due to objects' movement in the visual field. There must be mechanisms that process each object and its surroundings individually and independently from each other. Our approach is that the motion vector of each object serves as a local reference frame, and all necessary calculations are carried out according to this reference frame (Ogmen, 2007). In recent studies, experimental data supported this suggestion (Aydin et al., 2008; Aydin, Herzog, & Ogmen, 2009, 2011; Ogmen et al., 2006; Otto, Ogmen, & Herzog, 2008, 2009).

As a first innovation, we will combine anorthoscopic (slit) viewing (Parks, 1965; Rock, 1981; von Helmholtz, 1867; Zöllner, 1862) with the type of stimuli used in Ternus-Pikler display (Pikler, 1917; Ternus, 1926) to directly pit retinotopic and non-retinotopic representations against each other in processing stimulus features and correctly attributing them to objects. As a second innovation, we will use a variant of induced motion paradigm (Duncker, 1929; Johansson, 1975, 1976; Rock, Auster, Schiffman, & Wheeler, 1980) to explore the effects of non-retinotopic reference frames in a wide range of spatial locations.

In addition we will be able to modulate the local reference frames (formed by motion vectors) by varying the speed of a reference. The distinction between the two types of problems, the interpretation of the previous studies, and the pilot data led us to develop a qualitative theory. This theory can offer a unified solution to broad range of problems (e.g., motion blue, moving ghosts, occlusions, dynamic feature processing and attribution) hitherto given individual and independent accounts, which can be regarded as the third innovation.

Chapter 2. Non-Retinotopic Feature Processing in the Absence of Retinotopic Spatial Layout and the Construction of Perceptual Space from Motion

2.1. Introduction

The spatial representation of a visual scene in the early visual system is well known. First, the optics of the eye map the three-dimensional environment into two-dimensional images on the retina. The projections from retina to early visual areas preserve neighborhood relations to generate retinotopic representations of the stimulus (e.g., Sereno et al., 1995; Tootell et al., 1982). Retinotopic representations and processing are among the most prevalent concepts in current visual neuroscience. Most theories of vision involve computations where retinotopically-based receptive fields extract features, such as oriented boundaries, texture, color, etc., to synthesize the various attributes of the stimulus. On the other hand, it has long been known that a retinotopic representation of the stimulus is neither sufficient nor necessary for perception. In metacontrast masking (see reviews: Bachmann, 1994; Breitmeyer & Ogmen, 2000), a retinotopically non-overlapping “mask” stimulus can render a target stimulus completely invisible, showing that a retinotopic representation of the stimulus is not a sufficient condition for its perception. Anorthoscopic perception occurs when a stimulus moves behind a narrow slit, as shown in Fig. 2.1 (e.g., Morgan, Findlay, & Watt, 1982; Parks, 1965; Rock, 1981; Zöllner, 1862).

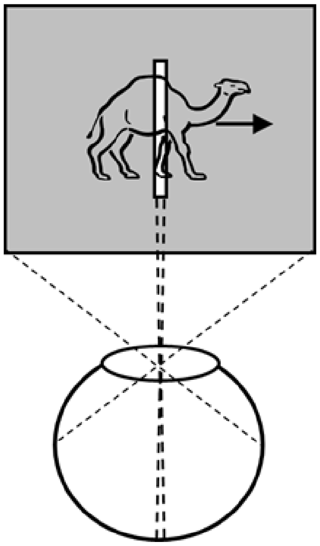

Figure 2.1. Anorthoscopic perception. A moving figure (the camel) is viewed behind a narrow slit. Observers report seeing a spatially extended coherent form instead of incoherently moving fragments confined into the narrow area inside the slit.

Under this viewing condition, all information about the stimulus falls onto the same narrow retinotopic strip over time. As a result, there is no spatially extended retinotopic representation of the stimulus. In other words, at any instant of time, the spatial layout of the stimulus does not have a complete retinotopic representation. However, observers report perceiving a spatially extended coherent form instead of a rapid succession of stimulus fragments confined into the area of the narrow slit (Fendrich, Rieger, & Heinze, 2005; Morgan et al., 1982; Parks, 1965; Rock, 1981; Zöllner, 1862).

On the other hand, one can appreciate the importance of retinotopic representations by considering tasks that are naturally suited for these representations. For example, a retinotopic representation provides a natural reference frame to drive eye movements. The retinotopic distance of a target from the fovea provides a direct error signal that can be used to position the fovea on a select target. Retinotopic representations also provide information about occlusions among stimuli and as such they can be used as a visibility map that distinguishes between modal and amodal percepts. Finally, basic computations in retinotopic representations, such as motion grouping, can provide the necessary transformational signals to build non-retinotopic representations.

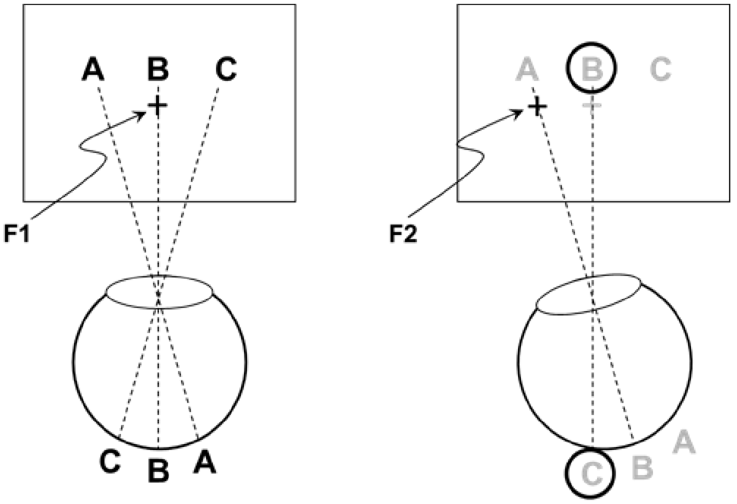

Figure 2.2. Retinotopic and spatiotopic representations differ in the way the two stimuli match: According to the retinotopic representation of the stimulus, the ring surrounds the letter C, while according to spatiotopic representation of the stimulus, the ring surrounds the letter B.

Given that retinotopic representations are neither necessary nor sufficient to support our perception, a fundamental question in vision science is to determine non-retinotopic bases of information processing in the visual system. The traditional experimental method to distinguish between retinotopic and non-retinotopic processing is the "Saccadic Stimulus Presentation Paradigm" (SSPP). Here, the observer is asked to make a saccade and two stimuli are presented, one before the saccade and the second after the saccade.

As shown in Fig. 2.2, retinotopic and spatiotopic representations differ in the way the two stimuli match: According to the retinotopic representation of the stimulus, the ring surrounds the letter C, while according to spatiotopic representation of the stimulus (e.g., a coordinate system based in space at the center of the display monitor), the ring surrounds the letter B. Although SSPP continues to be an informative paradigm in probing non-retinotopic representations and processes, it cannot be directly applied to cases where the eyes are stationary. This is because a variety of processes related to saccadic eye movements, such as saccadic suppression and efferent copy signaling, are active in SSPP. As a result, one cannot distinguish between these eye-movement related processes and other non-retinotopic processes that may be operating independently from eye movements. In order to investigate non-retinotopic processes in the absence of eye movements, we have introduced an alternative stimulus paradigm (Boi et al., 2009; Ogmen et al., 2006) based on the Ternus-Pikler display (Pikler, 1917; Ternus, 1926).

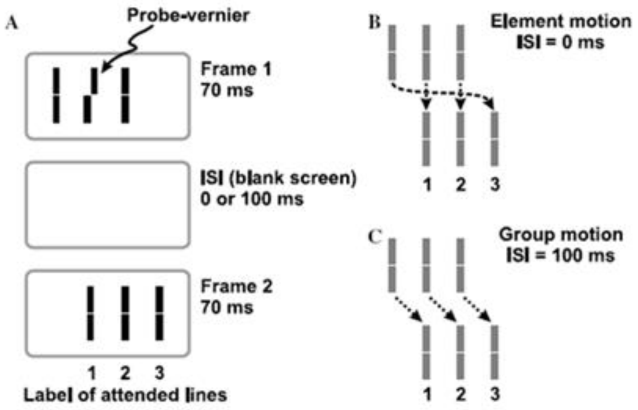

Figure 2.3. The Ternus-Pikler Display(A) and the associated percepts of "Element Motion" (B) and "Group Motion" (C). From (Ogmen et al., 2006).

Fig. 2.3A shows a basic Ternus-Pikler display. The first frame contains three elements. After an inter-stimulus interval (ISI), these three elements are shifted to the right by one inter-element distance so that two of the elements overlap retinotopically across the two frames. For short ISIs, observers report seeing the leftmost element of the first frame move to the rightmost element of the second frame, while the other two elements appear stationary (Fig. 2.3B). This percept is called “element motion”. For longer ISIs, all three elements move in tandem to the right as a group (Fig. 2.3C). This percept is called “group motion” (Pantle & Picciano, 1976). In order to study retinotopic versus non-retinotopic processes, we inserted a Vernier offset, called the “probe Vernier” to the central element of the first frame as shown in Fig. 2.3A. We asked observers to report the perceived offset direction for elements in the second frame, numbered 1, 2, and 3 in Fig. 2.3A. None of these elements contained a Vernier offset and naïve observers did not know where the probe Vernier was located. According to retinotopic relations, we expect the probe Vernier to be integrated with element 1 in the second frame for values of ISI that are within the window of temporal integration. On the other hand, if feature processing and integration take place according to motion grouping relations (Fig. 2.3B and C), instead of retinotopic relations, one would expect the probe Vernier to integrate with element 1 in the case of element motion and with element 2 in the case of group motion. Our results supported the predictions of the grouping-based non-retinotopic hypothesis (Fig. 2.4).

This grouping-based non-retinotopic feature processing provides strong support for the role of non-retinotopic mechanisms in processing dynamic stimuli. Based on this and several other studies using the Ternus-Pikler stimulus, we suggested that the visual system attributes features according to motion grouping relations across space and time (Boi et al., 2009; Ogmen & Herzog, 2010; Ogmen et al., 2006).

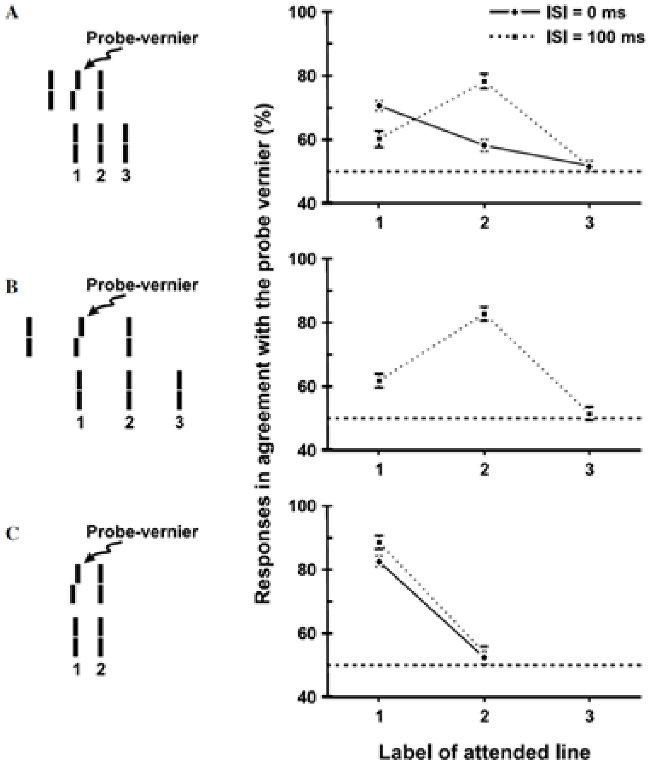

Figure 2.4. The percentage of responses in agreement with the probe Vernier as a function of the attended line for the Ternus-Pikler display shown in Fig. 2.3. From (Ogmen et al., 2006).

However, neither the SSPP nor the Ternus-Pikler paradigm completely rules out the contribution of retinotopic processing. In fact, in both SSPP and Ternus-Pikler displays, each stimulus frame contains a complete spatial layout of the stimulus: All elements and their spatial relations are represented in the retinotopic space. Retinotopic conflicts are created either by eye movements (SSPP) or by the retinotopically overlapping motion of the stimulus (Ternus-Pikler). Since retinotopic processes can take place during the presentation of each frame (e.g., retinotopic lateral interactions between elements may establish a spatial grouping of the elements), these paradigms do not completely eliminate potential contributions of retinotopic processes. In contrast, anorthoscopic viewing eliminates all retinotopic representations with the exception of a very narrow slit region. In addition to minimizing retinotopic representations, anorthoscopic stimuli also create retinotopic conflict in that different parts of the stimulus fall on the same retinotopic area (the interior of the slit) as the stimulus moves behind the slit. By using anorthoscopic viewing paradigm, the goal of this study was to investigate how stimulus features are processed and attributed in the absence of a retinotopic spatial layout and in the presence of retinotopic conflicts.

2.2. General Methods

2.2.1. Participants

Six observers (4 males) including the author participated in this study. The age of the participants ranged from 24 to 37 years and all participants had normal or corrected-to-normal vision. Experiments followed a protocol approved by the University of Houston Committee for the Protection of Human Subjects. Each observer gave written consent before the experiments and volunteers were paid $10/hour for their participation.

2.2.2. Apparatus and Stimuli

Visual stimuli were created via visual stimulus generator card VSG2/3 (Cambridge Research Systems) and displayed at a resolution of 800x600 with a refresh rate of 100 Hz. A head/chin rest was used. Observers reported their responses by a joystick. The stimulus configurations used are shown in Fig. 2.5. These stimuli were presented behind a narrow slit as in Fig. 2.1. The luminance of the opaque region was 20cd/m2 while the slit area was 40 cd/m2. The stimuli were black (1 cd/m2). The vertical and horizontal dimensions of the slit were 7.1° and 17’ (visual degrees and minutes of arc will be represented by ° and ’, respectively, in the remainder of this chapter), respectively. For each condition, the display consisted of three elements, each having two abutting lines with a small (6.2’) vertical gap between them. The size of each abutting line was 38.7’ for the longest element, 29.4’ for the medium sized element and 20.1’ for the shortest. The thickness of each line was 7.7’ and inter element distance was 51’. In each trial, a Vernier offset was inserted randomly to one of the elements (probe Vernier). The direction of the offset, i.e., position of the lower segment of a given element with respect to the upper one, was also random in each trial. The experiments consisted of “Connected Lines”, “Basic Display”, and the “Flash” conditions. The rationale of stimulus design for each of these conditions and their predictions are discussed in the next section. Vernier offset sizes were selected to obtain about 85% or higher correct responses when the Vernier offset was inserted to the element attended by the subject. This yielded a Vernier offset of 6.2’ for “Connected Lines” and “Basic Display” conditions and 3.1’ for the “Flash” condition. The speed of the stimulus behind the slit was 5.1°/sec. In the “Flash” condition, the duration of the flash period (70 ms) for each element and the inter-stimulus interval (100 ms) between the flashes were chosen to mimic the timings in the other two conditions. The direction of motion (rightward or leftward) was randomized from trial to trial. In all conditions, the shortest element (Element 1 in Fig. 2.5) was always presented first inside the slit; in other words, the configurations shown in Fig. 2.5 were used for rightward motion and their horizontally mirror-symmetric versions were used for leftward motion.

2.2.3. Procedures

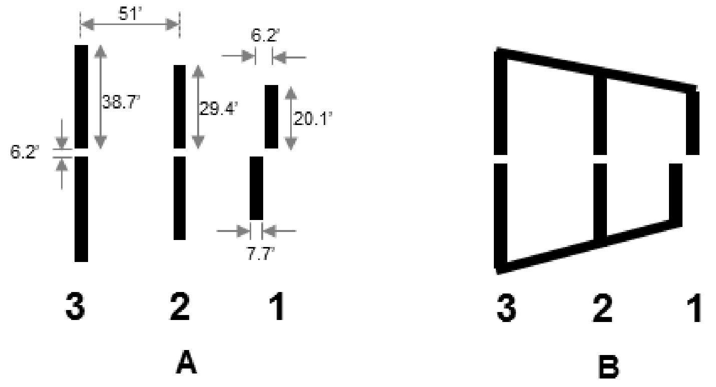

Figure 2.5. Stimulus configurations. (A) "Basic Display" and (B) "Connected Lines" conditions. These two stimuli are identical with the exception of the oblique lines connecting the three vertical line segments. A probe Vernier was inserted in one of the elements.

The task of the observers was to attend to a given element and report the perceived direction of the Vernier offset for that element by pressing a button. Observers attended to the same pre-determined element within a given block. Performance was evaluated as the percentage of responses in agreement with the actual direction of the probe Vernier, regardless of which element contained the probe Vernier. The next trial started by pressing another button. There were 120 trials in a block. Within each block, the same stimulus configuration was used and the number of trials in which the probe Vernier is assigned to a particular element was 40. All conditions were presented three times and the order of conditions was randomized. Training sessions were run before the actual experiment in order to familiarize the observers with the equipment and the task. The results of training sessions were not included in the data analysis.

2.3. Rationale of Stimulus Design and Predictions

We have designed three stimulus conditions based on the degree to which they produced an anorthoscopic percept. Motion information is crucial in generating anorthoscopic percepts (e.g., Aydin et al., 2008; McCloskey & Watkins, 1978; Morgan et al., 1982; Rieger & Grüschow, 2007; Shimojo & Richards, 1986). A stationary stimulus behind the slit, or a stimulus which is flashed behind the slit, generate a percept of stimulus fragments confined to the retinotopic area of the slit. On the other hand, when the stimulus moves behind the slit, a process akin to amodal completion (Michotte, Thinès, & Crabbé, 1964) takes place, where different parts of the stimulus are perceived in a spatially integrated form revealing the complete spatial layout of the stimulus.

Our first condition, the “Flash” condition, was a baseline control condition in that no anorthoscopic percept was generated: We flashed the three line segments shown in Fig. 2.5A, centered in the region of the slit one after the other. The temporal order of presentation was from the shortest to the longest segment.

Our second condition, the “Connecting Lines” condition was designed to generate a strong anorthoscopic percept. Here, the three line segments were connected by oblique lines to generate the spatial layout shown in Fig. 2.5B. This figure was moved behind the slit. Having line segments of different vertical length allowed us to connect them with oblique rather than horizontal lines. The advantage of an oblique over a horizontal line comes from the constraint imposed by the slit: If a horizontal line is moved behind a vertical slit, motion signals will be generated only when the line enters or leaves the slit. On the other hand, an oblique line generates a motion signal during all time instants when it moves behind the slit. As suggested by previous research (e.g., Sohmiya & Sohmiya, 1994) composite figures, such as our “Connecting Lines” configuration, that generate both vertical and horizontal motion signals are best suited for generating strong anorthoscopic percepts.

Finally, a third condition, the “Basic Display” condition was created as an intermediate between the previous two conditions. As shown in Fig. 2.5A, the stimulus was identical to that of Fig. 2.5B with the exception of the removal of oblique connecting lines. When this stimulus was moved behind the slit, the continuous motion signal generated by oblique lines was no longer present. Instead, only a brief horizontal motion signal was generated when each line segment passed through the slit. Thus anorthoscopic percept was either absent or very weak. In this sense, this condition is similar to the baseline “Flash Condition” with the exception that, the line segments moved in a way identical to the “Connecting Lines” conditions instead of being flashed. Our predictions were that, if present, non-retinotopic processes should occur prominently in the “Connecting Lines” condition but not in the “Flash” condition, where no anorthoscopic percept is generated. In the “Basic Display” condition, anorthoscopic percept is either weak or absent, therefore we expected weak, if any, non-retinotopic processing.

To measure non-retinotopic processes, we assessed the perception of the Vernier offset, as it was done in our previous study using the Ternus-Pikler paradigm (Ogmen et al., 2006) discussed in the preceding section. Predictions of the retinotopic and non-retinotopic feature processing and integrations are depicted in Fig. 2.6 for the stimuli used in this study. According to the retinotopic hypothesis, Vernier offset of a stimulus should be integrated with those stimuli that are presented at the same retinotopic region within the temporal integration period (ca. 120ms) (Coltheart, 1980). The integration is graded, strongest for short ISIs and weaker for longer ISIs. Accordingly, the prediction is that the accordance of observers' responses with the Vernier offset direction will be highest when the observers attend to the location of the element containing the probe Vernier and should be lower but significantly better than chance for other elements presented within approximately 120ms of the probe Vernier (Fig. 2.6A). The agreement between observers' responses and the probe Vernier should be at chance for elements presented outside the temporal integration window. The retinotopic processing hypothesis predicts that the same pattern of results will be obtained in all three conditions ("Connecting Lines", "Basic Display", and "Flash"). Note that in our depictions of the predictions, for simplicity, we assumed an equal level of performance for the attended line when the physical location of the probe Vernier is 1, 2, or 3. Moreover, when Element 2 is attended, we assumed an equally effective integration for the cases where the probe Vernier is located in Element 1 and Element 3. However, these assumptions may not hold since the Elements 1, 2, and 3 had different lengths and the effectiveness of integration may be influenced by temporal order (Scharnowski, Hermens, & Herzog, 2007). It is difficult to quantify a priori the potential contributions of these factors for our stimuli. Nevertheless, as we will discuss below, the exact shapes of these functions are not critical in comparing the predictions of the retinotopic versus non-retinotopic hypotheses.

According to non-retinotopic processing hypothesis, retinotopic integration should be prevented when the anorthoscopic percept indicates that the elements are spatially distinct parts of a spatially extended figure. In this case, the accordance of the observers' responses with the direction of the probe Vernier should be high when observers attend the element containing the probe Vernier and it should be at chance for all other elements (Fig. 2.6B). Since a strong anorthoscopic percept is generated in the "Connecting Lines" condition, the non-retinotopic hypothesis predicts the outcome shown in Fig. 2.6B for this condition. For the "Flash" condition it predicts the pattern shown in Fig. 2.6A since no anorthoscopic percept is generated in this case. Moreover, since the anorthoscopic percept is weak or absent in the "Basic Display" condition, a pattern more similar to that in Fig. 2.6A than that in Fig. 2.6B is predicted. Therefore, the critical comparison for our study is to determine whether the "Connecting Lines" condition is significantly different from the other two conditions and whether the pattern of results for this condition is more similar to Fig. 2.6B compared to Fig. 2.6A

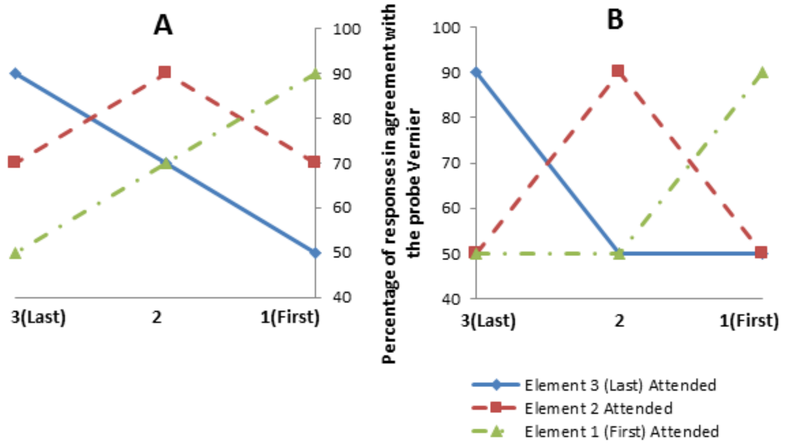

Figure 2.6. Predictions of the retinotopic and non-retinotopic feature processing hypotheses. The non-retinotopic hypothesis predicts the pattern shown in (A) for the "Flash" condition and the pattern shown in (B) for the "Connecting Lines" condition.

2.4. Results

Results for all conditions are given in Fig. 2.7. Percentages of responses in accordance with the actual direction of the Vernier offsets are plotted against the physical location of the Vernier offsets. Different line styles (colors) represent different attended lines. Both retinotopic and non-retinotopic processing hypotheses predict that the pattern of results for “Flash” and “Basic Display” conditions should be similar to the one shown in Fig. 2.6A. For these two conditions, when Element 3 (the longest element which always comes last in the sequence) is attended, we observe a decreasing performance as a function of the physical location of the probe Vernier. Similarly, when Element 1 (the shortest element that always appears first) is attended, we observe an increasing performance as a function of the physical location of the probe Vernier, in agreement with the qualitative predictions depicted in Fig. 2.6A. When Element 2 is attended, the resulting function is not an inverse-V function, as depicted in Fig. 2.6A, possibly due to the involvement of factors that may be influenced by the size and temporal order of the stimuli.

The critical comparison to distinguish between the retinotopic and non-retinotopic feature processing hypotheses is the comparison of the results for “Connected Lines” condition to Fig. 2.6A (retinotopic hypothesis prediction) versus Fig. 2.6B (non-retinotopic hypothesis prediction). Visual inspection suggests that the “Connected Lines” results are much more similar to Fig. 2.6B than to Fig. 2.6A and therefore supports the non-retinotopic hypothesis. A repeated-measures ANOVA with conditions (Flash, Basic Display and Connecting Lines), attended lines and physical locations of the probe Vernier as main factors showed that the physical location of the probe Vernier and the location of the attended line have no significant main effects ($F(2,10) = 0.125$, $p = 0.884$ and $F(2,10) = 0.937$, $p = 0.424$, respectively). However, as can be seen from Fig. 2.7, there is a significant interaction between the physical location of the probe Vernier and the location of the attended line ($F(4,20) = 111.525$, $p < 0.0001$). As mentioned in the previous section, the retinotopic hypothesis predicts that the results will be essentially the same for all three conditions while the non-retinotopic hypothesis predicts that the results for the “Flash” and “Basic Display” conditions should be similar to each other but different from the “Connecting Lines” condition. However, a repeated-measures ANOVA showed that stimulus condition (“Flash”, “Basic Display”, and “Connecting Lines”) does not have a significant main effect (F(2,10) = 2.274, p = 0.153). This is not an unexpected result since ANOVA lumps all points in each plot shown in Fig. 2.7 and compares these averaged values. According to both retinotopic and non-retinotopic hypotheses, averages should be similar, if not equal.

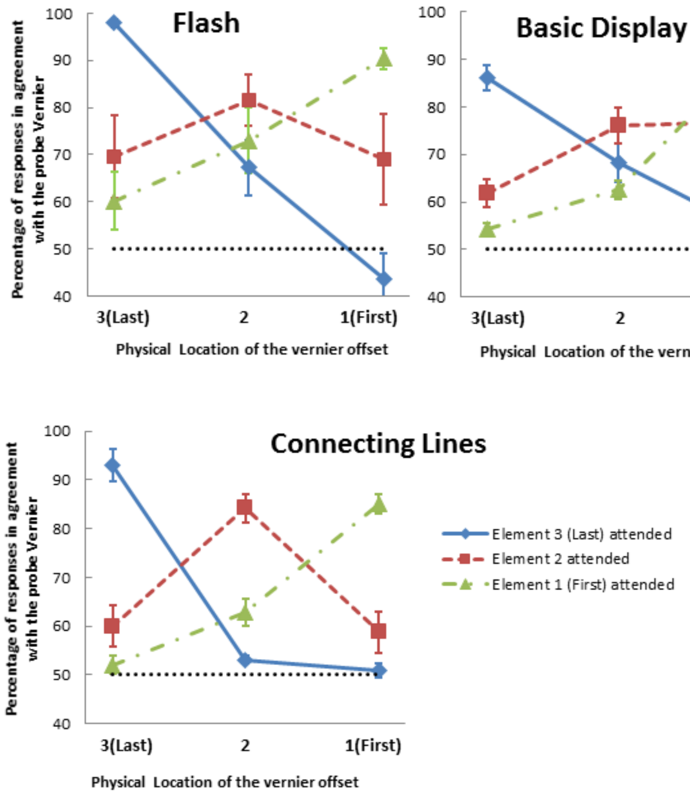

Figure 2.7. Experimental results for all stimulus configurations. The horizontal axes represent the actual physical location of the probe Vernier and the vertical axes show the percentage of responses in agreement with the probe Vernier. Error bars represent ±SEM (N=6).

An alternative way to investigate the differences among conditions is comparing the "relative retinotopy" for points at which retinotopic and non-retinotopic hypotheses predict different performances (compare Fig. 2.6A and 2.6B). There are four points (combination of factors) in each condition that meet this criterion: a) When observers attended to either Element 1 or Element 3 while the probe Vernier was at Element 2, b) When Element 2 was attended while either Element 1 or Element 3 carried the probe Vernier. The deviation of performance from chance level at these points signals the existence of residual retinotopic integration. Moreover, performance when an element carries the probe Vernier is attended (peak points in Fig. 2.7) depends on the temporal order of the element and visual masking that this element undergoes resulting in different peak values in Fig. 2.7. For instance, observers consistently perform slightly worse when they attend Element 2 while it also carried the probe Vernier (red dashed lines in Fig. 2.7) than when they attend Element 1 or Element 3 and probe Vernier is at the same element. Therefore we calculated the relative retinotopy as the ratio of deviation of performance from chance level and the difference of peak performance from chance level for that specific physical location of probe Vernier. Fig. 2.8 shows relative retinotopies for all conditions. As predicted by the non-retinotopic hypothesis, retinotopic integration is very effective in the Flash condition and it is reduced by the anorthoscopic percept in the Connecting Lines condition. A one-way repeated measures ANOVA shows a significant effect of condition on retinotopic integration ($F(2,10) = 5.728$, $p = 0.022$). Post-hoc multiple comparisons (without any correction) showed that Basic Display and Connecting Lines conditions are significantly different (mean difference = 0.330, std. error = 0.088, $p = 0.014$) and the difference between Flash and Connecting Lines conditions is only marginally significant (mean difference = 0.302, std. error = 0.119, p = 0.052). Considering the spatiotemporal layout of the stimulus display in this experiment, the difference between Basic Display and Connecting Lines conditions stress prominently the effect of non-retinotopic representations. In these conditions, everything else was equal except the two oblique lines that enhance the anorthoscopic percept. On the other hand, as opposed to predictions of the pure non-retinotopic hypothesis, there is some residual retinotopy left in the Connecting Lines condition. This is in agreement with our previous experiments which also showed residual retinotopy. This can be seen, for example, in the data from (Ogmen et al., 2006) study reproduced in Figure 2.4. The percentages of responses in agreement with the probe Vernier for Element 1 are larger than 50% (top and middle panels in Fig. 2.4).

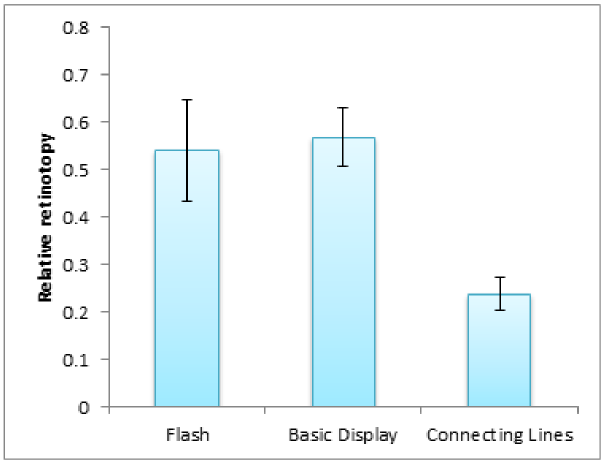

Figure 2.8. Relative retinotopies are given for all experimental conditions. Error bars represent ±SEM across observers (N=6).

2.5. Discussion

Retinotopic organization of the early visual system is well established. Most theories of vision are based on this retinotopic organization: Features are extracted through retinotopically localized receptive fields of neurons and these features are integrated through connections between neurons positioned at various retinotopic locations. However, as discussed in the introduction section, metacontrast masking and anorthoscopic perception demonstrate that retinotopic representations are neither sufficient nor necessary for the perception of spatially extended form. Several recent studies provided evidence for non-retinotopic bases for the computation of stimulus features such as form (Ogmen et al., 2006), color (Nishida et al., 2007), luminance (Shimozaki et al., 1999), size (Kawabe, 2008), motion (Boi et al., 2009; Melcher & Morrone, 2003), and position (Fischer, Spotswood, & Whitney, 2011). However, in all these studies retinotopic representations of the stimuli were available for visual processing. As a consequence, although these studies showed non-retinotopic processing, they did not completely exclude contributions of retinotopic processes. In this study, by using anorthoscopic perception, we minimized the retinotopic stimulus to a very narrow area corresponding to the inside region of the slit. Under this condition, we tested whether simple feature processing would follow retinotopic or non-retinotopic rules. Our results support the operation of non-retinotopic processes even when retinotopic representations are minimized.

To probe the broader significance of these findings, one may pose two inter-related questions:

1) If retinotopic representations are not necessary for form perception, why is the early visual system organized retinotopically?

2) If the early visual system is organized retinotopically, why does the visual system use non-retinotopic processes and representations?

We suggest that retinotopic organization emerges, on the one hand, from the physics of image formation apparatus and, on the other hand, from the fact that spatial contiguity is a common property of objects in our environment. Based on the physics of image formation, the retinotopic representations can be used to control eye movements by providing error signals of targets with respect to fovea. The spatial contiguity of physical objects and the attendant neighborhood relations are preserved by the optical characteristics of the eyes. In retinotopic representations, breakdown of neighborhood relations often correspond to occlusions; as a result retinotopic representations can be used as a visibility map to distinguish between modal and amodal percepts. However, the visual system is forced to transform these retinotopic representations into non-retinotopic ones as a result of ecological constraints. Under normal viewing conditions, our eyes undergo frequent movements making retinotopic representations highly unstable. Moreover, many objects in the environment are in motion and, given the visible persistence of vision, these objects would generate highly blurred percepts if the computations were to take place in retinotopic representations (Ogmen, 2007). Finally, occlusions are ubiquitous in our environment and occlusions often blank out significant parts of retinotopic representations both in a static and dynamic manner. Therefore the visual system needs to create representations where objects can be processed and represented in a way that is invariant with respect to dynamic changes and occlusions. The achievement of object continuity and permanence is a significant milestone in cognitive development (Piaget & Inhelder, 1969) and indicates the central role non-retinotopic processes play both in perception and cognition.

Anorthoscopic perception provides a window through which one can investigate how these non-retinotopic representations are constructed in the visual system. In particular, how is the dimension of space constructed to allow spatially extended percepts from stimulations that fall in a very limited retinotopic region? “Retinal painting” was one of the first hypotheses to explain anorthoscopic percepts (von Helmholtz, 1867), The explanation was based on the assumption that eye movements enable the spatial expansion of stimuli from the slit region to a broader retinotopic region: If the eyes move while viewing the stimulus, then successive parts of the stimulus fall on adjacent retinotopic loci thereby “painting” a retinotopic spatially extended picture of the figure. However, several studies showed that, while retinal painting can occur when the eyes move in a coordinated fashion, it cannot explain anorthoscopic perception in general: Measurement of eye movements and studies using retinal stabilization provide evidence that anorthoscopic perception does also occur without the contribution of eye movements (Fendrich et al., 2005; Morgan et al., 1982).

Another theory, proposed by Parks (1965), suggests that a post-retinal mechanism stores in memory the information available through the slit and reconstructs the spatial layout of the figure according to a “time-of-arrival coding.” Fig. 2.9 shows a stimulus which was used to test retinal painting and the time-of-arrival reconstruction theories (McCloskey & Watkins, 1978; Sohmiya & Sohmiya, 1992, 1994). The stimulus consists of two triangular shapes made up of dots moving in opposite directions. The tips of the triangles pass through the slit simultaneously, followed by the middle segments and finally the longest segments. Since the tips, the middle segments and the bases of the two triangles arrive to the slit region at the same time instants, the time-of-arrival coding theory predicts that observers will perceive two identical, instead of mirror image, triangles. Similarly, since eye movements affect the two triangles exactly the same way, the retinal painting theory makes the same prediction. In contradiction to this prediction, observers' perception corresponds to the actual stimulus configuration, i.e., two mirror image triangles (McCloskey & Watkins, 1978; Sohmiya & Sohmiya, 1992, 1994). Therefore, the space dimension is created neither through eye movements nor through a transformation of time dimension. Instead, the direction of motion information is critical in constructing spatial representations. If the direction of motion is not known, the stimulus is ambiguous in that a leftward moving image and its mirror-symmetric version moving rightward generate identical patterns in the slit. Therefore, the determination of the direction of motion is critical for anorthoscopic perception. We have proposed that the space dimension is created from motion dimension (Ogmen, 2007). This proposal has been supported by a recent study where spatial distortions in anorthoscopic perception were examined (Aydin et al., 2008). The figure moving behind the slit typically appears spatially compressed in the direction of its motion (e.g., Anstis & Atkinson, 1967; Aydin et al., 2009; Haber & Nathanson, 1968; McCloskey & Watkins, 1978; Morgan et al., 1982b; Parks, 1965; Rock & Sigman, 1973; Rock, 1981; von Helmholtz, 1867; Zöllner, 1862). Aydin et al., (2008) showed that this spatial compression can be explained by the differences in the perceived speeds of the leading and trailing parts of the anorthoscopic figure. If indeed, space is constructed from motion, one would expect the spatial metric (perceived distances and sizes) to depend on the motion metric (perceived speeds).

2.6. Conclusion

Taken together, these studies suggest that to process and represent information about the environment during ecological viewing conditions, the visual system uses non-retinotopic representations where space is constructed from motion. While the single slit configuration used in this study may appear an extreme form of occlusion, it has also been demonstrated in the case of multiple slits that spatial form is constructed by using motion signals (Nishida, 2004). As shown in the current study, the visual system is also capable of preserving figural, as opposed to purely retinotopic, relationships as features are processed in these motion-based non-retinotopic representations.

Chapter 3. The Effective Reference Frame for Motion Perception

3.1. Introduction

When an object moves during steady fixation, its projection on the retina also moves at a speed proportional to its physical speed. The perceptual system readily interprets this retinal motion as the motion of an object in the environment. However, when the observer’s eyes, head or body move, the retinal image motion does not directly correspond to a corresponding motion in the environment. In order to perceive veridically the motion of an object in the environment, the perceptual system needs to carry out coordinate transformations (Swanston, Wade, & Day, 1987; Wade & Swanston, 1987). In other words, the retinal motion due to self-motion or movement of the eyes need to be parsed out such that what is left directly corresponds to the motion of an object in the environment. Gibson argued that optic flow alone is sufficient to make the required transformations and to decompose retinal motion into self-motion and object motion relative to the scene (Gibson, 1979). Many psychophysical (e.g., Rushton, Bradshaw, & Warren, 2007; Warren & Rushton, 2009), neurophysiological (Duffy & Wurtz, 1991a, 1991b), functional imaging (e.g., Morrone et al., 2000), and modeling (Furman & Gur, 2003; Pack, Grossberg, & Mingolla, 2001) studies supported his position. However, early studies of motion perception during smooth-pursuit eye movements showed that the coordinate transform from retinocentric reference frame to head-centric one is not perfect. A stationary object is perceived to be moving in the direction opposite to the direction of the ongoing pursuit eye-movement (Filehne illusion. Filehne, 1922; Freeman & Banks, 1998; Mack & Herman, 1972, 1973; Wertheim, 1987) and a moving object is perceived to be slower when it is tracked than when it is viewed during fixation (Aubert-Fleischl effect. Aubert, 1886; Fleischl, 1882; Freeman & Banks, 1998). The perceived direction and the extent of motion of an object that moves non-collinearly with the pursuit target significantly deviate from corresponding physical quantities (Becklen, Wallach, & Nitzberg, 1984; Festinger, Sedgwick, & Holtzman, 1976; Furman & Gur, 2005; Kano & Hayashi, 1981; Souman, Hooge, & Wertheim, 2005, 2006b). Assuming perfect retinal gains (i.e., the ratio of perceived and actual retinal motion extents or speeds is one), these perceptual errors and illusions have been conventionally attributed to an under-registration of eye velocities. However, perceived retinal motion is strongly modulated by stimulus properties such as spatial frequency, dot density, contrast, stimulus scale and chromatic content (see review by Nishida, 2011) and hence, errors in estimating retinal motion should also be considered in the computations of head-centric motions (Freeman & Banks, 1998).

Many models of motion perception during smooth pursuit have been proposed to quantify the degree to which this coordinate transformation is complete. In most of these models, the observer's head and body are assumed to be stationary with respect to the outside world, and the perceived head-centered motion is a combination of retinal motion and eye velocity estimates (Freeman & Banks, 1998; Freeman, 2001; Souman, Hooge, & Wertheim, 2006a; Swanston et al., 1987; Turano & Massof, 2001; Wertheim, 1994). Models with non-linear motion transducers have been shown to perform slightly better than those with linear motion estimators for both terms (Freeman, 2001; Turano & Massof, 2001). The estimated eye velocity in some of these models is a function of both retinal and extra-retinal signals, whereas retinal motion estimates depend only on stimulus parameters and retinal motion itself (Freeman & Banks, 1998; Turano & Massof, 2001; Wertheim, 1994). Several studies have concluded that perceived motion during pursuit also depends on stimulus parameters including size (Turano & Heidenreich, 1999), spatial frequency (Freeman & Banks, 1998; Wertheim, 1994), speed (Pola & Wyatt, 1989; Turano & Heidenreich, 1996), and presentation duration (Mack & Herman, 1978; Souman et al., 2005; Wertheim, 1987).

When there are two objects in the scene and one of them is tracked, the relative motion between the objects may become a major determinant of perceived motion. In some studies, this fact was overlooked and the failure to discriminate relative motion from retinal motion led some researchers to conclude that the perceptual system has very weak (i.e., gains < 0.1) or no information at all about the ongoing pursuit eye movement (Dodge, 1904; Festinger et al., 1976; Stoper, 1973). For instance, when a small dot is pursued in a dark room and the motion of another (moving or stationary) dot is judged, the retinal motion of the target dot and its relative motion with respect to the pursuit dot are almost identical (assuming perfect smooth pursuit). It is impossible to decouple contributions of the retinal and relative motions in these displays. In fact, Mack and Herman (1978) showed that the relative motion between the pursuit target and the background object is one of the main factors influencing perceived motion. The contribution of the relative motion between the pursuit target and the background has been noted in several studies (Baker & Braddick, 1982; Brenner & van den Berg, 1994; Freeman, Champion, Sumnall, & Snowden, 2009; Freeman, Champion, & Warren, 2010; Hisakata, Terao, & Murakami, 2013; Mack & Herman, 1978; Mateeff, Hohnsbein, & Ehrenstein, 1990; Snowden, 1992; Turano & Heidenreich, 1999; Wallach, O'Leary, & McMahon, 1982; Wallach, 1959). In these studies, qualitative descriptions of how and when relative motion between the pursuit target and the background affects perceived motion have been given. Baker and Braddick (1982) argued that, at slow speeds, relative motion determines percepts whereas at high speeds, absolute motion (i.e., motion with respect to a spatiotopic reference frame such as stimulus display) takes over. Mack and Herman (1978) concluded that object-relative motion is only effective when the object of interest is in close proximity of the pursuit target. Brenner and van den Berg (1994) reported that the perceived target velocity does not change as long as the relative motion of the pursuit target with respect to a textured background is kept fixed.

When there are multiple moving objects in the scene, a typical scenario in normal viewing conditions, relative motions of these objects can fully determine the perceived motion. Duncker (1929) used displays generated by point-lights attached to an otherwise invisible rotating and translating circular cardboard (Duncker, 1929, pg. 240). When a point-light is attached to the rim of the cardboard, observers perceive cycloidal motion of the light, which corresponds to its trajectory on the retina if the observer's eyes are stationary. Percepts do not change when the point-light is tracked. However, when another point-light is added to the hub of the wheel, the central light is perceived to be translating linearly, whereas the peripheral light is perceived as rotating around the central light, regardless of whether the central light is tracked or not. In the latter case, the retinal trajectory of the point-light at the rim is again a cycloid but the percepts dominantly correspond to its relative motion with respect to the central light. Similar and more complex demonstrations of the superiority of relative motion were done by Johansson (Johansson, 1950, 1973). In line with this, it has been shown that the thresholds for detecting relative motion is much less than those for absolute motion (Snowden, 1992). Moreover, the movements of the eyes, head or body result in relative motions of objects at different depths in the environment.

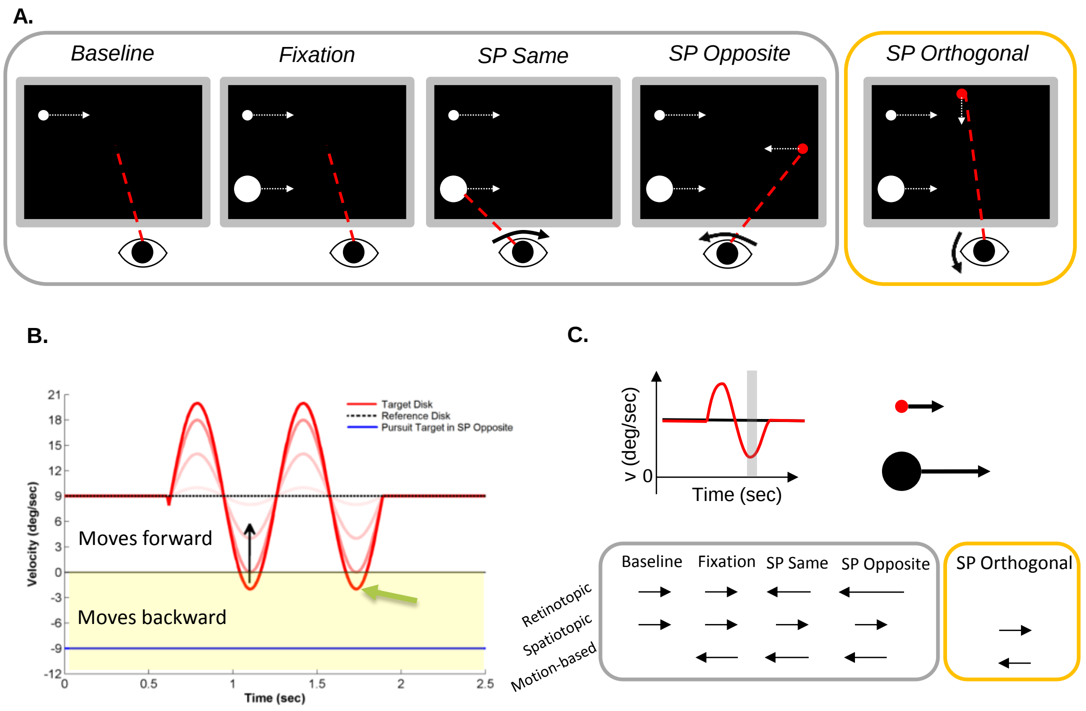

Figure 3.1. Spatial and temporal characteristics of the stimuli. A. Experimental conditions. B. Horizontal velocity profiles of the target and reference disks. C. Instantaneous horizontal velocity vectors (bottom) of the target disk shown for an example motion profile (top-left).

A complete theory of motion perception, therefore, must take into account the relative motion of objects with respect to each other. Wade and Swanston's quantitative model of motion perception (Wade & Swanston, 1987) explicitly includes a term for relative motion of objects with respect to each other. According to their model, the registered retinal motion undergoes a sequence of coordinate transforms to reach a geocentric representation. Estimated retinal motions are compensated for estimated eye movements at the orbital level, and the output of this process is combined with the "pattern-centric" signals (i.e., relative motion). Furthermore, they proposed that the two signals are not treated equally, but each has a weight. A similar approach was taken by Gogel (1977). He also argued that the relative motion has a greater weight compared to the other components (Gogel, 1977). Unfortunately, the weights of different terms have never been determined experimentally.

In contrast to the models of motion perception mentioned so far, we adopted a top-down approach and modeled the perceived motion as an interplay between various reference frames available to the perceptual system. By doing so, we remained agnostic as to how coordinate transforms outlined by previous models take place; instead, we sought to investigate how the perceptual system forms the “effective reference frame”. Let us assume that the head is kept still and two objects are moving in the fronto-parallel plane at different velocities. The perceived motion of each object depends on its motion on the retina (i.e., retinocentric or retinotopic reference frame), its motion on the display (i.e., space-centric or spatiotopic reference frame), and its motion relative to the motion of the other object (i.e., object-based or motion-based reference frame). The proposed model is given by

$$Perceived\ Motion = w_s(d,\varphi)P_s + w_r(d,\varphi)P_r + w_{mb}(d,\varphi)P_{mb} + c, \qquad (1)$$

where P_s, P_r, and P_{mb} represent the motion signals on spatiotopic, retinotopic and motion-based reference frames, and w_s, w_r, and w_{mb} represent the weights of each reference frame, respectively. The constant term c in the model captures the response bias of observers. The response bias represents byproducts of decision processes. Each P value represents also the predicted perceived-motion from a given reference frame. For instance, if observers perceive the motion direction solely based on retinal motion, (i.e., $w_s = 0$,

$w_{mb} = 0$, and $w_r = 1$), perceived motion would be equal to P_r. Note that each weight is modeled as a function of distance d between the two objects and some other potential factors φ (such as perceptual groupings, stimulus scale, attention, etc.).

Eqn. (8) contains four unknowns, namely the three weights and the constant term. In order to have a unique solution, at least four linearly independent equations (i.e., different combinations of P_s, P_r, and P_{mb} values) are needed. To this end, we designed four experimental conditions where observers judged whether or not a horizontally moving small (target) disk reversed its direction of motion (see Methods for details). Our stimulus is depicted in Fig. 3.1A. One small (target) and one large (reference) disk moved horizontally. The large disk moved with a constant velocity while the velocity of the small disk changed sinusoidally during the mid-course of its trajectory (Fig. 3.1B). The average velocity of the small disk was equal to the velocity of the large disk so as to keep the average distance between the two disks constant. This constraint was introduced because the distance between objects is known to affect the extent to which relative motion is perceived (Gogel & Koslow, 1972; Gogel, 1974; Mack & Herman, 1978; Mateeff & Hohnsbein, 1989; Mori, 1979; Shum & Wolford, 1983). In several studies of relative motion perception, stimuli consisted of objects with different velocities. As a consequence, the distance between the objects changed during the stimulation period. This resulted in a confound between the distance effect and the relative-motion effect (Festinger et al., 1976; Mack & Herman, 1978; Stoper, 1973). This issue becomes even more severe when the two objects move non-collinearly. To avoid this confound, we kept the *average* distance between the two objects constant.

A variety of stimuli, pursuit conditions, and tasks have been used in different studies to investigate perceived motion during eye movements. It is known that stimulus parameters such as size, speed, task, and attention affect performance in speed judgments (Baker & Braddick, 1982; Freeman et al., 2009; Gogel & Sharkey, 1989; Mateeff et al., 1990). Therefore, it is difficult to compare and reconcile results of different studies. Here, we investigated various pursuit conditions with the same task, stimuli and procedures as described below.

The rationale of our experimental conditions was based on four considerations: First, we wanted to quantify the contributions of different reference frames in determining the effective reference frame for motion perception. In order to dissociate between retinotopic and spatiotopic reference frames, we included conditions with and without eye movements. Second, we wanted to capture the quantitative contributions of different reference frames into a simple mathematical model, as given in Eqn. (1). Since the model in Eqn. (1) has four free parameters, four linearly independent equations are needed to estimate these parameters. Hence, we designed four conditions as follows: In the Baseline condition, a target disk was viewed during steady fixation (Fig. 3.1A, Baseline). In the Fixation condition (Fig. 3.1A, Fixation), another moving disk was also presented while observers kept fixation at the center of the display. In the Smooth Pursuit (SP) Same and SP Opposite conditions (Fig. 3.1A, SP Same and SP Opposite), observers tracked with smooth pursuit eye-movements another disk moving either in the same or in the opposite direction of the target disk, respectively. Third, we also varied the distance between the horizontally moving disks to characterize the weights in Eqn. (1) as a function of d to address previous accounts of distance dependent effects (Gogel, 1974; Mack & Herman, 1978). Finally, we wanted to test the predictive ability of the quantitative model. For this purpose, we have included a fifth condition, SP Orthogonal (Fig. 3.1A, SP Orthogonal), where observers pursued an additional vertically moving object with the goal of testing whether the weights obtained from the results of the first four conditions can be generalized to this new condition.

Fig. 3.1B illustrates the velocity profiles used in the experiments and in Fig. 3.1C we provide a specific example of velocity profiles for the target and reference disks, along with their horizontal instantaneous motion vectors on each reference frame when the target disk reaches its minimum velocity (gray shaded area, Fig. 3.1C top-left). Top-left plot in Fig. 3.1C shows the horizontal velocity profiles of the target (solid trace) and reference (dashed trace) disks. On the right, critical motion vectors corresponding to the time at which the target disk's velocity reaches its minimum are shown, with the target disk moving in the same direction as the reference disk but at a lower speed. At the bottom of panel C, the horizontal component of target's instantaneous velocity vector is shown according to different reference frame in each of the five experimental conditions. The spatiotopic motion is identical across all conditions. In the Baseline condition, since only the target is presented, there is no relative motion. However, in all other conditions, the relative motion of the target with respect to the reference disk is in the opposite direction of spatiotopic motion (in this particular example). The retinotopic motion is equal to the spatiotopic motion when observers fixate at the center (i.e., in the Baseline and Fixation conditions). It becomes equal to the relative motion in the SP Same condition whereas the difference between the relative and spatiotopic motions determines its value in the SP Opposite condition. Finally, since observers track a vertically moving object in the SP

Orthogonal condition, the *horizontal component* of the retinotopic motion is identical to the horizontal component of retinotopic motion in Baseline and Fixation conditions.

For the particular example described so far, if observers' perception is based solely on the spatiotopic reference frame, they would perceive forward movement for the target disk regardless of condition and report that the target did *not* reverse direction. On the other hand, if their perception is based only on retinal motion, they would report that the target reversed direction in the SP Same and SP Opposite conditions but not in other conditions. As this example demonstrates, observers' tendency to report motion direction reversal depends on how they combine these motion vectors and the aim of this study was to determine this strategy by quantifying the weights applied to each motion vector. Table 3.1 summarizes the predicted PSSs in all conditions for each reference frame assuming no influence (i.e., w=0) from the other two.

Table 3.1. Predictions of PSSs (in deg/sec) from spatiotopic (P_s), retinotopic (P_r), and motion-based (P_{mb}) frames of references in different experimental conditions.

	P_s	P_r	P_{mb}
Baseline	***0***	***0***	***0***
Fixation	***0***	***0***	***9***
SP Same	***0***	***9***	***9***
SP Opposite	***0***	***-9***	***9***
SP Orthogonal	***0***	***0***	***9***

3.2. Predictions Based on a Single Frame of Reference

There are three reference frames available to the observers in all experimental conditions: Retinotopic, spatiotopic, and motion-based frames of reference. Minimum velocities at which subjects perceive the target to be moving backwards (the PSSs) are obtained in all conditions as a function of target-reference center-to-center distance. If perception were solely based on the spatiotopic reference frames, subjects' percepts would be independent of experimental conditions (see Fig. 3.1C) and the PSSs would be zero in all conditions and for all target-reference distances. On the other hand, had observers perceived the direction of motion solely based on retinal motion, backward motion would have been reported whenever the target disk moves backwards on the retina. Therefore, in the Baseline, Fixation, and SP Orthogonal conditions, the PSS would be zero. However, in the SP Same condition in which the reference disk is stationary on the retina (let us assume for the moment that the smooth pursuit gain is 1.0), slightest reduction in speed of the target disk causes backward retinal motion, and hence, the PSS would be equal to the average speed of the target (i.e., 9 deg/sec). Moreover, in the SP Opposite condition, according to the retinotopic reference frame, backward motion can only be perceived when the minimum speed of the target disks goes below -9 deg/sec. If, as Johansson and many researchers have suggested, the relative motion between the target and reference disks drives perceptual judgments, the PSSs obtained in all conditions (except the Baseline condition in which there is no reference disk) should be 9 deg/sec because as soon as the target disk's speed falls below 9 deg/sec, it will create a relative motion with respect to the reference disk in the opposite direction of its motion. Predictions of each reference frame in all experimental conditions are summarized in Table 3.1.

3.3. General Methods

3.3.1. Participants

Three naive observers and the author (MNA) participated in the study. The age of the participants ranged from 26 to 29 years and all participants had normal or corrected-to-normal vision. Experiments followed a protocol approved by the University of Houston Committee for the Protection of Human Subjects and research was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Each observer gave written consent before the experiments.

3.3.2. Apparatus

Visual stimuli were created via a visual stimulus generator card (VSG2/5, Cambridge Research Systems) and displayed at a resolution of 800 x 600 with a refresh rate of 100 Hz on a Sony GDM-FW900 CRT monitor. Gaze position monitoring for both eyes was performed by means of an Eyelink-II eye-tracker at 250 Hz sampling rate. The distance between the observer’s eyes and the display was 1 m and the dimensions of the display at this distance were 22.7 x 17.0 deg^2. A head/chin rest was used to help stabilize fixation and to avoid nonlinearities in eye movement recording due to head movements. Observers reported their responses via a joystick.

3.3.3. Stimuli

Spatial configurations and temporal characteristics of all conditions are given in Fig. 3.1. White (56 cd/m^2) horizontally moving disks against a black (<0.5 cd/m^2) background were utilized. Experiments were conducted in a normally illuminated room. In some studies, a dark illumination is used when the rationale of the study necessitates a complete elimination of spatiotopic references. However, as discussed in the Introduction, in our study our goal is not to eliminate spatiotopic reference frames. In fact, we are studying the prevailing reference frame in the presence of *all three reference frames*, a situation closer to everyday viewing since under ecological viewing conditions, a static background filled with trees, buildings, or other contextual elements is often present while viewing objects' movements. In order to illustrate that the edges of the display were visible, stimulus configurations shown in Fig. 3.1 were drawn with gray boundaries. Note also that since observers make motion judgments in the *horizontal* direction, the upper and lower edges of the screen do not provide a reference. The left and right edges do; however, as shown in Figure 3.1B, velocity modulations occur only in the central part of the screen where the stimulus is relatively far from the left and right edges of the screen. A fixation point was provided at the beginning of each trial and it was turned off during the motion of the disks to avoid confounding an additional relative motion factor.

We investigated the contributions of three reference frames in perceived motion-direction. Assuming that reference frames combine additively, perceived motion can be modeled as a weighted sum of motion vectors in each reference frame plus a constant term, which represents response bias (Eqn. (1)). Since we have four unknowns (three weights and a constant term), we designed four conditions (i.e., four equations) to solve this problem. In order to have a fully determined system of linear equations, the motion of the target disk was represented by a distinct set of vectors according to these reference frames in each condition. In the Baseline condition, a small target disk (0.5 deg) moved horizontally at various vertical eccentricities (2, 5, 8, and 11 deg) with the velocity profile shown in Fig. 3.1B (for rightward motion). The direction of motion was randomized across trials. Motion duration was 2520 ms. The velocity of both disks was constant during the first and last 630 ms. From 630 ms to 1890 ms, the velocity of the target disk was modulated by a sine wave. Note that as long as the amplitude of the sine wave is smaller than the magnitude of the constant speed component (9 deg/sec), the target never moves backward physically (by physically moving backwards, we mean moving backward according to a spatiotopic reference frame, for example, a reference frame centered on the computer monitor). However, if the amplitude of the sine wave exceeds the magnitude of the constant component, the minimum combined velocity of the target will fall below zero and the target will physically move backward (the green arrow in Fig. 3.1B). In the Fixation condition, the target disk was accompanied by a larger (2.0 deg) disk (referred to as the "reference disk" hereafter) with a constant velocity profile (the dotted horizontal line in Fig. 3.1B). The presence of a constant-speed reference disk along with the target in the display induces the perception of backward movement in the target, even when the target never moves backward physically (See video demo). The target and reference disks were always located at equal vertical eccentricities from the center of the display but in opposite parts of the display. Which one of the disks is presented in the upper half of the screen was also randomized across trials. In the Baseline and Fixation conditions, observers' eyes remained fixated at the remembered location of the fixation spot (i.e., the center of the screen) throughout the motion of the disks. If the left eye moved outside a 2x2 deg virtual window centered at the center during a trial in these conditions, the trial was discarded and repeated immediately. In the Smooth Pursuit (SP) Same condition, the reference and target disks had the same velocity profiles shown in Fig. 3.1B, but observers were asked to pursue the reference disk. In the SP Opposite condition, an additional disk (0.5 deg) with red color, which served as the pursuit target, was presented (Fig. 3.1A, SP Opposite). In this condition, the pursuit target was always shown at 0 deg vertical eccentricity and moved in the opposite direction (-9 deg/sec) of the target and reference disks (the red horizontal line in Fig. 3.1B). The task of the observers was to report whether the target disk moved back, i.e., the sign of its instantaneous velocity vector has ever changed, at any instant of its motion on the display.

The four conditions described so far were intended to quantify the contribution of three frames of reference: retinotopic, spatiotopic, and motion-based reference frames. After obtaining the weights of each reference frame as a function of distance, in order to test whether they can be generalized to other potential situations as well, we devised another experimental condition as a verification step. In the SP Orthogonal condition (Fig. 3.1A), the target and reference disks were presented in the same way as in the previous conditions. However, in contrast to the SP Opposite condition, the pursuit target (the red disk) moved vertically. The direction of motion (upward vs. downward) was randomized across trials. The speed of the pursuit target was again 9 deg/sec but now in the vertical dimension. The timing of the disk was arranged such that all disks crossed the symmetry axis of the display orthogonal to their motion direction at the same time.

3.3.4. Procedures

At the beginning of a trial, a fixation spot (0.2 deg) was shown at the center of the screen. Observers were required to press a button on the joystick after establishing proper fixation to carry out drift correction for better accuracy. After drift correction (or 1000±500 ms after the fixation spot was shown in the Baseline and Fixation conditions), the trial started. As soon as all disks completed their motion and disappeared, subjects pressed a button to indicate whether they perceived the target disk to be moving backward or not, in an adaptive staircase design. The amplitude of the sine wave in the target disk's velocity profile was varied by the staircase algorithm across trials. Various realizations of sine modulation are illustrated in Fig. 3.1B by gray curves. The dependent variable in this study was the "minimum velocity" of the target disk, which corresponds to the dip of the sine wave, or the global minimum in its velocity profile. For instance, a minimum speed of 9 deg/sec for the target disk corresponds to 0 deg/sec amplitude for the sine wave, i.e., no modulation, whereas a minimum speed of 0 deg/sec corresponds to 9 deg/sec amplitude for the sine wave. Different target-reference center-to-center distances (vertical separation) were blocked, and each block had four independent staircases interleaved. Each staircase had an initial minimum speed randomly chosen between -10 to 9 deg/sec (corresponding to 0 to 19 deg/sec amplitude for the sine wave) and was terminated after ten reversals in subjects' responses (a reversal is a response change from Yes (it moved back), to No (it did not move at all or it moved forward) or vice versa in two consecutive trials within the same staircase). The average of last eight reversals within a staircase was calculated and taken as the point of subjective motion direction reversal or as the point of subjective stationarity (PSS). In other words, staircases converge to a sine wave amplitude which corresponds to the minimum velocity at which the target disk is no longer perceived to be going backwards. As seen from Fig. 3.1B, if perception were veridical, (i.e., spatiotopic motion on the display were perceived), the PSS would be zero. The minimum step size in the staircase was 0.2 deg/sec. A staircase was completed in 15-40 trials depending on the subject and experimental conditions, thus a block of trials could be finished in 60 or as many as 150 trials. The order of blocks (different distances within a given condition) was randomized across subjects. The order of conditions was the same for all subjects (Baseline, Fixation, SP Same, SP Opposite, and SP Orthogonal) and each condition was run on separate days. In order to familiarize the subjects with the stimulus conditions and experimental setup, each subject ran 1 or 2 blocks of trials before collecting data for each condition. One of the practice blocks was always the Baseline trials. This was done to make sure that allocation of attention was similar among conditions. We told observers to spread their attention to the whole display during the experiments to have equivalent allocation of attentional resources among different conditions. However, we cannot completely rule out the use of different strategies in different conditions. In pilot experiments on two observers, we confirmed that the order of conditions and whether staircases for different conditions are interleaved within a block of trials or not, did not affect the results.

3.4. Results

Fig. 3.2 shows the PSSs as a function of target-reference center-to-center distance in the Fixation, SP Same, and SP Opposite conditions. Since there was no reference disk in the Baseline, the PSS values in this condition are plotted as a function of vertical eccentricity of the target disk. A two-way repeated measures ANOVA with experimental conditions and distance as the main factors showed significant effect of conditions (F(3,9)=11.471, p=0.002, η_p^2=0.793) and distance (F(3,9)=9.656, p=0.004, η_p^2=0.763). A significant effect for conditions indicates that perception is not veridical and may depend on factors such as eye movements and relative motion. The significant effect of distance confirms previous accounts and shows that the effective reference frame depends on distance. The interaction between condition and distance factors was also significant (F(9,27)=4.250, p=0.002, η_p^2=0.586) implying that distance has different effects in different conditions. Therefore, we carried out one-way repeated measures ANOVAs on each condition to assess the effect of distance in each case. We found a significant effect of distance in the Fixation condition (F(3,9)=30.571, p<0.001, η_p^2=0.911) whereas in the Baseline, SP Same, and SP Opposite conditions, the effect did not reach significance (F(3,9)=0.194, p=0.898, η_p^2=0.061; F(3,9)=3.136, p=0.080, η_p^2=0.511; F(3,9)=2.985, p=0.089, η_p^2=0.499, respectively). We also fitted linear regression lines to the results in each condition to get quantitative measures of how much distance affects the PSSs. In the Baseline condition, the percepts are veridical (i.e., the PSSs are around zero), and vertical eccentricity does not have any effect (slope = 0.010; intercept = -0.580). In the Fixation condition, for each degree of separation between the target and reference disks, the effect size drops by 0.4 deg/s (slope = -0.400; intercept = 6.170). In the SP Same condition, interestingly, the effect reaches 90% (intercept = 8.520) of the physical speed and becomes immune to changes in target-disk separation (slope = -0.020) within the range of distances used in this study. This may be due to the fact that retinotopic and motion-based frames of references reinforce the same percept (both predict an effect size of 9 deg/sec). In the SP Opposite condition, the effect of distance is reduced. Nevertheless, the distance is marginally significant (slope=-0.293, p=0.059; intercept = 6.734).

We also carried out regression fits for each subject, individually. Fig. 3.3 shows the slopes and intercepts from all subjects in all experimental conditions. Different marker shapes (and colors) represent various experimental conditions. Each point represents data

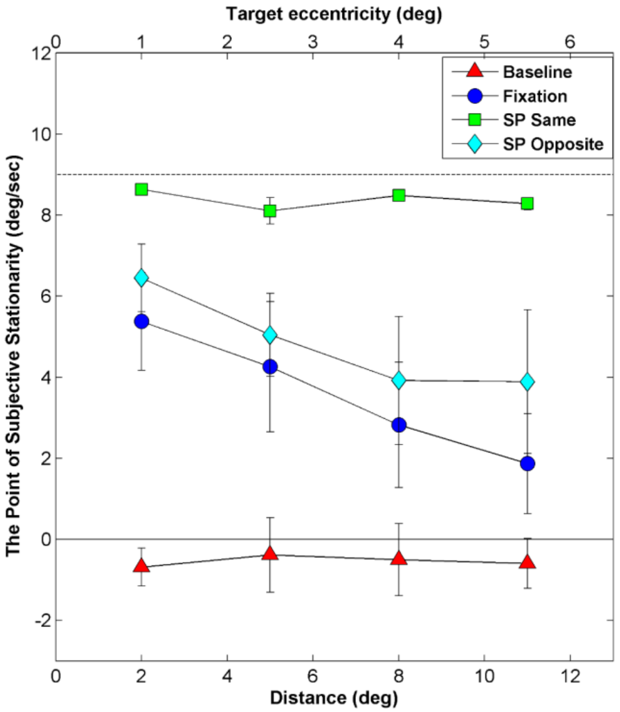

Figure 3.2. *The points of subjective stationarity in all conditions are plotted against center-to-center distance between the target and reference disks. Horizontal dotted line represents the average velocity of the target disk and the reference disk. Error bars indicate ±SEM (n=4).*

from a single subject and error bars are the standard errors of the parameter estimations. The vertical dashed line represents the average physical speed, i.e., the ceiling for the perceptual effect. Note that all the points in Fig. 3.3 fall on the left of this physical limit, i.e., intercepts smaller than 9 deg/sec. This finding itself lends support to the hypothesis that each reference frame contributes to the percept with varying weights. Furthermore, most of the points fall below the horizontal axis, which is indicative of distance-dependent contribution of at least some of the reference frames. Data from different subjects for each condition form clusters with few exceptions. Nevertheless, subject-to-subject differences do not prevent generalization of the slope and intercept values to the population of observers. Therefore, the following analyses are done on average data.

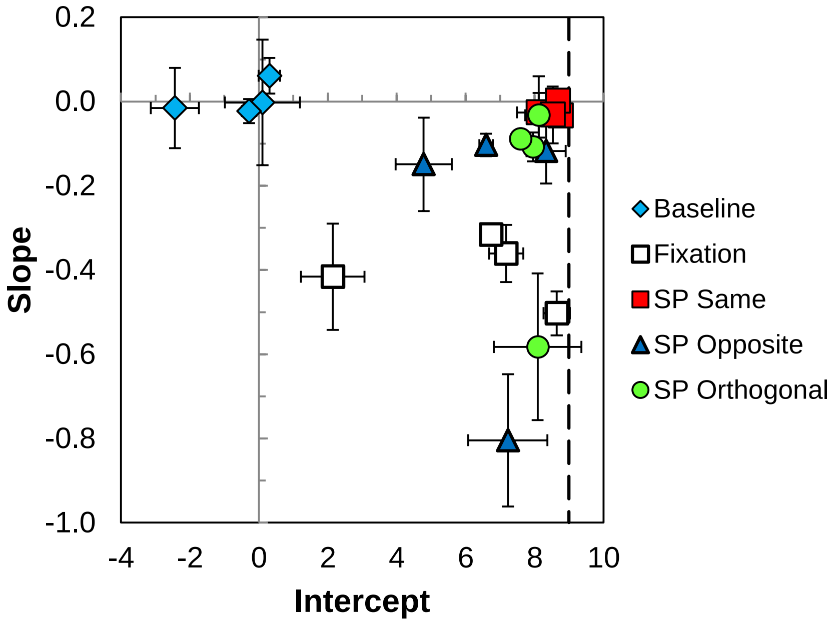

Figure 3.3. *Slope and intercept values obtained by linear regression of the PSSs obtained in different conditions. Each marker represents a single observer. Different marker types (or colors) represent different conditions. Error bars show ±SEM across staircases.*

3.4.1. Quantifying the Weight of Each Frame of Reference

We fitted Eqn. (8) to data in Fig. 3.2 and estimated the weights of each reference frame at various target-reference distances. So far, we have assumed perfect fixation stability and ideal smooth-pursuit gain (i.e., equal to one) in outlining the predictions of each reference frame. However, in practice, subjects often make micro-saccades and drifts during fixation. Stability worsens especially when there is no fixation target during stimulus presentation, as it was the case in our experiments. In addition, smooth pursuit gains are generally lower than one, which may lead to overestimation of predictions of the retinotopic reference-frame. For example, assume that a subject has a smooth pursuit gain of 0.7. This means that the retinal speed of the reference disk will not be 0 but (1-0.7)x9=2.7

deg/sec. This would decrease the PSS. In order to take these into account, we measured eye movements during the experiments.

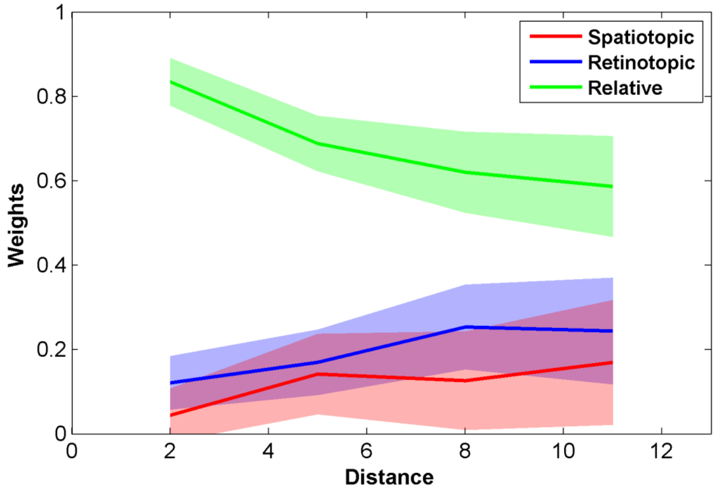

Figure 3.4. Weights of retinotopic, spatiotopic, and motion-based frames of reference in determining the perceptual judgments of motion direction. The X axis is distance, and the Y axis represents the weights. Shaded regions around the curves represent standard errors.

Fig. 3.4 shows the median horizontal eye velocities from each subject in smooth pursuit experiments. Error bars represent one interquartile range from medians. Eye movement speeds in the SP Same and Opposite conditions were not significantly different from the physical speed of the pursuit target (Note that the direction of pursuit was in the opposite direction in SP Opposite condition). Horizontal eye velocities in the SP Orthogonal condition were not significantly different than zero for all subjects. We fine-tuned the predictions of each reference frame given in Table 3.1 by using average eye velocities recorded during smooth pursuit and fixation. Fig. 3.5 shows the weights of each reference frame as a function of distance. The X and Y axes represent the distance and the weights, respectively. Different marker types (and colors) indicate different reference frames (retinotopic, spatiotopic, and motion-based). At all distances, the motion-based reference frame dominates such that the relative speed of the target disk with respect to the reference disk is perceived as backward motion even though the physical speed of the target disk was always in the same direction. There is a clear drop in the weight of the motion-based reference frame with increasing distance. Moreover, retinotopic and spatiotopic reference frames have similar weights. Their weights stay relatively constant with increasing distance.

3.4.2. Putting the Weights to the Test

We have quantified the weight of each reference frame as a function of distance, i.e., $w(d)$. In order to test whether the perceptual system uses the same weight functions in other situations as well, we designed another experimental condition, namely the SP Orthogonal condition, in which the target and the reference disks moved as in previous conditions while the pursuit target (the red disk) moved vertically. Since horizontal eye movements are assumed to be negligible during vertical pursuit, this condition leads to predictions very similar, if not identical, to those in the Fixation condition. Assuming perfect vertical pursuit without any horizontal component in the eye movements, the prediction coefficients in Eqn. (1) for the SP Orthogonal condition are $P_s = 0, P_r = 0,$ and $P_{mb} = 9$ deg/sec; in other words, identical to those in the Fixation condition (see Table 3.1). Imperfect pursuit performance will only affect P_r, which would then be equal to the average horizontal speed during vertical pursuit. The only difference between the two conditions is the existence of an ongoing vertical eye movement in the SP Orthogonal condition. Data from this experiment are shown in Fig. 3.6. A two-way repeated measures ANOVA with main factors of distance and conditions (the Fixation vs. the SP Orthogonal) showed no significant difference between the two conditions (F(1,3)=2.343, p=0.223, η_p^2=0.439). This allows us to test the weights we estimated previously in the SP Orthogonal condition. Distance had a significant effect (F(3,9)=10.949, p=0.002, η_p^2=0.785). The interaction between conditions and distance was not significant (F(3,9)=2.015, p=0.182, η_p^2=0.402). Linear regression resulted in an intercept of 7.939 and a slope of -0.203.

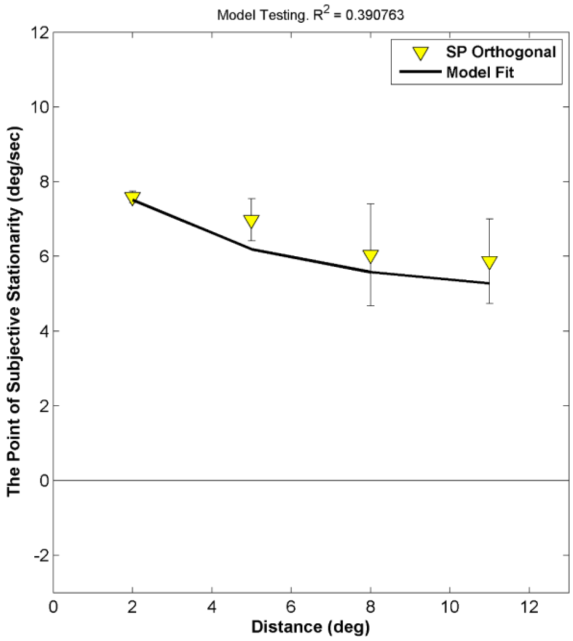

Figure 3.5. Model fit and experimental data in the SP Orthogonal condition. Error bars represent SEM (n=4).

We used the predictions and the weight functions estimated from the other four conditions to predict the results in the SP Orthogonal condition (Fig. 3.3 also shows individual slopes and intercepts in the SP Orthogonal condition). Fig. 3.6 shows experimental data (markers) and the model predictions (solid line). The model fit follows a similar pattern with the data. In fact, bootstrapping for paired samples t-test resulted in no significant difference between the model fit and the experimental data (t(3)=0.555, p=0.618, d=0.377). However, there is a slight underestimation of the effect size, which is reflected by the coefficient of determination (R^2=0.391).

3.5. Discussion

The aim of this study was to investigate the strategies used by the perceptual system to cope with the dynamic changes in the retinal stimulation. More specifically, we sought to understand the way various reference frames are utilized for motion computations. We used a variant of the induced motion paradigm, and in four conditions, we measured how direction of motion of a target disk is perceived when there are several reference frames that could be utilized in motion calculations. Specifically, we quantified the contributions of retinotopic, spatiotopic, and motion-based reference frames in a simple model. We found that the relative motion of the target disk with respect to the reference disk (i.e., motion-based reference frame) outweighs the other two, and its effectiveness is dependent on the distance between the target and the reference disks. This finding suggests that relative motion constitutes a significant part of the percepts within the range of distances tested here (up to 11 degrees). Next, we ran another condition to see whether these partial contributions are specific to the first four conditions, or they generalize to other possible situations as well. The predictions of each reference frame was the same in the Fixation and the SP Orthogonal conditions, allowing us to directly test the performance of our model. Results in these two experiments were not statistically different from each other, implying that the use of the proposed model is warranted. The proposed model successfully predicted the experimental results in the SP Orthogonal condition. There was, though, a small underestimation of the effect size, which suggests that there might be additional factors that influence perceptual judgments of motion direction. Nevertheless, the weighted combination of retinotopic, spatiotopic, and motion-based reference frames performed reasonably well in explaining the percepts.

We do not, however, overgeneralize these findings because there are still undetermined factors, which became evident with small underestimation of the effect size in the SP Orthogonal condition. One of the potential factors may be the mere existence of an ongoing eye movement. In other words, the presence of ongoing eye movements might have influenced motion detection mechanisms in a complex manner. During self-motion or eye movements, the perceptual system may rely more on relative motions. Ecologically speaking, while an organism is on the move in its habitat or moves its eyes, retinal images undergo global changes and relative motion of objects might be more relevant for survival. Therefore, it is a reasonable strategy to put more weight on relative changes during eye movements or self-motion of an observer. However, as statistical tests confirmed, we did not find any difference between the Fixation and SP Orthogonal conditions, conditions in which the only difference was the mere existence of eye movements in the latter. Contrasting these two conditions show that the sole existence of an ongoing motor action is not the missing factor. Allocation of spatial attention during fixation vs. smooth pursuit may also be different. Gogel and Sharkey (1989) measured the perceived motion trajectory of a vertically moving object in the presence of one or two objects (i.e., inducers) moving horizontally. Observers were tracking the vertically moving object. In the case of a single inducer, they found a substantially larger tilt in the trajectory of the tracked object when the inducer spot was attended than ignored. Similarly, when there were two inducer spots moving horizontally in opposite directions, attending to one of the inducers clearly increased effectiveness of the attended object. In our experiments, we asked observers to spread their attention to the whole display; however, we did not have any control over allocation of spatial attention on a trial-by-trial basis. Therefore, observers might have used various strategies to allocate their attention. This point needs further experimentation.

3.5.1. The Effective Reference Frame in Perception

In general, the choice of a reference frame cannot be done at will, and is neither a result of an all-or-none process nor an outcome of a winner-take-all competition among neural representations. Each reference frame exerts its effect with varying weights depending on its relevance and the spatio-temporal characteristics of the retinal stimuli. There are several studies supporting partial contributions of different exogenous (space-based, object-based, motion-based etc.) reference frames depending on the spatial structure and geometrical organization (Farrell-Whelan & Brooks, 2013; Magnussen, Orbach, & Loffler, 2013; Shum & Wolford, 1983; Tadin, Lappin, Blake, & Grossman, 2002), depth (Gogel & Koslow, 1972; Gogel, 1974), belongingness (DiVita & Rock, 1997), speed (Hisakata et al., 2013; Léveillé & Yazdanbakhsh, 2010; Mori, 1984), lighting conditions (Shum & Wolford, 1983), eccentricity (Thurman & Lu, 2013), and even interactions among different modalities (Avillac, Denève, Olivier, Pouget, & Duhamel, 2005). Furthermore, studies on perception during voluntary movement of eyes, head, or body indicate varying contributions of endogenous (retinocentric, headcentric, etc.) and exogenous reference frames (Agaoglu et al., 2012; Becklen et al., 1984; Brenner & van den Berg, 1994; Durgin, Gigone, & Scott, 2005; Hisakata et al., 2013; Johansson, 1976; Souman et al., 2006a; Turano & Heidenreich, 1999).

Durgin, Gigone, and Scott (2005) measured perceived speeds of visual flow under the influence of these factors, i.e., while walking on a treadmill, during physical translation (without biomechanical self-motion), and during normal walking. They found that each factor (walking without translation and translation without walking) reduces the perceived speed, and they approximately add up, i.e., the reduction is greatest during normal walking (Durgin et al., 2005). Brenner and van den Berg (1994) asked subjects to pursue a target moving against a textured background. In the midst of its trajectory, the target could increase or decrease its velocity but the background texture could also move with the target. Subjects were asked to indicate whether the target moved faster, at the same speed, or more slowly during the final interval than it had in the initial interval. They found that, as long as the target-background relative motion is kept constant, i.e., the retinal slip of the background is in the opposite direction of the eye movement, perceived target velocity does not change even if the physical speed of the target is increased or decreased. In addition, when the target speed specified by retinal signals is slower than what extra-retinal signals indicate, or is in the opposite direction, extra-retinal signals dictate perceived speed judgments (Brenner & van den Berg, 1994). Hisakata, Terao, and Murakami (2013) investigated motion-induced position shifts during smooth pursuit eye movements. In different conditions, they varied the motion of the carrier of a moving Gabor patch to pin down the critical reference frame for the illusion to occur. They found that the illusion occurs according to the envelope-relative velocity of the carrier (Hisakata et al., 2013). In other words, a motion-based reference frame drives the illusion, which again emphasizes the dominance of relative motion over retinotopic and spatiotopic motions. Anstis and Casco also demonstrated in the “flying bluebottle illusion” (see also: Furman & Gur 2005; Kano & Hayashi, 1981) that relative motion drives perception, which is reflected in shape and size judgments in their experiments (Anstis & Casco, 2006). However, the effect sizes they obtained were much larger than what would be expected by just relative motion between objects and background. This indicates that there are still other factors that might shape percepts, which is also in line with what we have found in our study. Of course, studies demonstrating the dominance of relative motion are not limited to those cited here. All these studies imply that the visual system, although it is mostly organized retinotopically in its early areas (Sereno et al., 1995; Tootell et al., 1982), chooses alternative representations over retinotopic ones, as necessitated from an ecological point of view. How does, then, the visual system choose the effective reference frame for perception? Is this process determined by some parameters of the visual stimuli or is it a result of selective allocation of attention on specific features?

3.5.2. Non-retinotopic Processes

Dynamic changes in the environment and the movement of the eyes, head, and body necessitate representations of objects and events (changes over time) that are invariant to such changes. Indeed, there have been many psychophysical (Agaoglu et al., 2012; Boi et al., 2009; Boi, Vergeer, et al., 2011; Kawabe, 2008; Léveillé, Myers, & Yazdanbakhsh, 2014; Nishida et al., 2007; Ogmen et al., 2006; Otto, Ogmen, & Herzog, 2010b; Shimozaki et al., 1999), functional imaging (Galati et al., 2000; Maus, Fischer, & Whitney, 2013; Neggers, Van der Lubbe, Ramsey, & Postma, 2006; Yin, Shimojo, Moore, & Engel, 2002; Zaehle et al., 2007), and neurophysiological studies (Bremner & Andersen, 2014; Moorman & Olson, 2007; Olson, 2003) on the existence of robust non-retinotopic processes and representations. Since early visual areas are organized retinotopically (Sereno et al., 1995; Tootell et al., 1982), these studies indirectly suggest that representations based on multiple reference frames are constructed by different levels of processing in the brain. In fact, it has been shown that a spectrum of neural representations of endogenous (e.g., retinocentric motion) and exogenous (e.g., relative motion, optic flow etc.) motions coexist in the brain (Arnoldussen, Goossens, & van den Berg, 2011; Avillac et al., 2005; Colby & Goldberg, 1999; Inaba, Shinomoto, Yamane, Takemura, & Kawano, 2007; Malkinson, McKyton, & Zohary, 2012; Neggers et al., 2006; Takemura, Ashida, Amano, Kitaoka, & Murakami, 2012). Under some conditions, the observer may be able to combine different reference frames according to task demands; however, in many cases, it is not possible for the observer to pick a reference frame at will (try for example to perceive a stationary object in motion according to a retinotopic reference frame while you are moving your eyes). A variety of factors such as adjacency, similarity, belongingness, center of gravity, figural organization, and attention, influence how reference frames are selected or combined. Further, when all parameters are kept the same, the same stimuli could be perceived differently by different observers. For instance, the classical experiment on induced motion with two dots (one moving, one stationary) showed that some observers perceive motion on one dot, some attribute motion to the stationary one, and some perceive both dots as moving at intermediate values (Day, 1978; Mack, Fisher, & Fendrich, 1975; Wallach, 1959).

3.5.3. Implications for Perceptual Vector Decomposition

Distance dependent changes in motion perception, like we have shown in this study, have been attributed to imperfect extraction of common motion vectors when there are multiple objects in a dynamic scene (Hochberg & Fallon, 1976; Shum & Wolford, 1983). These studies have shown a finite spatial distance between moving dots, in which there is a linear drop in extracted common-motion components. On the contrary, Johansson in his *theory of perceptual vector decomposition*, implicitly claimed that extraction of common motion is complete (Johansson, 1950, 1973, 1976). For instance, he explained the rotary motion of the peripheral light in Duncker's wheel stimuli by perceptual subtraction of the common motion vector from its cycloidal motion (Johansson, 1950, 1976). Johansson also demonstrated a unique ability of the human perceptual system with his biological motion displays (Johansson, 1973, 1976). Constructed by only 5-10 point-lights placed at the joints of an otherwise invisible actor, biological motion displays contain highly complex motion patterns with respect to a retinotopic or spatiotopic reference frame. Interestingly, observers can still clearly identify the type of motion even if stimulus duration is very short, e.g., 100 ms (Johansson, 1973). Bardi, Regolin and Simion (2011) provided evidence supporting the view that this ability is inborn in humans by demonstrating preference of newborns to look at biological motion displays. Johansson attributed this ability to perceptual vector decomposition. In fact, with his theory of vector analysis, biological motion displays could be described by a hierarchy of moving reference frames, thus simplifying the motions of knees and feet as simple harmonic motion of a pendulum (Johansson, 1973, 1976). Supporting his suggestions, other examples of hierarchical reference frames and potential neural processes that give rise to decomposition of these reference frames have been investigated (Bertamini & Proffitt, 2000; Grossberg, Léveillé, & Versace, 2011; Sokolov & Pavlova, 2006). Following the Gibsonian approach to the problem of perceptual stability (Gibson, 1979), Johansson also proposed that frequent changes in the proximal stimuli (due to motion of the eyes, head, or body, and motion of objects in the environment) do not pose any problems because the visual system is tuned to "abstracting" information from change in retinal stimuli over time (Jansson, Bergstrom,

& Epstein, 1994; Johansson, von Hofsten, & Jansson, 1980). This abstraction involves extracting common motion in three dimensions and using it as a reference frame to represent other changes in the retinal flow. This way, Johansson argued that motion resulting from self-movement can be distinguished from the motion of objects in the environment.

Johansson’s theory of perceptual vector analysis (Johansson, 1973) had three principles: (i) elements under motion are always perceptually related, (ii) simultaneous motion of elements form rigid perceptual groups, (iii) the decomposition of motion vectors into equal and simultaneous motion vectors leads to the perception of “common motion”, and the residual motion vectors will be perceived as “relative motion”. Although this seems logical, biological implementation of vector decomposition is not as simple as it appears. First, the extraction of common motion vectors may not always be the same for a given combination of point lights. For instance, when presented with the Duncker’s wheel stimuli, while some observers reported a rotating wheel, others reported that the motion of two point lights resembled more to a tumbling stick (Cutting & Proffitt, 1982; Proffitt, Cutting, & Stier, 1979). Different speeds along the same motion trajectories have been shown to result in both subject-to-subject and trial-to-trial differences in perceived common motion in two point light displays (Mori, 1984). More importantly, in mathematical terms, vector decomposition is an ill-posed problem: Infinitely many pairs of common and relative motions can produce exactly the same absolute motion, i.e., motion with respect to a stationary exogenous reference frame. Therefore, a fundamental question in vector decomposition has been to determine which of the infinitely many solutions is adopted by the visual system.

Several heuristics have been proposed as to how vector decomposition takes place in the brain. Here, we describe only one of them in detail. In Wade and Swanston's (1987) formulation, *"... finding the common motion is equivalent to a comparison of pairs of relational* [i.e., relative] *motions, and selecting those which have the same value".* And *"it seems reasonable to adopt those points that have the same relational motion values as a patterncentric frame of reference for allocating other relational motions"* (Wade & Swanston, 1987, p. 564). According to the proposed method, the perceptual system has to do pairwise comparisons for all the points (objects or features). If this process were to be carried out in an iterative way, it would run into combinatorial explosion in a natural environment. A parallel implementation of this process would require a complex neural architecture that needs to be spelled out.

3.5.4. An Alternative to Vector Decomposition: The Reference-Frame Metric Field Theory

Recently, we proposed the Reference-Frame Metric Field (RFMF) theory (Ogmen, Herzog, & Noory, 2013b) in order to study how reference frames are established and how dynamic computations of form is carried out. Here we describe only the part of RFMF that is relevant to this study. Fig. 3.7 provides a schematic description of the RFMF theory. At the bottom, the retinotopic space is illustrated. Here, several simple stimuli (dots) are in motion. These motion vectors are grouped locally to generate local motion vectors. These local motion vectors serve as *local reference frames***.** According to our theory, *a field is created around each local reference frame* (like an electromagnetic field) and fields of different reference frames *interact* to establish a global equilibrium in the retinotopic space. Therefore, the visual system does not need to solve explicitly the ill-posed common motion extraction problem. Instead, motion-based reference frames result from a process of *reference field* interactions. Interactions of various fields give rise to perceptual organizations such that different frames of reference may dominate different regions in the perceptual space. The degree to which extraction process takes place may be incomplete (i.e., the perceived and physical common motion vectors may not be equal) and can vary with spatio-temporal properties of the stimuli (DiVita & Rock, 1997; Gogel & Koslow, 1972; Johansson et al., 1980; Mori, 1979, 1984; Poljac, Verfaillie, & Wagemans, 2011; Shum & Wolford, 1983). This finding rejects perfect vector decomposition and rather calls for a finite region within which the extraction of common motion component can be carried out. Yet within this finite region many moving stimuli can exist, each providing a potential reference frame. This region might depend on stimulus scale (Maruya, Holcombe, & Nishida, 2013) and other parameters which need further investigation. In terms of neural implementation, field interactions can be realized by distance-dependent isotropic (in case of no direction bias) or anisotropic (in case of direction bias) connections among units carrying out the computations.

3.6. Conclusion

Taken together, our results suggest that the effective reference frame for motion perception involves a combination of motion-based, retinotopic and spatiotopic reference frames. Relative motions defined by motion-based reference frames dominate retinotopic and spatiotopic motions. Effectiveness of the motion-based reference frames drops substantially as the distance between objects of interest increases, indicating a finite region within which each motion-based frame of reference operates. Such dramatic changes are not found for retinotopic and spatiotopic reference frames. Contributions of retinotopic and spatiotopic frames of reference are minimal when there is an ongoing smooth pursuit eye movement. From the perspective of the RFMF theory, motion-based reference frames emerge from field-like interactions of local motion vectors, thereby providing an alternative account to the vector decomposition approach. Distance-dependent changes in perceived motion result from interactions between multiple motion-based reference frames. These interactions determine an effective reference frame whereby information from retinotopic representations can be mapped into non-retinotopic ones. In our previous studies, we have suggested that such a mapping allows dynamic computation of form while avoiding motion blur, moving ghosts, and occlusion problems (Ogmen & Herzog, 2010; Ogmen, 2007).

Chapter 4. Field-Like Interactions between Motion-Based Reference Frames

4.1. Introduction

4.1.1. Relativity of Perceived Motion

The perception of a stimulus does not only depend on its own individual properties but also on the properties of other spatiotemporally neighboring stimuli (Koffka, 1935). Karl Duncker was one of the first scientists who examined this issue for the perception of motion (Duncker, 1929; Ellis, 1938). In one of his experiments, he used displays generated by point-lights attached to an otherwise invisible rotating and translating circular cardboard (Duncker, 1929, pg. 240). He found that when a single point-light was attached to the rim of a cardboard, observers reported seeing the point-light moving along a cycloidal trajectory. On the other hand, when he added a second point-light to the hub of the cardboard, observers often reported perceiving the point-light attached to the rim undergoing circular motion around the point-light on the hub, which itself was perceived to move horizontally. These results can be understood in terms of the reference-frame against which the motion of the point-light is perceived. Cycloidal motion corresponds to a trajectory relative to a stationary geocentric reference while the rotation corresponds to a trajectory *relative* to a moving reference-frame positioned on the point-light at the hub.

4.1.2. The Theory of Perceptual Vector Decomposition

In order to demonstrate the power of the visual system in determining behaviorally relevant reference frames, Johansson devised the “biological motion” paradigm in which point-lights are attached to few joints of humans undergoing complex motion (Johansson, 1973). In viewing these displays, observers readily perceived the underlying biological motion. In other words, observers were able to select the appropriate reference frame(s) that revealed the underlying biological motion. In order to explain the selection of the reference frame, Johansson proposed a theory of vector decomposition (Johansson, 1976). This theory is based on three principles: (i) elements under motion are always perceptually related, (ii) simultaneous motion of elements form rigid perceptual groups, (iii) decomposition of motion vectors into equal and simultaneous motion vectors leads to perception of “common motion”, and the residual motion vectors will be perceived as “relative motion”. Fig. 4.1 illustrates these concepts. Assume that three vertically positioned dots shown in Fig. 4.1A have the corresponding motion vectors associated with them individually. The dots at the top and bottom move back and forth along the horizontal dimension whereas the dot in the middle moves along an oblique trajectory. What observers perceive is shown in Fig. 4.1B: all three dots moving in tandem to the right and to the left (blue arrows) as a group and the one in the middle moves up and down (black arrows). By the first principle, individual trajectories of dots are not perceived in isolation but related to each other. By the second principle, the top and bottom dots form a perceptual group due to their simultaneous motions. Finally, by the third and the most important principle, the motion of the middle dot is decomposed into a horizontal component, which is common to other dots, and a vertical component, which makes it look like moving up and down. Fig. 4.1C illustrates common and relative motion components of each dot resulting in the percept shown in Fig. 4.1B (review: Herzog & Öğmen, 2014; Shum & Wolford, 1983).

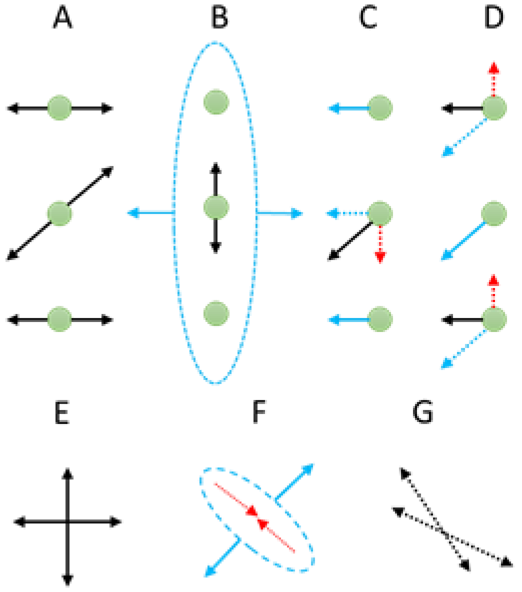

Figure 4.1. A. The three element display used by Johansson. B. Typical percepts C. The vector decomposition in B. D. An alternative decomposition. E. The motion trajectories in the two-dot display used by Johansson (1950). F. Vector decomposition for E. G. Reported percepts.

4.1.3. Vector Decomposition: An Ill-posed Problem

Although the processes illustrated in Fig. 4.1 seem straightforward, in mathematical terms, vector decomposition is an ill-posed problem: Infinitely many pairs of common and relative motions can produce exactly the same absolute motion. Fig. 4.1D shows an alternative set of common and relative motion components corresponding to the same physical motion in Fig. 4.1A. Johansson recognized this ambiguity in some of his studies (Johansson & Jansson, 1968; Johansson, 1950, 1958). For example, in his Exp. 19 (1950, p. 89), in which he presented two-dot displays where one of the dots oscillated horizontally while the other oscillated vertically (see Fig. 4.1E), he reported that subjects did not always experience the same motion configuration. If they attended to one of the dots, they perceived the geocentric motion of that dot while perceiving the other dot as moving along a slanted trajectory. There were even reports of 3D rigid motion of a rotating rod. Similarly, with Duncker's wheel stimulus, while some observers reported a rotating wheel, others reported that motion of two point lights resembled more to a tumbling stick (Cutting & Proffitt, 1982; Proffitt et al., 1979).

Since observers do not perceive all possible solutions, but instead a rather small subset (e.g., Johansson, 1950; Proffitt, Cutting, & Stier, 1979; Wallach, Becklen, & Nitzberg, 1985), the fundamental questions are to determine which subset of solutions is selected by the visual system and why. Mathematically, the number of solutions of an ill-posed problem can be reduced by introducing additional information or constraints, an approach known as *regularization* (Marr & Ullman, 1981).

4.1.4. Which Subset and Why: Regularization Approach in Vector Decomposition

A variety of constraints were proposed to explain how the visual system regularizes vector decomposition. Hochberg and McAlister (1953) argued that the perceptual system chooses the simplest solution in terms of information required to define the pattern when it encounters an ambiguous stimulus with multiple potential interpretations. Börjesson and von Hofsten (1972) proposed as constraint that residual motion vectors sum to zero. Gogel proposed the "adjacency principle" according to which the relative motion determination is restricted only to nearby objects (Gogel & Koslow, 1972; Gogel, 1974). Proffitt and colleagues proposed that the common motion is determined by the motion of the center of gravity of the dots (Proffitt et al., 1979). Restle (1979) proposed “information load” as the constraint to be minimized in determining the prevailing solution. A hybrid system which minimizes either the common or relative component, depending on which process (assuming that the common and relative motion calculations are done via independent processes) is completed first, is shown to account for some of the classical findings in dot motion experiments (Cutting & Proffitt, 1982). More recently, building upon Johansson’s original study of vector analysis, a Bayesian framework with a set of probabilistic constraints is also introduced (Gershman, Jäkel, & Tenenbaum, 2013). In sum, the constraints introduced in regularization approaches to vector decomposition provide heuristics to explain, at least partially, *why* the human visual system selects a particular vector decomposition in motion perception.

To put this approach in perspective, consider its use in physics. In physics, the principle of minimum total potential energy (which states that particles move so as to minimize the total potential energy) is formulated to explain *why* things move the way they do. Hence, based on this constraint, a global energy function can be minimized to determine the motion of particles in a medium. An alternative perspective is to express *how* this particular solution emerges in real-time through interactions between particles. In this case, one uses forces and fields applied to particles. In mathematical terms, these two approaches are related and can be expressed as the energy (or Ljapunov) function of a system and the differential equations governing the system.

4.1.5. Reference-Frame Metric Field (RFMF) Theory

To complement regularization approaches, we have taken the alternative perspective with the goal of expressing reference frame selection in terms of interactions between motion vectors. Our goal in formulating RFMF was not just to explain how motion is perceived but, more generally, how the visual system computes stimulus attributes in the presence of relative motion between the observer and the environment.

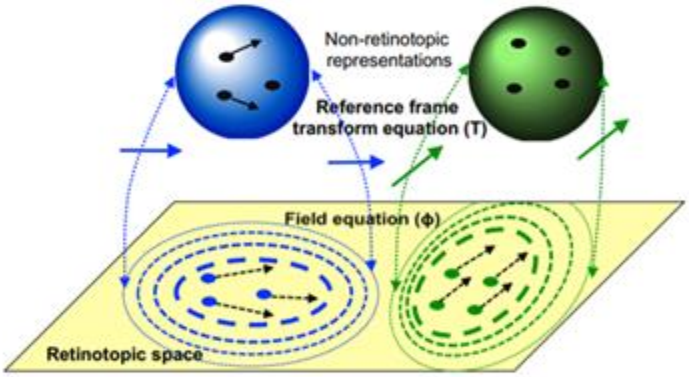

Figure 4.2. Schematic illustration of the RFMF theory. For more details, see (Ogmen & Herzog, 2010; Ogmen, 2007).

The early visual system is retinotopically organized. Furthermore, a briefly presented stimulus remains visible for about 100ms after its offset, a phenomenon known as visible persistence (Coltheart, 1980; Haber & Standing, 1970). Based on retinotopic representations and visible persistence, one would expect moving objects to appear extensively blurred without any form information (Ogmen, 2007; Ogmen & Herzog 2010). However, under normal viewing conditions, moving objects appear relatively sharp and clear (Hammett, 1997; Ramachandran et al., 1974). To explain how dynamic form is computed while avoiding these problems, we proposed RFMF (Ogmen, 2007; Ogmen & Herzog, 2010; Ogmen et al., 2013). The main idea is to compute dynamic form not according to a retinotopic reference-frame, but instead according to a reference-frame that moves according to the motion of the stimuli. Fig. 4.2 provides a schematic description. The yellow plane at the bottom depicts retinotopic representations. Two groups of dots, shown in blue and green, move in different directions in retinotopic coordinates. A common motion is determined for each group and this common motion is used to map retinotopic representations into distinct non-retinotopic representations, depicted as spheres at the top of the figure. According to RFMF theory, each retinotopic motion vector creates a field in the retinotopic space (like an electromagnetic field). Fields created by different motion vectors interact in order to determine a motion vector that will serve as the reference-frame at a given point and time in space. Previously, we have demonstrated the field-like nature, i.e., distance-dependent influence, of motion-based reference frames on the perception of motion direction (Noory, Herzog, & Ogmen, 2013; Ogmen, Herzog, & Noory, 2013; Agaoglu, Herzog & Ogmen, submitted). This concept of distance-dependent reference-frame determination is also consistent with earlier observations on relative motion. For example, Mori (1979) found a linear decrease in the perceived horizontal component of the middle dot in Fig. 4.1A with increased separation between the dots.

In the present study, we investigated whether and how different reference-frame fields interact. We sought to determine (i) whether reference-frame fields are actually generated by local motion vectors (i.e., to replicate previous findings with a variant of the stimuli previously used), (ii) whether stationary landmarks generate reference fields as well, (iii) whether fields created by different motion-based reference frames actually interact and, if so, (iv) how they interact. To this end, we probed perceived direction of a moving dot with and without drifting gratings (which produce motion-based reference frames) at various distances from the dot and we quantified the field effect and relativity of perceived motion.

4.2. General Methods

4.2.1. Participants

Three naive observers and the author (MNA) participated in the study. The age of the participants ranged from 26 to 29 years and all participants had normal or corrected-to-normal vision. Experiments followed a protocol approved by the University of Houston Committee for the Protection of Human Subjects. Each observer gave written consent before the experiments.

4.2.2. Apparatus

Visual stimuli were created via visual stimulus generator card VSG2/5 (Cambridge Research Systems) and displayed at a resolution of 800 x 600 with a refresh rate of 100 Hz. Gaze position monitoring for both eyes was performed by means of an Eyelink-II eye-tracker at 250 Hz sampling rate. The distance between observer's eyes and the display was 1 m and the dimensions of the display at this distance were 22.7 x 17.0 deg. A head/chin rest was used to help stabilize fixation. Observers reported their responses via a joystick. All experiments were done in a normally illuminated room.

4.3. Experiment 1: The Reference Field Effect

The purpose of the experiment was to demonstrate the distance dependent effects of a motion-based reference frame, i.e., the reference field effect. In Experiment 4.1a, we established a baseline for the judgments of motion direction in the absence of any dynamic reference frames. In Experiment 4.1b, we added a motion-based reference frame to the display and varied the distance of the target from this reference.

4.3.1. Stimuli and Procedures

Previously, we have used a variant of the induced motion paradigm to see how various reference frames influence perception of motion (Agaoglu, Herzog, & Ogmen, in press, see Chapter 3). Similar to the stimuli used in that study, we used a small white square (56 cd/m^2) as the target on a black background (<0.5 cd/m^2) (Fig. 4.3A). Each trial started off with a presentation of a fixation cross at the center of the display for a randomly chosen duration (400-700 ms). As soon as the target appeared, the fixation cross was turned off to avoid additional reference frames, which might be used in motion judgments. Observers were required to keep fixation at the remembered location of the fixation cross during the trial. Trials during which gaze positions of observers deviated more than 2 deg from the fixation cross were discarded and repeated immediately. The target square moved horizontally at a fixed vertical eccentricity in a given trial. There were five possible vertical eccentricities for the target square in all conditions: 0 deg, ±2.75 deg, and ±5.5 deg (positive values represent upper visual field). Direction of the target motion was randomized across trials. An example velocity profile of the target, where it moved from left to right, is given in Fig. 4.3F (thin red curves). The velocity of the target was constant during the first and last 630 ms. From 630 ms to 1890 ms, the velocity of the target square was modulated by a sine wave. As long as the amplitude of the sine wave is smaller than the average speed of the target (9 deg/sec), the target might either decelerate or accelerate but never moves backward in spatiotopic (e.g., screen based) coordinates (light thin red curves in Fig. 4.3F). However, if the amplitude of the sine wave exceeds the average speed, the target stops and reverses its direction for a short amount of time (dark thin red curves). Fig. 4.3G shows the space-time diagram of the target's motion for various amplitudes of the sine modulation. Note that the x-axis represents roughly the second quarter of the target motion on the display. When the amplitude of the sine wave is equal to the average speed of the target, i.e., 9 deg/sec, then the target slows down and stops completely and starts accelerating until it reaches 9 deg/sec speed again. When the amplitude of the sine modulation exceeds 9 deg/sec, dips in the space-time curves, indicating *spatiotopic* backward motion, become apparent (modulation amplitudes 12 and 15 in Fig. 4.3G).

In Experiment 1a (Fig. 4.3A), the target square was presented alone at various vertical eccentricities across blocks. In Experiment 1b, a square wave grating (dimensions: 23 deg x 1 deg, spatial frequency: 0.25 cpd, duty cycle: 50%, Michelson contrast: 0.98) was always presented at 6.5 deg vertical eccentricity either in the upper or lower part of the display. The drift speed of the grating was equal to the average speed of the target square (9 deg/sec) and the drift was always in the same direction as the target square's motion (see the thick red arrow in Fig. 4.3B and the thick red line in Fig. 4.3F). Vertical eccentricities of the target were blocked and the order of blocks was randomized.

Observers were asked to spread their attention to the entire display and, as soon as the target (and the grating) completed their motion and disappeared, to report via a joystick whether the target has moved backward ever during the trial (Yes/No). The amplitude of the sine modulation in the target's velocity profile was varied across trials by an adaptive staircase algorithm (see various thin red lines showing different modulation amplitudes in Fig. 4.3F). For each reversal in observers' responses, the step size in the staircase was halved. Four independent staircases with randomly chosen initial amplitudes (within the range 0-19 deg/sec) were interleaved in a block of trials. A single staircase was completed in 15-25 trials. A staircase was considered "converged" when it encountered ten reversals in observer's responses and the last eight reversals were used to calculate the threshold for perceiving backward motion. The minimum velocity of the target corresponding to this threshold amplitude was taken as the reference field effect. For instance, if the staircase converges to 9 deg/sec amplitude for sine modulation, it corresponds to the minimum target velocity of 0 deg/sec (the reference field effect = 0 deg/sec). This would mean that backward motion is perceived only when the target velocity goes below 0 deg/sec (veridical percept). On the other hand, if, for instance, the staircase converges to 6 deg/sec corresponding to the minimum target velocity of 3 deg/sec, it would mean that as soon as the target velocity falls below 3 deg/sec (the reference field effect = 3 deg/sec), backward motion is perceived (illusory percept) although it may never spatiotopically move backward. For each vertical position of the target, each observer ran one block of trials (four staircases).

Experiment 1a was designed to determine quantitatively for each observer the ability to detect a reversal in the direction of motion of a simple stimulus as a function of eccentricity. Ideally, the thresholds for detecting backward motion should be zero or close to zero. However, since we used a yes/no task, bias may occur. In Experiment 1b, the reference field generated around the drifting grating should cause illusory percepts. To be more specific, the RFMF theory predicts that backward motion should be perceived even at positive minimum target velocities (field effects > 0 deg/sec). In order to show that percepts are not simply due to subtraction of a common motion component and that a motion-based reference frame is effective only within a limited spatial region, we analyzed the field effects as a function of grating-to-target distance.

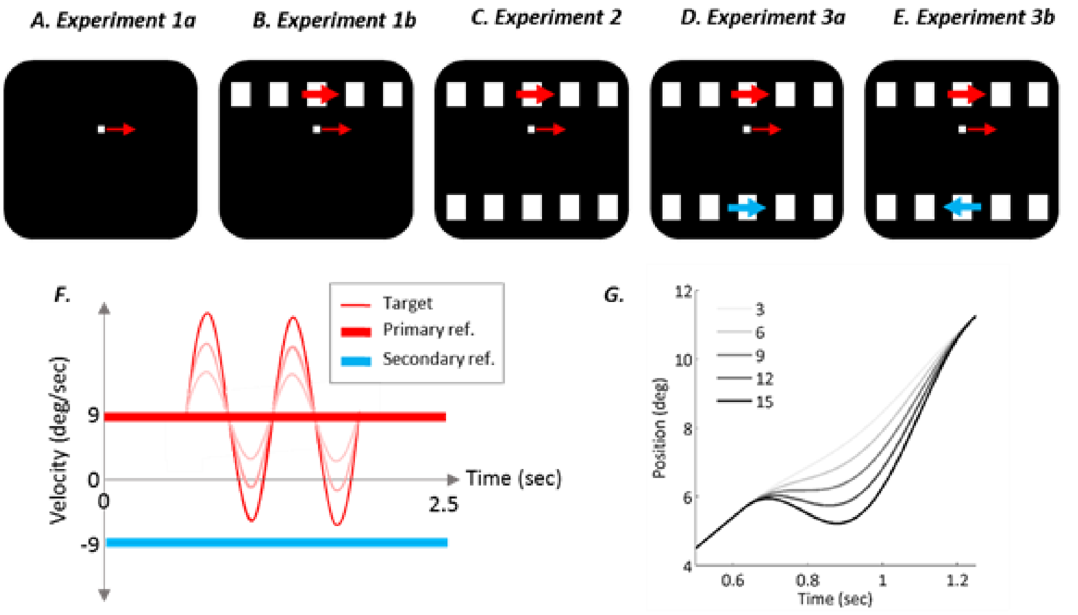

Figure 4.3. Spatial and temporal characteristics of the stimuli. A-E. Experimental conditions F. Velocity profiles of the target dot and drifting gratings. G. Close-up view of the space-time graphs of the target dot for various magnitudes of velocity modulation.

4.3.2. Results and Discussion

Fig. 4.4A shows the baseline thresholds as a function of vertical eccentricity in Experiment 1a. There is no indication of perceptual biases at a particular eccentricity. A one-way repeated measures ANOVA confirms that the vertical eccentricity has no significant effect ($F_{2,6}$=0.338, p=0.726, η_p^2=0.101) on perceived direction of the target's motion. However, we found a small but a significant bias towards negative reports, i.e., to say “No, it did not move back”, (mean across observers and vertical eccentricities = -0.620 deg/sec, one sample t-test: t_{11}=-4.018, p=0.002).

The presence of the drifting grating significantly changed the pattern of results. Fig. 4.4B shows the reference field effects (defined as the minimum speed at which the target is perceived to be moving backward, see Stimuli and Procedures) as a function of target-grating distance. Although we found a small but significant perceptual bias in Experiment 1a in the absence of the drifting grating, we did not take this bias into account for the results of Experiment 1b since this bias was not eccentricity dependent. Factoring out these biases would only cause an overall upward shift, not affecting statistical analysis. The drop in the effect size with increasing target-grating distance, shown in Fig. 4.4B, is significant ($F_{4,12}$=13.550, p<0.001, η_p^2=0.819). The results replicate the distance dependent influence of motion-based reference frames on perceived motion of nearby objects that has already

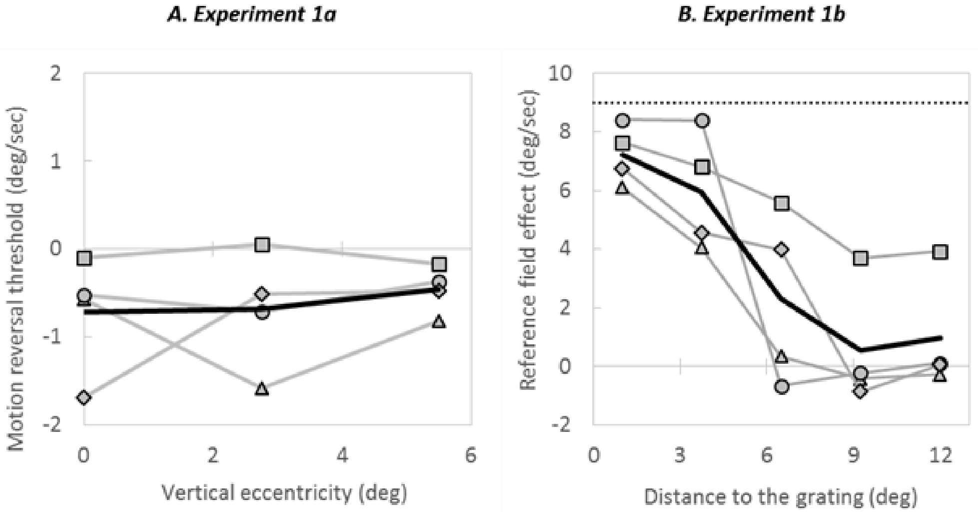

Figure 4.4. A. In Experiment 1a, minimum velocity of the target dot at which it appeared to move backward is taken as motion reversal thresholds and plotted against vertical eccentricity. B. Motion reversal thresholds (referred to as the reference field effects).

been shown by several studies (DiVita & Rock, 1997; Hochberg & Fallon, 1976; Mori, 1979; Shum & Wolford, 1983). All these studies lend support to the claim that common motion might serve as a reference frame but its effectiveness is limited to a spatial region.

4.4. Experiment 2: No Motion, No Interaction

The RFMF theory predicts that reference-field interactions occur only when there is motion. In this experiment, we tested this prediction by presenting a static grating in addition to the dynamic grating. The static grating was in the other half of the visual field than the dynamic grating. It has been shown that the presence of a stationary landmark (such as fixation point, a surrounding frame etc.) substantially influences the perceived motion (e.g., Wallach et al., 1985). The static grating used in Experiment 2 provides an additional reference frame (along with the display borders) for motion computations; however, whether it can generate interactions with a reference field needs to be investigated. If the presence of the static grating modulates the strength of a motion-based reference-field, we should see distance dependent drops in the reference field effect compared to the case where only the dynamic grating is presented.

4.4.1. Stimuli and Procedures

Stimuli and procedures were identical to those used in Experiment 1b with the following exceptions. In addition to the dynamic grating which always moved in the same direction as the target, a second grating was presented at the same vertical eccentricity as the dynamic one but in the opposite half of the screen (Fig. 4.3C). The second grating was stationary (drift velocity = 0 deg/sec). Which one of the gratings was presented in the upper visual field was randomized across trials.

4.4.2. Results and Discussion

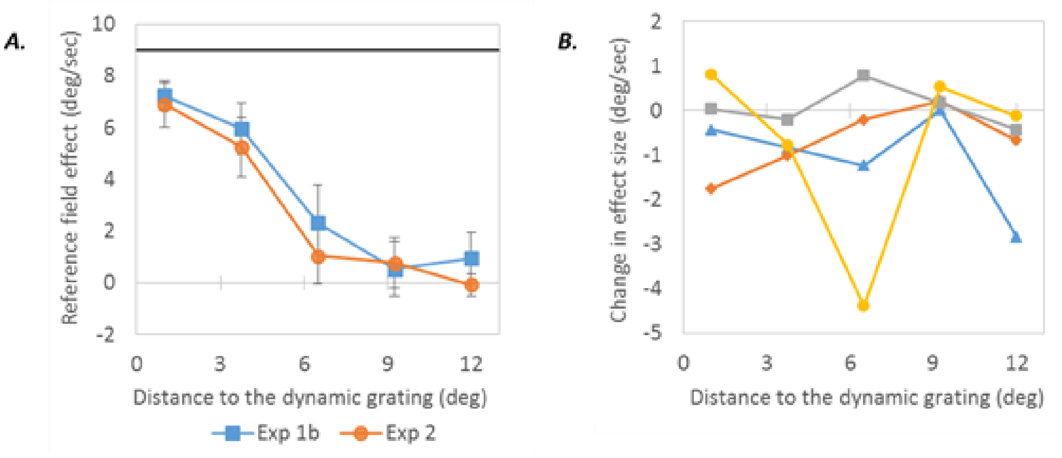

Figure 4.5. A. Reference field effects average across observers (n=4) in Experiment 1b and Experiment 2. Error bars represent ±SEM. B. Change in effect size with the addition of a static grating. Each line represents a different observer.

Fig. 4.5A shows the reference field effects measured in the presence (Experiment 2) and the absence (replotted from Experiment 1b) of the static grating. Similar to Experiment 1b, a one-way repeated measures ANOVA on the reference field effects in Experiment 2 revealed a significant effect of the distance between the target and the dynamic grating ($F_{4,12}=18.355$, $p<0.001$, $\eta_p^2=0.860$). A two-way ANOVA with target to dynamic-grating distance and experiments as the main factors showed, once again, a significant effect of distance ($F_{4,12}=18.792$, $p<0.001$, $\eta_p^2=0.834$). However, the main effect of experiments did not reach significance ($F_{1,3}=6.346$, $p=0.086$, $\eta_p^2=0.679$). More importantly, the interaction of the main factors was not significant either ($F_{4,12}=0.824$, $p=0.535$, $\eta_p^2=0.215$). These results suggest that static reference-frames might influence motion perception (though its effect did not reach significance here) but they do not generate field-like spatial zones within which their influence is modulated by distance. The failure to obtain significant main effect of the experiments (i.e., the addition of a static grating) might be due to a floor effect: Experiments were done in a normally illuminated room where screen edges and other potential spatial landmarks were visible. Adding another spatial landmark might not be very effective.

One might also argue that failure to find significant interaction between distance and experiments here might be due to the floor effect. At around 9-12 degrees of distance from the dynamic grating, the effect size is already very close to zero and any potential drop in the effect size due to the presence of the static grating might be clouded. In order to address this issue, we took the three closest distances to the dynamic grating and repeated the statistical analysis. The main effect of distance was again significant ($F_{2,6}=10.240$, $p=0.012$, $\eta_p^2=0.773$) whereas the main effect of experiments was not ($F_{1,3}=4.775$, $p=0.117$, $\eta_p^2=0.614$). The interaction of the main factors was, once again, not significant ($F_{2,6}=0.358$, $p=0.713$, $\eta_p^2=0.106$). Moreover, within observer differences between the two experiments do not show any trends whatsoever (see Fig. 4.5B). If there was indeed field associated with static references, we would see positive slopes, i.e., increased effect size difference with distance.

4.5. Experiment 3: Interacting Dynamic Reference Fields

In the first experiment, we replicated the basic finding that perceived motion of a stimulus can be influenced by motions of nearby objects and that this effect spreads over space in a field-like manner. In the second experiment, we showed that the reference fields are generated only when motion is present. In ecological viewing conditions, a multitude of objects might move in various directions. According to the RFMF theory, in order to perceive sharp and clear forms of these objects, each object needs to be processed in a proper reference frame (determined by local motion vectors). Further, selection of a certain reference frame is not done in all-or-none manner. Instead, each reference frame has a reference field associated with it and exerts its effect within this spatiotemporally limited field. A question, then, arises: What if several reference fields come into close proximity with each other? The RFMF theory suggests that reference fields interact to reach an equilibrium in the retinotopic space. In this experiment, we tested this prediction.

4.5.1. Stimuli and Procedures

Stimuli and procedures were identical to those used in Experiment 2 with the following exceptions. The static grating was replaced by a drifting grating having the same spatial characteristics as the dynamic grating in the Experiments 1b and 2. In Experiment 3a, the drift velocities of the two gratings were identical and equal to the average velocity of the target (Fig. 4.3D). Since these gratings are identical in all aspects, target-grating distances of 1 deg and 12 deg, and 3.75 deg and 9.25 deg are essentially the same. Therefore, we had effectively only three target-grating distances in Experiment 3a (1, 3.75, and 6.5 deg). In Experiment 3b, one of the gratings (primary) always drifted in the target's motion direction while the other (secondary) drifted in the opposite direction but with the same speed (Fig. 4.3E). As in Experiment 2, the target was presented at five different vertical eccentricities (corresponding target-primary distances were 1, 3.75, 6.5, 9.25, and 12 deg). Which one of the gratings was presented in the upper visual field was randomized across trials.

4.5.2. Results and Discussion

Reference field effects averaged across observers in Experiments 3a and 3b are plotted in Fig. 4.6. The primary x-axis represents the distance to the primary grating (which moves in the same direction as the target's average velocity) and the secondary x-axis represents the distance to the secondary grating. In the case of Experiment 3a, the distinction between the two gratings is void since they both drift in the same direction, and in order to facilitate visual comparison of results in the Experiments 3a and 3b, data at distances 1 and 3.75 deg are replotted at distances 9.25 and 12, respectively (Fig. 4.6 dashed lines and open symbols). However, all of the following statistical analyses are carried on the reference field effects at the three closest distances (i.e., 1, 3.75, and 6.5 deg). A one-way repeated measures ANOVA showed that the effect of distance is only marginally significant ($F_{2,6}=4.774$, $p=0.057$, $\eta_p^2=0.614$). When the second grating drifted in the opposite direction, the effect of distance became highly significant ($F_{4,12}=27.265$, $p<0.001$, $\eta_p^2=0.901$). Moreover, a two-way repeated measures ANOVA with distance to the primary grating and drift direction of the secondary grating as the main factors yielded significant main effects for both factors (distance: $F_{2,6}=37.715$, $p<0.001$, $\eta_p^2=0.921$; drift direction $F_{1,3}=24.517$, $p=0.016$, $\eta_p^2=0.891$). Interaction between factors was also significant ($F_{2,6}=7.450$, $p=0.024$, $\eta_p^2=0.713$). The significant main effect of drift direction of the second grating suggests that reference fields generated by two gratings interact, and the nature of this interaction (being constructive or destructive) depends on the direction of motion within the second grating. Significant interaction of the main factors, as manifested by distance dependent differences between the reference field effects in Experiments 3a and 3b, also lends support to the reference field hypothesis.

Experiments 3a and 3b allow us to draw conclusions on the interactions taking place. Fig. 4.7A illustrates the reference fields generated by the drifting gratings. Assuming that different fields linearly add up, when both gratings drift in the same direction, their respective reference fields should lead to accumulation as illustrated in Fig. 4.7B. However, when one of the gratings drifts in the opposite direction, the respective fields should not add up linearly. The "sign" of the reference field of the second grating is negative and corresponding fields are illustrated in Fig. 4.7C. If two fields were to simply add up, illusory perception of backward motion should have decreased with distance and, perceptual responses to say "No, the target did not move backward" (although it physically does) should have taken over and became stronger as the target became closer to the second grating (Fig. 4.7C(i)). The resultant reference field effects, hence, should start at a positive value (indicating illusory percepts) and go down to negative values (indicating perceptual biases to say No) roughly linearly. However, our results do not confirm these predictions.

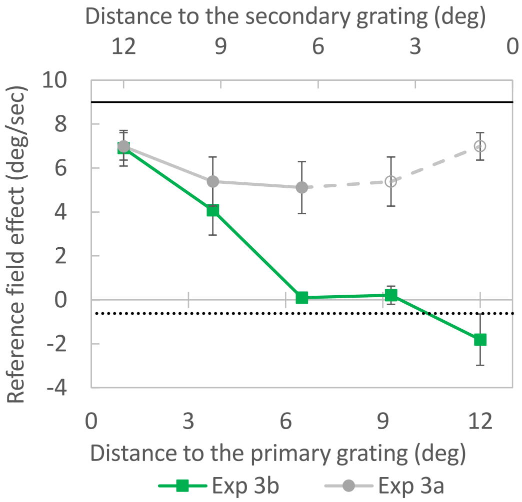

Figure 4.6. Results from Experiments 3a and 3b. Circles represent results from Experiment 3a whereas square symbols represent data from Experiment 3b. Error bars represent ±SEM (n=4).

A faster drop in the field effect was followed by veridical percepts (i.e., zero effect size) as the distance between the target and the second grating became smaller. This effect is illustrated in Fig. 4.7C (ii). Therefore, the interactions of different reference fields cannot be explained by simple linear summation of individual fields. Rather, the asymmetry underlying the nonlinear effect needs to be understood in terms of the coherence between the motion of the target and those of the reference frames. As illustrated in Fig. 4.7, the average motion of the target is in the same direction as the grating shown in yellow but in opposite direction to the grating shown in blue. As depicted in Fig. 4.2, according to RFMF theory, motion-based grouping takes place first. Hence the target is grouped with the grating drifting in the same direction but not with the grating drifting in the opposite direction. Hence the results can be incorporated to the theory as follows: A motion signal that is not grouped with the target can decrease the strength of the field but it cannot revert it so as to cause a sign change in the effect. In other words, it can null the effect of a reference-frame but it cannot become the reference frame itself since it is not grouped with the target.

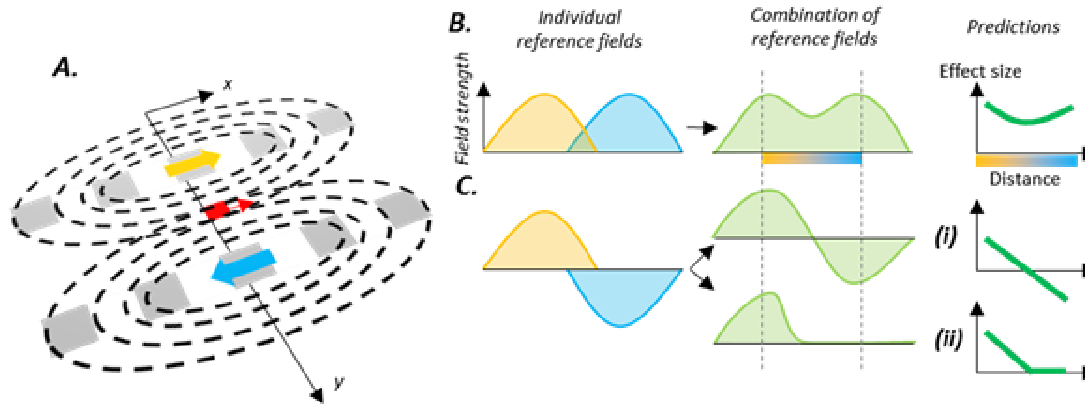

Figure 4.7. *An illustration of the reference fields and their interactions. A. Motion within each grating generates a reference field, which spreads over space and time. B. When both gratings drift in the same direction. C. When the gratings move in opposite directions.*

4.6. General Discussion

When we compute the motion of an object in everyday life, we generally use the static environment as a reference frame. However, perceived motion of an object corresponds to its motion with respect to a static reference frame only in special, simple cases. When we see a friend waving his/her hand on a moving bicycle, the hand undergoes a complex motion trajectory with respect to the static background. But, in fact, we perceive his hand moving vertically up and down discounting the motion of the bicycle. Likewise, we see the wheels of the bicycle rotating around their axles with the horizontal motion of the axles discounted. Hence, the circular motion of the wheel is perceived with respect to the translatory motion of the bicycle. The inadequacy of using the static environment as the single reference frame and the roles of moving reference frames have been systematically investigated by Johansson (1950, 1958, 1973, 1976, 1986) and many others (Duncker, 1929; Gogel, 1974; Hochberg & Fallon, 1976; Mori, 1979; Wallach et al., 1985, 1982).

Johansson claimed that the rotary motion of any point on the wheels can be deduced by perceptual subtraction of the translatory motion vector, common to both the hub and the wheel, from its “real” cycloidal motion (Johansson, 1950, 1976). The theory of perceptual vector analysis can explain rapid perception of highly complex motion patterns such as biological motion displays by a hierarchy of moving reference frames, thus, simplifying the motions of knees and feet as simple harmonic motion of a pendulum (Johansson, 1973, 1976). The gist of Johansson’s theory lies in the extraction of common and relative motion components. However, many studies demonstrated that the extraction of common motion is not always perfect (DiVita & Rock, 1997; Gogel, 1974; Hochberg & McAlister, 1953;

Johansson, 1974; Mori, 1979, 1984; Wallach, 1959). More importantly, in mathematical terms, vector decomposition is an ill-posed problem and needs additional information to be solved. Several constraints to limit the number of solutions have been introduced (Borjesson & von Hofsten, 1972; Cutting & Proffitt, 1982; Gershman et al., 2013; Gogel & Koslow, 1972; Gogel, 1974; Hochberg & McAlister, 1953; Proffitt et al., 1979; Restle, 1979). In short, these constraints provide heuristics to explain *why* the human visual system selects a particular solution. In the present study, we have taken the alternative perspective and looked at *how* a particular solution emerges through interactions between motion vectors.

Previously, we have shown that perceived motion of a target stimulus can be influenced by nearby motion of another object and that this object needs not to be surrounding the target stimulus as in the induced motion paradigm (Agaoglu, Herzog & Ogmen, in press). In this study, we started off by replicating our previous findings that each local motion vector has a reference field associated with it and this is manifested by increased illusory percepts of backward motion with decreasing distance to this moving reference frame. We then sought to determine whether these field-like effects of motion-based reference frames can also be extended to stationary landmarks. We presented a highly salient stationary grating along with a drifting one to examine if the effect of the latter is in any way modulated by the presence of the former. Although there was a consistent trend of reduced the effect size in the presence of a stationary grating, this reduction did not reach significance. More importantly, we did not find any significant interaction between distance and presence/absence of the stationary grating, suggesting that reference fields interact only when there is motion.

In order to investigate whether and how different reference fields interact with each other, we presented two drifting gratings at various distances from the target square. In different experiments, we manipulated the drift direction of the secondary grating while the primary grating always drifted in the same direction as the average target velocity. We found that when both gratings drift in the same direction, their effects combine and strengthen the illusory backward motion percepts. When the secondary grating drifts in a direction opposite to the direction of both the target and the primary grating, we found a significant drop in the illusory percept, i.e., the reference field effect. These drops, however, cannot be explained by linear summation of the reference fields. Taken together, these findings suggest that reference fields do interact and the way their effect combine is nonlinear and depends on how motion vectors are grouped.

These results clarify the details of interactions posited in the RFMF theory. The RFMF theory has been developed to explain how the visual system computes the attributes of stimuli under ecological conditions, i.e., when the observer and objects in the environment are in motion. Due to visible persistence, moving objects should appear extensively smeared, but under normal viewing conditions they do not (Hammett, 1997; Ramachandran et al., 1974). In addition, a moving object activates a retinotopically anchored receptive field only briefly, hence not allowing sufficient time for computation of stimulus attributes. As a result, one would expect moving objects to have a featureless “ghost-like” appearance (Ogmen, 2007; Ogmen & Herzog, 2010); however, under normal viewing conditions moving objects appear sharp and clear. RFMF suggests that the visual system avoids these problems by computing attributes of moving objects, not based on retinotopic coordinates, but instead according to a reference frame that moves with the object. To achieve this, as depicted in Fig. 4.2, a first stage of processing groups motion information and extracts reference frames that are used to compute attributes of moving objects. The use of non-retinotopic motion-based reference frames has been supported by several studies (Agaoglu et al., 2012; Boi et al., 2009; Hisakata et al., 2013; Kawabe, 2008; Nishida et al., 2007; Ogmen et al., 2006; Yamada & Kawabe, 2013). Based on our results, we can summarize reference frame rules as follows: First individual motion vectors are grouped according to their similarities (law of common fate). Each individual motion vector creates a field whose effect decays with distance. Fields of vectors that are grouped together reinforce each other. The field of a vector can weaken the effect of fields of vectors that form a different group; however, without being grouped with these motion vectors, it cannot revert the effect.

There are also other studies that focus more on how, rather than why, a particular solution emerges as a result of vector decomposition. Wallach and colleagues (1985) rejected the idea of perceptual vector analysis and interpreted the percepts which supposedly result from imperfect vector decomposition as a consequence of “process combination” (see also Johansson, 1985 and Wallach & Becklen, 1985). They claimed that there is no need for extraction of common motion and that the perceived motion patterns are nothing but incidental results of the sensory apparatus. In other words, component motions activate different kinds of motion processes and sometimes these processes can combine, resulting in motion percepts that deviate from what vector decomposition theory would predict. Take, for example, the two-dot display shown in Fig. 4.1E. According to process combination theory, individual motions of the dots, displacements of the group as a whole, and motion within the group activate different motion sensors. When a stationary landmark (a fixation point or a rectangular frame) is presented along with the two-dot display shown in Fig. 4.1E, observers mostly perceive the absolute motion paths, i.e., one dot oscillating vertically and the other horizontally because motions with respect to the stationary reference are enhanced and grouping of the dots is weakened (Wallach et al., 1985), which is also in line with RFMF theory.

Although indirect, there is some partial evidence supporting this position. The existence of motion sensors that are tuned to various types of moving patterns has been shown: The dorsal portion of the medial superior temporal area (MST) is known to be sensitive to global motion patterns whereas the ventrolateral portion is more sensitive to within-configuration or object motions in the scene (Duffy & Wurtz, 1995; Eifuku & Wurtz, 1998), and some MT (area V5) neurons are responsive to global motion of a plaid whereas others respond to motion of its individual sinusoidal components (Rust, Mante, Simoncelli, & Movshon, 2006). If a given type motion sensor is activated more compared to others, perceived motion can be mostly determined by the outcome of this process as a result of process combination. For instance, during steady fixation, relative motion determines percepts at slow speeds whereas absolute motion takes over at high speeds (Baker & Braddick, 1982; Snowden, 1992), which might be due to different levels of activation of corresponding motion sensors at different speeds. However, in contrast to this perspective built on hard-wired motion mechanisms, we suggest that the formation of reference frames is a dynamic process which is *adaptable* in real-time. The rationale for this is that under ecological viewing conditions, trajectories can be arbitrarily complex and it is not possible for the visual system to build hard-wired motion sensors for all possible trajectories. Hence, real-time field interactions between activities generated by a small set of canonical motion mechanisms can represent as a *neural-network state* solution that prevails under the given specific stimulus conditions.

A constraint that can play an important role in disambiguation vector decomposition is prior knowledge of the observers. For example, observers can readily recognize biological motion when the stimulus is presented in correct orientation but they fail to do so when it is inverted (Pavlova & Sokolov, 2000). This suggests that templates from memory can also help resolve ambiguity. Part of the reason why stimuli like simple dots shown in Fig. 4.1 generate multiple percepts may be due to the fact that they are not rich enough to engage specific memory patterns. Hence, in general learned figural configurations need also to be taken into account. Recently, Grossberg and colleagues (Grossberg et al., 2011) proposed a neural network model to explain how vector decomposition might occur by taking into account figural factors. According to their model, figure-ground separation and inhibition between neural populations, which represent motion at different depths, play the critical role. Near-to-far inhibition and the resultant peak-shift in the population activity leads to vector decomposition. Consistent with this model, it has been shown that surface decomposition leads to velocity decomposition (Watanabe, 1997). An important but rather implicit assumption of the model, which has not been tested formally, is that common (or coherent) motion is perceived to be at a different depth or scale than relative (or incoherent) motion: The former resides in a nearer depth or in a larger scale. Moreover, as acknowledged by the authors, their model in its current form cannot account for induced motion where a stationary object is seen to be moving in the opposite direction of a surrounding (or neighboring) moving object. This is partly due to a claim that vector decomposition and induced motion arise from different neural mechanisms (DiVita & Rock, 1997). For instance, induced motion is not perceived after the motion threshold (up to 3 deg/sec) for the frame is reached. Also in induced motion, inducer is perceived to be either moving at a lower speed or not moving at all whereas in vector decomposition stimuli, common and relative parts are perceived simultaneously. However, we obtained strong illusory perception of backward motions in the presence of a grating with a drift velocity of 9 deg/sec. Therefore, one can argue that vector analysis effect and induced motion stem from the same neural mechanism and that the reference-field effect reported here constitutes a special case of induced motion. Last but not the least, figure-ground segregation via perceptual grouping operations require at least some sort of form computations. Grossberg and colleagues' model extracts the form of the group of two dots via illusory contours, only then can common motion be calculated. It would be interesting to see how their model would respond to a "formless" motion stimuli. The RFMF theory predicts that retinotopic motion (without any form) is sufficient to generate a reference field.

Chapter 5. Motion-Based Nearest-Vector Metric for Reference Frame Selection in Perception of Motion

5.1. Introduction

Motion is defined as a change of position over time. Position is defined based on a reference frame (coordinate system) and, hence, the analysis of motion requires a reference frame. In physics, what makes a certain reference frame preferable or more convenient over others is its relevance to the context or its ability to represent phenomena in simpler terms. For instance, a major revolution in astronomy occurred with the shift from the geocentric to the heliocentric model. This new reference frame simplified the expressions for planets' motions by eliminating complex epicycles that were introduced to account for irregularities that arise in the geocentric reference frame. As in physics, in order to make sense of the complex motion trajectories of multiple objects and their parts, the perceptual system needs to choose an appropriate reference frame according to task demands. In psychology, reference frames can be classified into two broad types: endogenous and exogenous reference frames (Piaget & Inhelder, 1969; Swanston et al., 1987; Wade & Swanston, 1987).

Endogenous reference frames are those that are positioned on the subject. For example, a retinotopic reference frame is an endogenous reference frame positioned on the retina of the observer. Similarly, neurophysiological studies revealed endogenous reference frames positioned on the head, hands, etc. (Bremner & Andersen, 2012; Duhamel, Bremmer, Ben Hamed, & Graf, 1997; Galletti, Battaglini, & Fattori, 1995). The reference frame can also switch dynamically following behavioral demands. For example, Bremner and Andersen (2014) showed that the area 5d in monkey posterior parietal cortex codes the position of the hand relative to the gaze before the presentation of a reach target, but switches to coding the target location relative to the hand soon after the target is presented.

Exogenous reference frames are those that are external to the one's body. Exogenous reference frames can be positioned on stationary landmarks in the environment (e.g., in natural conditions: mountains, trees, horizon, etc., in laboratory conditions: screen edges, fixation points etc.), as well as on moving objects (e.g., a moving train, a running horse, clouds etc.) or motion itself (i.e., without reference to the object's figural properties) can serve as a reference frame.

In contrast to physics, the selection of a reference frame for perception cannot be done at will in most cases (try for example to perceive a stationary object in motion according to a retinotopic reference frame while you are moving your eyes). In addition, the selection of a reference frame is not simply a result of a winner-take-all competition among available frames of reference, but multiple reference frames can be utilized in combination by the perceptual system according to the demands of the task and the relevance of the prevailing the context. For instance, several studies showed that the effective reference frame can be expressed as a weighted combination of multiple reference frames (Agaoglu, Herzog, & Ogmen, in press; Freeman & Banks, 1998; Freeman, 2001; Souman, Hooge, & Wertheim, 2006; Swanston et al., 1987; Turano & Massof, 2001; Wade & Swanston, 1987; Wertheim, 1994).

In general, our perception is anchored in exogenous reference frames since, in most cases, the perception of our environment remains stable despite the movements of our eyes, head, and body. The early visual system is organized retinotopically and hence the initial coding of visual information is in retinotopic coordinates. Thus, the early retinotopic representations need to be transformed to representations based on exogenous reference frames. In the case of self-generated movements, these necessary transformations can be carried out by means of efference-copy signaling (see reviews: Bridgeman, Van der Heijden, & Velichkovsky, 1994; Wurtz, 2008). The neural motor-planning signals provide the brain with a means to predict retinotopic motions *before* they occur. Furthermore, the observer’s motion generates global and stereotypical retinotopic motion patterns such as translating, expanding optic flow which can be used to carry out reference-frame transformations (e.g., Gibson, 1986; Morrone et al., 2000; Rushton, Bradshaw, & Warren, 2007; Warren & Rushton, 2009). A more challenging situation arises in the case of motions of the objects in the environment since the retinotopic changes that occur due to the movements originating from objects external to the observer are neither predictable nor global. The brain has absolutely no information *in advance* about the changes in retinal motions as a result of motions of external objects. This needs to be computed online in real-time. Furthermore, due to visible persistence, moving objects would appear extensively blurred unless they are processed according to reference-frames based on their motion trajectories (Ogmen & Herzog, 2010; Ogmen, 2007). Since each object in the environment may move in a different direction and since motion trajectories can be arbitrarily complex, there must be reference-frame selection mechanisms that process visual motion. Recently, several studies showed visual information processing according to motion-based non-retinotopic reference frames (Agaoglu et al., 2012; Boi, Ogmen, et al., 2011; Boi et al., 2009; Kawabe, 2008; Léveillé et al., 2014; Nishida et al., 2007; Ogmen et al., 2006; Otto et al., 2009, 2010a; Shimozaki et al., 1999). The broad aim of this study was to characterize the processes underlying motion-based exogenous reference frames.

Recently, we have quantified the relative contributions of endogenous (retinotopic) and exogenous (spatiotopic and motion-based) reference frames on judgments of motion direction (Agaoglu, Herzog, & Ogmen, in press; submitted. See also Chapters 3 and 4). By doing so, we showed that the perceptual system selects/forms the effective reference-frame for the perception of motion through an amalgamation of both endogenous and exogenous reference frames, and that among the latter group, motion-based reference frames dominate the other available reference frames within a finite spatial region. In this study, we investigated how various spatiotemporal factors influence the selection and/or effectiveness of exogenous reference frames for motion perception during steady fixation. To this end, we used two concentric arcs (Fig. 5.1) undergoing circular motion around the center of display, where observers remained fixated. Using rotating arcs allowed us to keep the eccentricity of the stimuli constant. The outer arc’s angular velocity profile was modulated by a sine wave midflight whereas the other arc moved at a constant angular speed. The task of the observers was to report whether the outer arc reversed its direction of motion (from clockwise to counter clockwise or vice versa) at any point during its motion on the display. In order to study the effect of spatial factors, we varied the radial (Experiment 1) and the angular distance between the two arcs (Experiment 2). In order to investigate figural factors, we varied the relative size (Experiment 3) between the two arcs. Finally, to study the effect of velocity, we introduced varying fluctuations to the velocity of the inner arc also (Experiment 4) and quantified how this manipulation affects its ability to serve as a reference frame.

5.2. General Methods

5.2.1. Participants

Five naive observers and the author (MNA) participated in the study. The age of the participants ranged from 26 to 30 years and all participants had normal or corrected-to-normal vision. The experiments followed a protocol approved by the University of Houston Committee for the Protection of Human Subjects and were in accordance with federal regulations, the ethical principles established by the Belmont Report, and the principles expressed in the Declaration of Helsinki. Each observer gave written informed consent before the experiments.

5.2.2. Apparatus

Visual stimuli were created via visual stimulus generator card VSG2/5 (Cambridge Research Systems) and displayed at a resolution of 800 x 600 with a refresh rate of 100 Hz on a gamma-corrected Sony GDM-FW900 CRT monitor. Gaze position monitoring for both eyes was performed by means of an Eyelink-II eye-tracker at 250 Hz sampling rate. The distance between observer's eyes and the display was one m and the dimensions of the display at this distance were 22.7 x 17.0 deg^2. A head/chin rest was used to help stabilize fixation. Observers performed the task via a joystick.

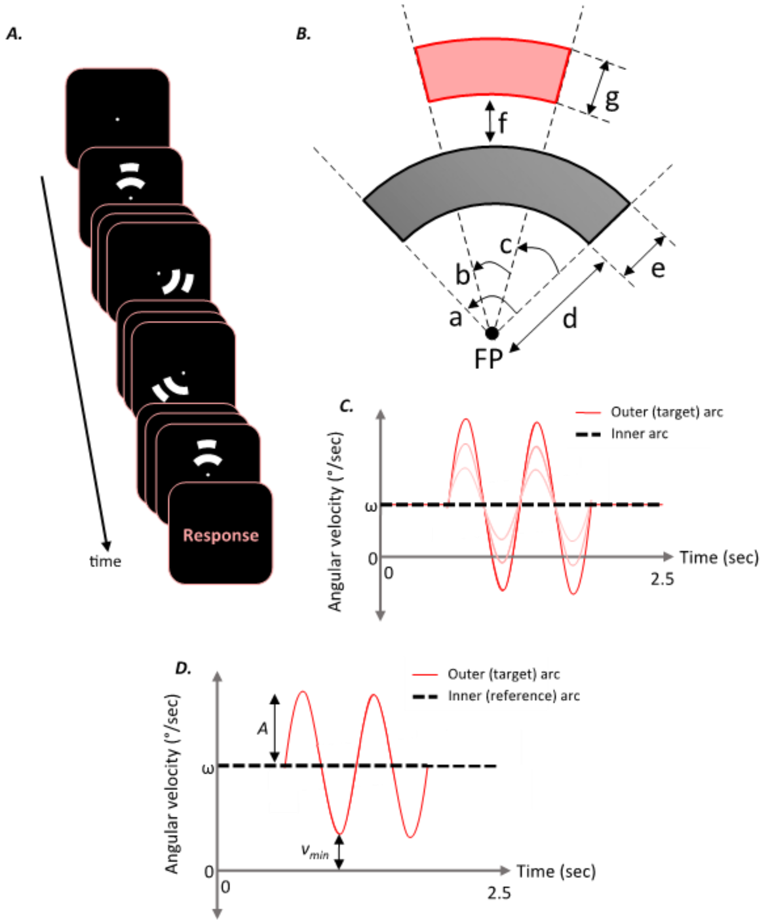

Figure 5.1. Spatial and temporal characteristics of the stimuli. A. An example stimulus presentation over time. B. The spatial and figural parameters of the two arcs. C. The angular velocity profiles. D. Given a specific modulation amplitude A, $v_{min}=w-A$ *and* $\kappa=A/w$.

5.2.3. Stimuli and Procedures

In all experiments, we investigated how perceived motion of an arc is influenced by another moving concentric arc. Fig. 5.1 shows the spatial and temporal characteristics of the stimuli used. Two white (56 cd/m^2) arcs moving along a circular path around the center of display against a black (<0.5 cd/m^2) background were used (Fig. 5.1A). The task of the observers was to report if the outer arc (hereafter referred as the target arc) reversed its direction of motion (from clockwise to counter clockwise or vice-versa) anytime during its presentation in the trial. The average angular speed of both arcs was 143°/sec so that one full cycle of circular motion was completed in 2.5 sec (all speeds are expressed in this manuscript in terms of rotational angles per second; rotational angles and visual angles are denoted by ° and deg, respectively). Since the distance between stimuli is known to affect reference frame selection (Gogel, 1974; Hochberg & McAlister, 1953; Mack & Herman, 1978; Mori, 1979), by using the same average speed for the two arcs, we kept the average distance between the arcs constant. The starting position of the arcs along the circular trajectory was selected randomly in each trial. The arcs were presented for only one full cycle of rotation. The direction of motion (i.e., clockwise or counter clockwise) was randomized across trials. An example of velocity profile for the target (outer) arc in a trial, where it rotated clockwise, is given in Fig. 5.1C (the thin red curves). The angular velocity of the target arc was constant during the first and last 630 ms. From 630 ms to 1890 ms, the velocity of the target arc was modulated by a sine wave. As long as the amplitude of the sine wave is smaller than the average angular velocity of the target, the target might decelerate and accelerate but never moves backward according to a retinotopic/spatiotopic reference frame (the light thin red curves in Fig. 5.1C). However, when the amplitude of the sine wave is chosen larger than the average speed, the target stops and reverses its direction for a short amount of time during its motion (dark thin red curves). The inner arc (hereafter referred as the reference arc) had a constant angular velocity profile in all experiments but Experiment 4, where its angular velocity profile was also modulated with varying amounts (denoted by κ in Table 5.1) in different blocks. Experiments were conducted in a normally illuminated room. A small point at the center of the display was provided throughout each trial to maintain proper fixation during stimulus presentation.

Trials during which gaze positions of observers deviated more than 2 deg from the fixation point were discarded and repeated immediately.

As soon as both arcs disappeared, observers were asked to report via a joystick whether the target reversed its direction of motion (from clockwise to counter clockwise or vice-versa) anytime during its presentation in the trial. The amplitude of the sine modulation in the target's velocity profile was varied across trials by an adaptive 1-up/1-down staircase algorithm (see various thin red lines showing different modulation amplitudes in Fig. 5.1C). For each reversal in observers' responses, the step-size in the staircase was halved. Four independent staircases with randomly chosen initial amplitudes were interleaved in a block of trials. Each staircase was completed in 15-35 trials. A staircase was considered "converged" when it underwent ten reversals and the last eight reversals were used to calculate the threshold for perceiving backward rotation. The minimum velocity of the target corresponding to this threshold amplitude (v_{min} in Fig. 5.1D) was taken as the point of subjective stationarity (PSS). For instance, if staircases converge to A= ω °/sec amplitude for sine modulation, it corresponds to the minimum target velocity of $v_{min}=(\omega-\omega) = 0$ °/sec. This would mean that backward rotation is *perceived* only when the target velocity goes below 0 °/sec ("veridical", i.e., spatiotopic percept), and hence, the PSS would be 0 °/sec. On the other hand, if, for instance, staircases converge to β °/sec, where $\beta < \omega$, corresponding to the minimum target velocity of $v_{min} = \omega-\beta$ °/sec, it would mean that as soon as the target velocity falls below $\omega-\beta$ °/sec, backward motion is *perceived* (illusory percept) although it never moves backwards in spatiotopic coordinates. Therefore, the PSS in this case would be $v_{min} = \omega-\beta$ °/sec. For each unique combination of stimulus parameters (a-g illustrated in Fig. 5.1B and modulation factor for the inner arc in

Experiment 4), each observer ran one block of trials (four staircases). Fig. 5.1B illustrates the arcs along with the parameters manipulated in different experiments and Table 5.1 summarizes the parameter sets used in all experiments.

In all experiments, we also ran a single block (four staircases) of baseline condition where the outer arc was presented alone. The PSS values obtained in the baseline conditions represent the response bias of observers. All PSS values reported here are corrected for observer bias by subtracting the PSS values in the baseline conditions from corresponding effect sizes in each experiment.

Table 5.1. Summary of parameter values used in all experiments. <c> denotes the average value of c. The parameters are defined in Fig. 5.1.

	a (°)	*b (°)*	*<c> (°)*	*d (deg)*	*f (deg)*	*(e,g) (deg,deg)*	*κ*
Exp. 1	90	30	30	5.5	1, 4, 7, 10.5	(0.5, 0.5)	0
Exp. 2	15, 45, 90, 180, 270, 360	30	-7.5, 7.5, 30, 60, 120, ring	5.5	1	(0.5, 0.5)	0
Exp. 3	90	30	30	5.9, 5.5, 5.5, 5.5, 5.5, 2.75, 0	1	(0.1,0.5), (0.5,0.5), (0.5,3), (0.5,6), (0.5,10), (3.25,0.5), (6,0.5)	0
Exp. 4	90	30	30	5.5	1	(0.5, 0.5)	0, 0.1, 0.2, 0.3, 0.5

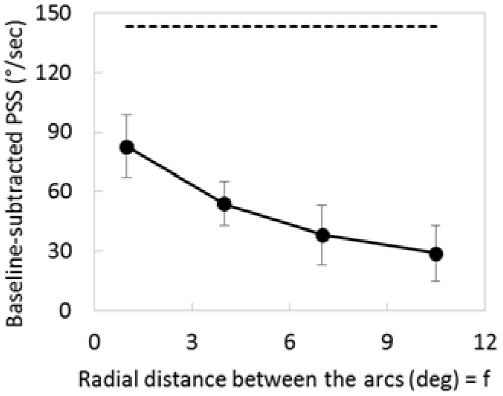

Figure 5.2. Baseline-subtracted PSS values as a function of the radial distance between the arcs in Experiment 1 (markers). The horizontal dotted line represents the prediction of perfect vector decomposition. Error bars represent ±SEM across observers (n=6).

5.3. Experiment 1: Spatial Factors: Radial Distance

In order to investigate how spatial factors influence the selection of motion-based reference frames, we varied the radial distance between the two arcs (f in Fig. 5.1B, the radial distance between the closest contours of the two arcs).

5.3.1. Results and Discussion

Table 5.1 summarizes the parameter values used in this experiment. Fig. 5.2 shows the baseline-subtracted PSS values (see General Methods) averaged across observers as a function of the radial distance between the rotating arcs. The results show a decreasing trend with increasing spatial separation between the two arcs. A one-way repeated measures ANOVA showed a significant effect of radial distance ($F_{3,13}$=9.969, p=0.001, η_p^2=0.697). Note that all data points are above zero, indicating that within the range of distances tested here, percepts were never veridical. These results are consistent with previous accounts of distance-dependent effects of moving reference frames (Agaoglu, Herzog, & Ogmen, submitted; Agaoglu et al., in press; Gogel & Koslow, 1972; Gogel, 1974; Hochberg & Fallon, 1976; Mori, 1979; Shum & Wolford, 1983). However, these results are inconsistent with complete extraction of common motion from the outer arc as predicted by the perceptual vector decomposition theory (Borjesson & von Hofsten, 1975; Johansson, 1950, 1973). According to the vector decomposition theory, the common angular velocity of the two arcs should be perceptually subtracted from the outer target arc, and hence, a slight deceleration in the angular velocity of the target arc should lead to backward motion percepts. In other words, all PSS values should lie on the horizontal dashed line in Fig. 5.2. However, all data points are well below of what is predicted from a perfect common motion extraction point of view.

Previously, we have shown similar distance-dependent effects with a variant of the stimuli used here (Agaoglu et al., submitted; in press). Instead of rotating concentric arcs, we used two horizontally moving disks, one translating with a constant velocity profile whereas the other's velocity profile was modulated by a sine wave as was the case for the outer arc in the present study. The PSS values showed a linear distance-dependent decrease with horizontally moving disks as well; however, the overall extent to which the extraction of common motion occurs (measured by the ratio of empirical PSS values and those predicted from perfect vector decomposition for the closest spatial separation) was significantly larger than what is reported here (~0.85 with translational motion vs. ~0.55 with rotational motion). Mori (1984) reported that speed also can influence the selection of the reference frame. However, the quantitative difference between the distance effects in the two studies cannot be explained by different speeds because the average linear speed of the moving elements were roughly the same in these studies. Bertamini and Proffitt (2000) assessed the degree to which different types of motion can serve as a reference frame, and found that translation and divergence are superior to rotation. Hence, a plausible explanation for the quantitative difference between the distance effects could be the ability of the perceptual system to establish reference frames based on translational vs. rotational motion in the fronto-parallel plane.

5.4. Experiment 2: Spatial Factors: Angular-contour Distance

While the contours of the arc with elongations that are parallel to the direction of motion cannot provide a reference frame for that motion, the contours perpendicular to the direction of motion can. In fact, since the surface of the arc is uniform, the rotational motion information is generated at the leading and trailing contours that are perpendicular to motion direction. Hence, another way to manipulate the distance between the motion vectors of the reference and the target arcs is to vary the angular distance between the edges of the arcs. In this experiment, the radial distance between the two arcs was kept fixed at 1 deg and the angular-contour distance between the edges (see Fig. 5.1A) was varied systematically. The angular size of the target arc was always 30^o whereas the angular size of the reference arc took one of the following values in a block of trials: 15^o, 45^o, 90^o, 180^o, 270^o, 360^o. The corresponding angular-contour distances are -7.5, 7.5, 30, 75, and 120. When the inner arc's angular span was 360°, it becomes a ring, therefore there is no angular-contour in this case. The parameter values used are summarized in Table 5.1.

5.4.1. Results and Discussion

Fig. 5.3 shows the baseline-subtracted PSS values averaged across observers as a function of mean angular-contour distance. When the inner arc's angular span was 360°, it becomes a ring. In this case, the rotation of the ring cannot be perceived since its surface is homogenous. In addition, since the contours of the ring are parallel to the direction of motion, they cannot serve as reference and hence the results are, as expected, identical to the baseline condition, yielding a zero baseline-subtracted PSS. Whenever the angular extent of the inner arc was less than 360^o, the reference arc appeared to rotate and illusory percepts of direction reversals were perceived, as indicated by positive PSS values in Fig. 5.3.

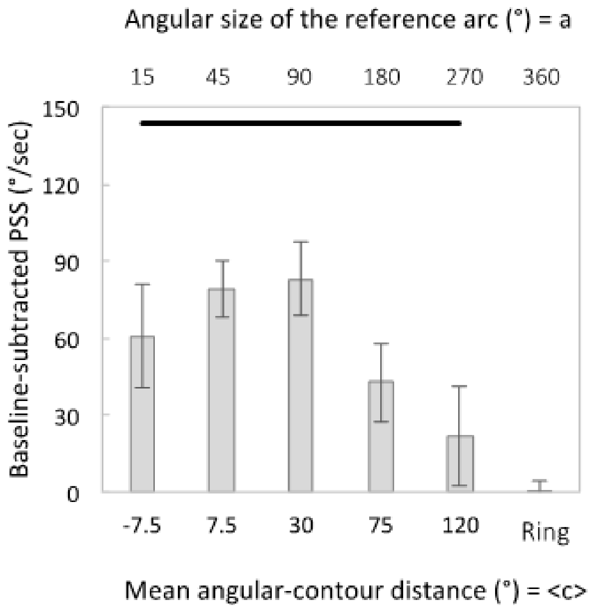

Figure 5.3. Baseline-subtracted PSS values in Experiment 2 are plotted as a function of mean angular-contour distance. On a secondary x-axis, the angular size of the reference arc is also shown. The horizontal line represents again the predictions of perfect vector decomposition.

A one-way repeated measures ANOVA showed a significant effect of angular-contour distance of the PSS values ($F_{5,25}$=11.718, p<0.001, η_p^2=0.701), indicating that the effectiveness of the inner arc as a reference frame for the motion of the outer (target) arc was strongly modulated by the changes in angular-contour distance between the two. The condition with <c> = 30^o is identical to the condition in Experiment 1 with f = 1 deg and we found similar PSS values in the two Experiments. Increasing <c> beyond 30^o caused a decrease in the reference frame effect, consistent with the distance-dependent decreases observed in Experiment 1. However, bringing the contours of the reference disk closer to those of the target disk (<c> < 30^o) did not cause a further increase in the effect size. Moreover, particularly interesting is the comparison of the two cases when the angular size of the reference arc is smaller than the target arc, and when it is larger (i.e., with radial-contour distances of -7.5° and 7.5° respectively). Although there is an apparent drop in the effect size when the inner arc is smaller than the target, this difference did not reach significance (paired t-test: t_5=-1.349, p=0.235).

5.5. Experiment 3: Figural Factors: Radial Size

Duncker (1929) proposed a principle called “the stationarity tendency of large stimuli” and suggested that large stimuli tend to serve as reference. In Experiment 2, we looked at the effect of radial-contour distance between the two arcs on how the target arc’s motion is perceived. A change in radial-contour distance was accompanied by a change in the radial size of the reference arc. In order to investigate the effect of size more directly, we kept the closest-contour radial distance and the angular-contour distances between the arcs constant and varied the sizes of the arcs by changing their thickness. Parameters used in this experiment are also summarized in Table 5.1.

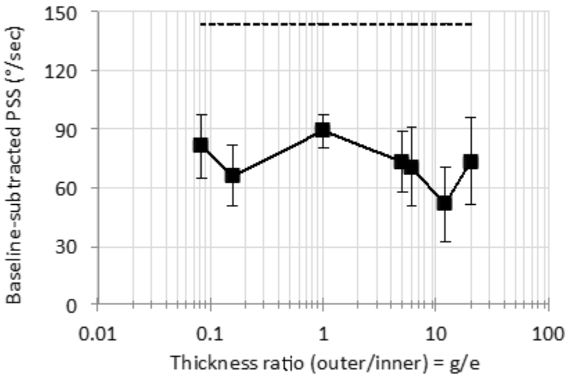

Figure 5.4. Baseline-subtracted PSS values in Experiment 3 are plotted as a function of the thickness ratio of the two arcs. The horizontal dashed line represents the prediction of perfect vector decomposition. Error bars represent ±SEM across observers (n=6).

5.5.1. Results and Discussion

Fig. 5.4 shows the average baseline-subtracted PSSs as a function of thickness ratio of the two arcs. There is no discernable pattern in the results suggesting that relative size of the arcs does not influence the perceived motion. A one-way repeated measures ANOVA revealed no significant effect of relative thickness ($F_{6,30}$=1.038, p=0.421, η_p^2=0.172).

5.6. Experiment 4: Constant Motion

The results of Experiment 3 suggest that the figural aspects of stimuli do not play a systematic role in determining the selection of reference-frames and the results of Experiments 1 and 2 suggest that the distance with respect to *motion* vectors has an important influence. The goal of the fourth experiment was to examine further how the motion of the stimuli influences reference frame selection. In Experiment 4, the velocity profile of the inner arc was also modulated by a sine wave (Fig. 5.5B) to see whether a reference frame is required to have a constant motion. In fact, previous studies suggested that constant motion is more likely to serve as a reference frame (Cutting & Proffitt, 1982; Rubin & Richards, 1988). In different blocks, the amplitude of modulation and correspondingly the minimum velocity was different. Phases of sine wave modulations of the velocity of both arcs were equal so that they decelerated and accelerated with the same time course.

5.6.1. Results and Discussion

Results are given Fig. 5.5. If common motion of the arcs is extracted perfectly, the target arc should be perceived as reversing its direction of rotation only when its angular velocity goes below that of the inner arc. Therefore, PSS values should be equal to the minimum velocity of the inner arc, as depicted by the dashed line in Fig. 5.5. However, all PSS values fall well below this line indicating, once again, that common motion extraction is incomplete. A one-way repeated measures ANOVA showed a significant effect of level of velocity modulation of the inner arc (i.e., its minimum velocity) ($F_{4,20}=9.246$, $p<0.001$, $\eta_p^2=0.649$). These results suggest that in order for the inner arc to serve as a reference frame for the motion of the target arc, it need not to have a constant velocity profile: in all levels of velocity modulation used here, PSS values were significantly different from zero, which indicates illusory percepts of rotation reversals. Moreover, the fact that PSS values increase with increasing minimum velocity of the inner arc (i.e., decreasing modulation in its

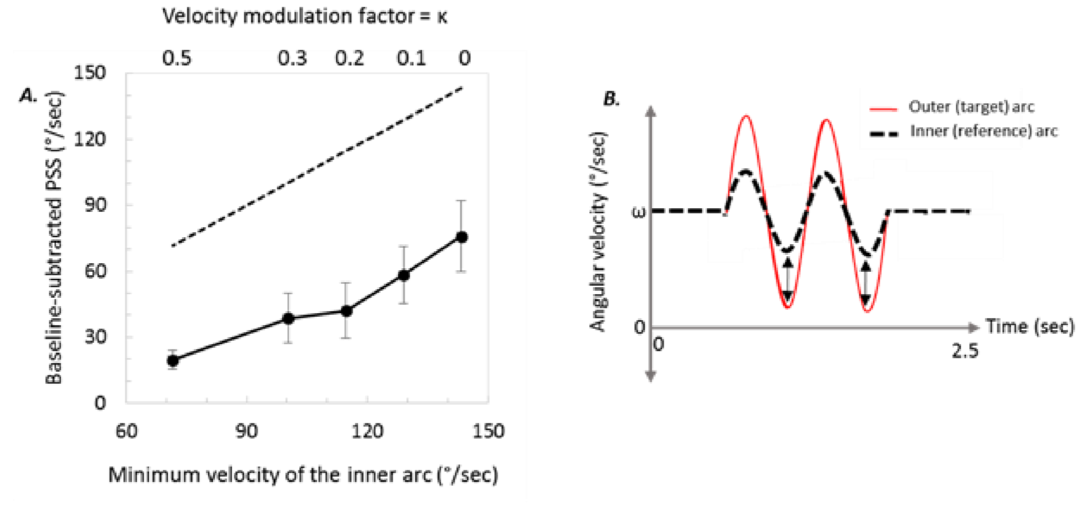

Figure 5.5. A. Baseline-subtracted PSS values as a function of the minimum angular velocity of the reference arc. On a secondary x-axis, the corresponding velocity modulation factors (κ) are given. B. An example of velocity profiles of the target and reference arc in Experiment 4.

velocity), does show that the relevant parameter of the reference arc's motion in serving as a reference frame is not its average velocity but its time-modulated velocity profile. In fact, a one-way repeated measures ANOVA on reference frame effectiveness, defined as the difference between the empirical PSS values (markers in Fig. 5.5) and those predicted from perfect vector decomposition (dashed line in Fig. 5.5), revealed no effect of level of velocity modulation of the reference arc ($F_{4,20}$=1.413, p=0.266, η_p^2=0.220). These results suggest that as long as the difference between the minimum angular velocities of the two arcs (double-headed arrows in Fig. 5.5B) is kept constant, the effectiveness of the inner arc as a reference frame does not change.

After completion of all experiments, all observers were asked to verbally report whether they were aware of the fact that in Experiment 4, the inner arc's velocity profile was also modulated by various amounts. Surprisingly, all five naïve observers reported that they were not, which suggests that the inner arc was perceived as rotating at a constant angular velocity. When attention was allocated to the target arc, the reference arc *appeared* to move with a constant angular velocity while *its actual time-varying velocity profile* determined how the motion of the target arc was perceived. This observation suggests that *the variations of the velocity profile* for the reference arc serve as reference for both the target and the reference arc itself. The average angular velocities of the two arcs are equal to each other and remain constant. This common motion is attributed to both arcs, as they are perceived rotating with this average velocity. The variations of velocity are judged with respect to variations of the reference arc in a distant-dependent manner. Since the distance of the reference arc to itself is zero, the variations of the reference frame match perfectly the variations of its own velocity and hence it appears to move at a constant velocity. On the other hand, the target arc being distant from the reference frame, the effect of the reference frame is only partial, resulting in perceived variations in its velocity profile. For the velocity modulation factor values used here, it was difficult to perceive the modulations in the velocity of the reference arc. In a previous study, we have shown that attention modulates the strength of reference frames (Noory, Herzog, & Ogmen, under review), hence we would predict with focused attention and large modulation factors, one can also observe the modulations of the reference arc.

5.7. A Unifying Metric of Non-retinotopic Reference Frames for Motion Perception

In the experiments presented in this study, we characterized non-retinotopic reference-frames for motion perception. In summary, we found that the effect of the reference arc's motion on the perceived motion of the target arc decreases with increasing radial and angular-contour distances, and is independent of the relative radial size of the two arcs. We sought to find a metric which can unify all of the findings reported in this study. To this end, we considered the four metrics illustrated in Fig. 5.6. The first two of these metrics, namely the object-centered and the object-nearest-contour, are form-based (Fig. 5.6A-B) while the other two are motion-based (Fig. 5.6C-D) metrics. If reference frames for motion perception are object-based, changes in the center-to-center or the radial distance between the arcs should account for the changes in the perceived motion of the target arc. Although these two metrics can explain the results in Experiment 1, they fall short in explaining the data of Experiment 2 in which the angular size of the reference arc is varied while distance according to these two metrics is kept constant. Both the object-centered and the object-nearest-contour metrics predict no change in the effect-sizes in this case. However, as our results in Experiment 2 show, that is not case. Furthermore, the object-centered metric can also be ruled out by the results of Experiment 3, since the changes in the radial thicknesses of each arc did not result in changes in the effect-size, although the center-to-center distance between the arcs did change.

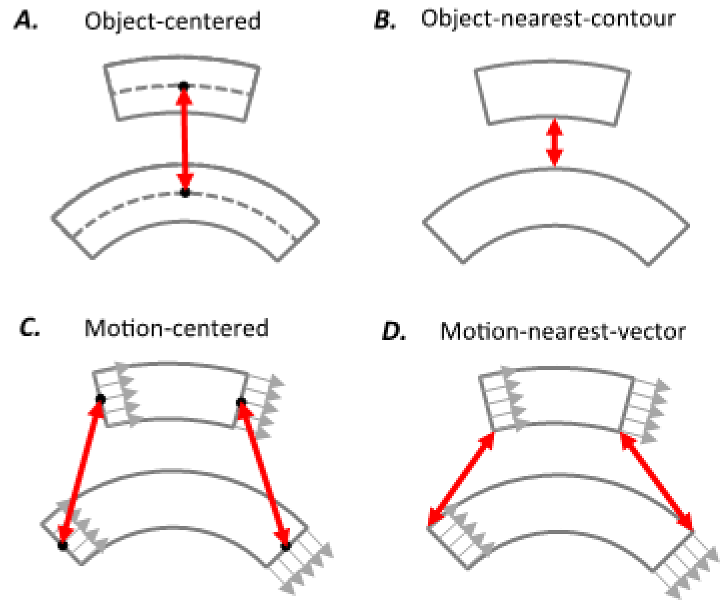

Figure 5.6. Illustration of various metrics that are considered in this study. The red double-headed arrows represent the corresponding metric in each part. In parts C and D, the gray arrows indicate the motion vectors when the direction of rotation of the arcs is clockwise.

Since the arcs underwent rotational motions, motion vectors are only generated at the leading and trailing contours that are perpendicular to motion direction (Fig. 5.6C-D). Another way of defining a metric is to consider the distance between the midpoints of the leading or trailing edges of the two arcs. A motion-centered distance metric is ruled out by the results of Experiment 3 for a similar reason given for object-centered metric: the distance between the midpoints of the leading or trailing edges of the arcs changes while

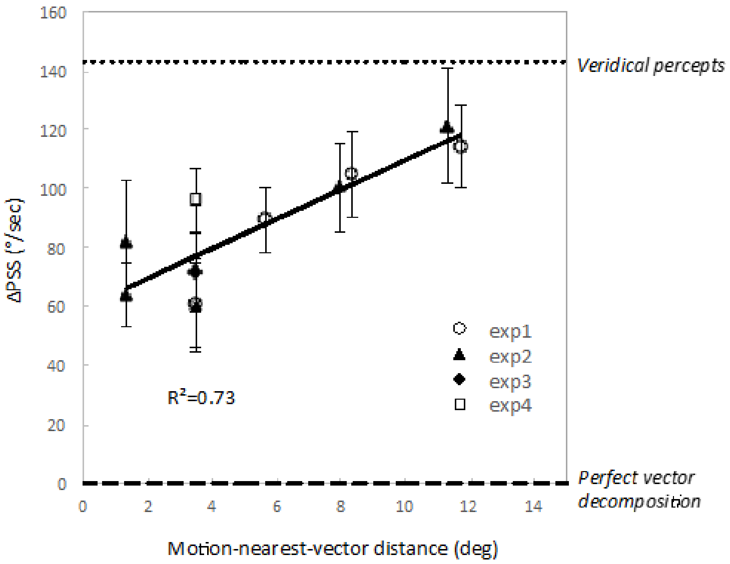

Figure 5.7. The differences between the predictions of perfect vector decomposition and the baseline-subtracted PSS values in each experiment are plotted as a function of the motion-nearest-vector metric. Error bars represent SEM across subjects (n=6).

the motion reversal thresholds in Experiment 3 do not. The last metric we considered was the motion-nearest-vector metric (Fig. 5.6D). This metric is defined as the distance between the nearest motion vectors of the two rotating arcs (denoted by the red double-headed arrow in Fig. 5.6D). Since the distance defined by this metric does not change with changing relative thicknesses in Experiment 3 and varying modulations in the reference arc's angular velocity in Experiment 4, this metric predicts no change in effectiveness of the inner arc as a reference in these experiments.

The effectiveness of the inner arc as a reference frame can be quantified by subtracting the empirical PSS values from those that are predicted from perfect vector decomposition. In order to illustrate how well the motion-nearest-vector metric accounts for all the data reported here, we plotted the effectiveness of the reference frame from all experiments against this metric in Fig. 5.7. A regression analysis revealed a simple linear relationship between the motion-nearest-vector metric and the reference frame effectiveness for motion perception.

5.8. General Discussion

A representation can be conceptualized by its coordinate system (reference frame). The human visual system is one of the primary sensory systems in which the physics of the sensory organ shapes the corresponding early representations. Neighboring elements in the environment stimulate neighboring photoreceptors in the retina, and these *retinotopic* relations are preserved in early visual cortices (Sereno et al., 1995; Tootell et al., 1982). Most theories of vision are based on computations on a retinotopic reference frame and/or make use of features extracted by retinotopically organized receptive fields. However, under normal viewing conditions, retinotopic representations are highly dynamic and unstable due to object and observer movements, which render retinotopically based theories insufficient to explain clarity and stability of perception under dynamic conditions (Ogmen & Herzog, 2010; Ogmen, 2007). In fact, many visual processes, which have been previously thought to occur in retinotopic coordinates, have been shown to result from computations in non-retinotopic reference frames (e.g., form: Agaoglu et al., 2012; Ogmen et al., 2006; Otto, Ogmen, & Herzog, 2006; luminance: Shimozaki et al., 1999; color: Nishida et al., 2007; attention: Boi, Vergeer, Ogmen, & Herzog, 2011; size: Kawabe, 2008; and motion: Boi et al., 2009). Therefore, the following questions become central for an adequate understanding of vision under ecological viewing conditions:

(1) What are the non-retinotopic reference frames?

(2) Why and how are particular non-retinotopic reference frames constructed?

(3) What factors influence the selection of non-retinotopic reference frames?

These have been enduring questions in ecological perception literature, in particular for motion perception (Gibson, 1979; Johansson et al., 1980; Johansson, 1950). As the well-known Gestalt principle of common fate suggests, motion signals influence form processing; objects or elements that share a common motion component are grouped into a single Gestalt. On the other hand, the perceived motion of a stimulus depends also on its own spatiotemporal properties. Motion trajectories and object properties such as elongation, symmetry axes, closest oriented element, spatial frequency, and presentation duration have been shown to affect perceived direction of motion (e.g., Freeman & Banks, 1998; Löffler & Orbach, 1999, 2001; Magnussen, Orbach, & Loffler, 2013). Moreover, perceived motion of a stimulus is also influenced by the properties of spatiotemporally neighboring stimuli. Karl Duncker was one of the first scientists to investigate this issue systematically. He found, for instance, that a slow movement of a large surrounding frame induces an illusory motion in the opposite direction in a stationary dot while its own motion is often barely visible (Duncker, 1929). He explained his findings by a Gestalt-like principle called the “stationarity tendency of large stimuli”. In other words, larger stimuli tend to be taken as a reference frame. However, later studies revealed that the inducer object does not have be larger in size to produce the illusion (e.g., Day, 1978; Wallach, 1959). Johansson used point-light displays to investigate how motion is perceived in the presence of a multitude of moving objects (Johansson, 1950, 1973). He proposed that the perceptual system decomposes the motion of each element in the display into common and relative components and the common component serves as the reference frame. In short, a possible answer to the first question has been the common motion among moving elements in the display so far.

Although it seems intuitive, in mathematical terms, vector decomposition is an ill-posed problem. Infinitely many different pairs of common and relative motion components can make up the same absolute motion vector. However, the percepts form only a small subset of possible solutions. For instance, when presented with the Duncker's wheel stimuli, where two point-lights are attached to the hub and the rim of an otherwise invisible rotating and translating wheel (Duncker, 1929), some observers reported rotational motion for the light at the rim whereas some others reported cycloidal motion (Johansson, 1974; Proffitt et al., 1979; Proffitt & Cutting, 1980; Shum & Wolford, 1983). A "tumbling stick" percept has also been reported (Mori, 1984). Therefore, it is evident from a mathematical point of view that additional information (or constraints) is needed to reduce the number of solutions to the vector decomposition problem. Various constraints (such as minimum information load, minimum relative motion, and zero sum of residual motion vectors) have been proposed to explain how the visual system regularizes vector decomposition (Borjesson & von Hofsten, 1972; Cutting & Proffitt, 1982; Gershman et al., 2013; Gogel & Koslow, 1972; Hochberg & McAlister, 1953; Proffitt et al., 1979; Restle, 1979). In other words, the constraints introduced in these studies provide heuristics to explain why the visual system selects particular non-retinotopic reference frames for motion perception. Moreover, Grossberg, Léveillé, & Versace (2011) showed how vector decomposition might emerge as a result of figure-ground segmentation and inhibition between neuronal populations, which represent motion at different depths. Recently, taking an alternative approach to vector decomposition, we have proposed that non-retinotopic reference frames are generated as a result of field-like interactions between motion vectors (Agaoglu et al., submitted; Noory et al., under review). In summary, the second question outlined above has at least a partial answer.

Although motion is ubiquitous under ecological viewing conditions and despite the need for non-retinotopic reference frames due to lack of clarity and stability in the retinotopic representations, studies examining the factors which influence the selection of non-retinotopic reference frames for motion perception (the third question given above) have been limited. Mori (1984) investigated how a reference frame for motion is affected by velocity by using the Duncker's wheel stimuli. He found that, at slow speeds (0.1-0.7 deg/s), the motion of the midpoint between the lights at the rim and the hub was the reference frame such that the resultant percepts can be described as a tumbling stick motion. At medium-speeds (0.8-40.0 deg/s), he found that the motion of the light at the hub served as the reference frame, which is indicated by the perception of rotational motion for the light at the rim. Lastly, for higher speeds (>40.0 deg/s), an absolute coordinate system, i.e., a spatiotopic reference frame, was chosen and percepts corresponded to veridical motions of the two lights. Yoshimoto, Uchida-Ota, and Takeuchi (2014) recently examined the effect of light level on reference frame selection for motion processing by using a motion priming task. They found positive motion priming in spatiotopic coordinates in photopic and scotopic conditions but not in mesopic conditions.

The aim of the present study was to further characterize non-retinotopic reference frames for motion perception under various manipulations in the spatiotemporal properties of the stimuli. We used two concentric arcs rotating with the same average angular velocity, and asked observers to report whether the target arc has ever reversed its direction of rotation throughout its motion on the display. We found that the radial and angular-contour distances between the target and the reference arc significantly affect the perceived motion of the former. We found a linear drop in the effectiveness of the inner arc as a reference frame with increase in both distance measures. Drastic changes in the relative sizes of the arcs did not influence motion reversal thresholds, suggesting that the strength of a reference frame for motion processing does not depend on stimulus size. Finally, we found no effect of whether a reference frame is defined by constant motion or not. What mattered was the difference between the minimum values of the velocities of the two arcs. We examined several metrics, namely object-centered, object-nearest-contour, motion-centered, and motion-based-nearest-vector, which could potentially unify all of our findings. We concluded that the motion-based-nearest-vector distance metric can fully account for all the data reported here.

Recently, we have taken an approach alternative to Johansson's vector decomposition theory with the goal of explaining reference frame selection as a result of interactions between motion vectors (Ogmen & Herzog, 2010; Ogmen, 2007). According to Reference-Frame Metric Field (RFMF) theory, each retinotopic motion vector generates a field in the retinotopic space (like an electromagnetic field). Fields created by different motion vectors interact in order to determine a motion vector that will serve as the reference-frame at a given point in space and time. Previously, we adopted a variant of a Ternus-Pikler display to investigate the spatial extent of a non-retinotopic reference frame field (Noory et al., under review; Ogmen, Herzog, & Noory, 2013a, 2013b).We also examined whether the size of the elements producing the non-retinotopic reference frame influence the strength of the field effect. In agreement with the present findings, our results showed that the field effect of a non-retinotopic reference-frame is independent of the size of the inducing elements. Taken together, these findings suggest that the selection of non-retinotopic reference frames for motion processing is not a result of an all-or-none or winner-take-all process, but instead, can be explained by a field whose strength decreases linearly as a function of the distance between the nearest motion vectors.

Chapter 6. Summary and Conclusions

The spatial representation of a visual scene in the early visual system is well known. The optics of the eye map the three-dimensional environment onto two-dimensional images on the retina. These retinotopic representations are preserved in the early visual system. Retinotopic representations and processing are among the most prevalent concepts in visual neuroscience. However, it has long been known that a retinotopic representation of the stimulus is neither sufficient nor necessary for perception. Saccadic Stimulus Presentation Paradigm and the Ternus-Pikler displays have been used to investigate non-retinotopic processes with and without eye movements, respectively. However, neither of these paradigms eliminates the retinotopic representation of stimulus. In Chapter 2, we investigated how stimulus features are processed in the absence of a retinotopic layout and in the presence of retinotopic conflict. We combined anorthoscopic viewing (slit viewing) and the Ternus-Pikler display and pitted a retinotopic feature-processing hypothesis against a non-retinotopic feature-processing hypothesis. Our results support the predictions of the non-retinotopic feature-processing hypothesis and demonstrate the ability of the visual system to operate non-retinotopically at a fine feature processing level in the absence of a retinotopic spatial layout. Our results suggest that perceptual space is actively constructed from the perceptual dimension of motion. In other words, these non-retinotopic representations uses motion as a reference frame to process form.

The retinotopic projection of stimulus motion depends both on the motion of the stimulus and the movements of the observer. In experiments reported in Chapters 3, 4, and 5, we investigated how the visual system selects a reference frame for the perception of motion. In Chapter 3, we aimed to quantify the contributions of endogenous (retinotopic) and exogenous (spatiotopic and motion-based) reference frames on judgments of motion direction. We used a variant of the induced motion paradigm and we created different experimental conditions in which the predictions of each reference frame were different. Finally, assuming additive contributions from different reference frames, we used a linear model to account for the data. Our results suggest that the effective reference frame for motion perception emerges from an amalgamation of motion-based, retinotopic and spatiotopic reference frames. In determining the percept, the influence of relative motion, defined by a motion-based reference frame, dominates those of retinotopic and spatiotopic motions within a finite region. We interpret these findings within the context of the Reference Frame Metric Field (RFMF) theory, which states that local motion vectors might have perceptual reference-frame fields associated with them, and interactions between these fields determine the selection of the effective reference frame.

A reference-frame is required to specify how motion is perceived. For example, the motion of a part of an object is usually perceived relative to the motion of the object itself. Johansson proposed that the perceptual system carries out a vector decomposition, which results in common and relative motion percepts. Because vector decomposition is an ill-posed problem, several studies introduced constraints by which the number of solutions can be substantially reduced. In Chapter 4, we adopted an alternative approach and studied how, rather than why, a subset of solutions is selected by the visual system. We propose that each retinotopic motion vector creates a reference-frame field in the retinotopic space and fields created by different motion vectors interact in order to determine a motion vector that will serve as the reference-frame at a given point and time in space. To test this theory we performed a set of psychophysical experiments. The field-like influence of motion-based reference frames was manifested by increased illusory percepts of backward motion of a target square with decreasing distance to a drifting grating. We then sought to determine whether these field-like effects of motion-based reference frames can also be extended to stationary landmarks. Our results suggest that reference field interactions occur only between motion generated fields. Finally, we investigated whether and how different reference-fields interact with each other, and found that different reference field interactions are nonlinear and depend on how motion vectors are grouped. These findings supports the perspective of the RFMF theory on dynamic form perception (see Chapter 1).

Finally, in Chapter 5, we examined the effect of several spatiotemporal factors on the perception of motion to determine the best metric which can explain how the reference frame selection is done in the brain. Two concentric arcs underwent circular motion around the center of the display, where observers remained fixated. The outer (target) arc's angular velocity profile was modulated by a sine-wave midflight whereas the inner (reference) arc moved at a constant angular speed. The task of the observers was to report whether the outer arc reversed its direction of motion (from clockwise to counter clockwise or vice versa) at any point during its motion on the display. In order to study the effect of spatial factors, we varied the radial and the angular distances between the two arcs. In order to investigate figural factors, we varied the relative size between the two arcs. Finally, to study the effect of velocity, we introduced varying fluctuations to the velocity of the inner arc also. We found that the radial and angular-contour distances between the target and the reference arc significantly affect the perceived motion of the former. A decrease is observed in the effectiveness of the inner arc as a reference-frame with an increase in both distance measures. Drastic changes in the relative sizes of the arcs did not influence motion reversal thresholds, suggesting that the strength of a reference-frame for motion processing does not depend on stimulus size. Finally, we found no effect of whether a reference frame is defined by constant motion or not. What mattered was the difference between the minimum values of the velocities of the two arcs. We examined several metrics, namely object-centered, object-nearest-contour, motion-centered, and motion-based-nearest-vector, which could potentially unify all of our findings. We found that the *motion-based-nearest-vector* metric can fully account for all the data reported here. These findings suggest that the selection of non-retinotopic reference-frames for motion processing is not a result of an all-or-none or winner-take-all process, but instead, can be explained by a field whose strength decreases linearly as a function of the distance between the nearest motion vectors (Agaoglu, Herzog, & Ogmen, submitted-b).

Taken together, our results from all motion experiments suggest that the reference frame for the perception of motion emerges as a result of interactions between motion vectors in a field-like manner. The strength of these interactions depends on the distance between the nearest motion vectors and the similarity of their directions. Finally, antagonistic interactions are nonlinear such that the fields generated by motion vectors in opposite directions exert inhibitory effects but they do not simply add up or cancel out each other.

Chapter 7. Future Directions

Early vision is retinotopic but perception is non-retinotopic. Therefore, a reference-frame transform should take place to construct non-retinotopic representations. The choice of the reference frame and the required coordinate transformations occur automatically in perception. The exact nature of this coordinate transform can only be established by determining the non-retinotopic reference frames and how they are constructed. In this dissertation, we investigated the non-retinotopic reference frames and how they are selected by the visual system. We proposed a qualitative theory of dynamic form perception and tested its predictions in a series of experiments. According to the RFMF theory, non-retinotopic representations are constructed from local reference frames that are generated by local motion vectors, and features are attributed based on these local reference frames. Our results confirmed these predictions and generated new insights on reference frame transformations.

Based on these findings, a quantitative formulation of the RFMF theory can be established in the future. In mathematical terms, the field equation for a reference frame can be expressed as

$$v_{ref}(x_r, y_r, t_r) = \Phi(x_r, y_r, t_r; p), \tag{2}$$

where v_{ref} is the reference frame that is generated from retinotopic coordinates x_r and y_r at time t_r. $\Phi(.)$ is the field equation that depends on a set of parameters p, such as common motion vectors, their location with respect to each other, their strengths etc.

In mathematical terms, a change in reference frame (i.e., coordinate transform) can be expressed as

$$\boldsymbol{c_{nr}} = \boldsymbol{T}\boldsymbol{c_{r}}, \tag{3}$$

where $\mathbf{c_{nr}}$ and $\mathbf{c_{r}}$ are non-retinotopic and retinotopic coordinate vectors and **T** is the transform matrix. **T** is a function of reference frame vector v_{ref}. For simplicity let us assume one dimensional space. Then if the transform is Galilean, then

$$\mathbf{T} = \begin{bmatrix} 1 & -v_{ref} \\ 0 & 1 \end{bmatrix}, \tag{4}$$

resulting in,

$$x_{nr} = x_r - v_{ref}t_r \text{ and} \tag{5}$$

$$t_{nr} = t_r, \tag{6}$$

which represent a velocity dependent position shift. The Galilean transform also implies independence of time dimension from speed. However, it has long been known that the perceived locations and durations vary as a function of movement of the objects (Baro, Brzezicki, Lehmkuhle, & Hughes, 1992; Bhatia & Verghese, 1964; Brown, 1931; Hughes, Lishman, & Parker, 1992; Kanai, Paffen, Hogendoorn, & Verstraten, 2006; Kaneko & Murakami, 2009). In addition, perceptual distortions of objects forms (compression) and speeds (overestimation) have been observed (Anstis & Atkinson, 1967; Aydin et al., 2009; Haber & Nathanson, 1968; Müsseler, 1999; Parks, 1965; Rock, 1981; von Helmholtz, 1867; Zöllner, 1862), and motivated the non-Euclidean and non-Galilean approaches in perception. Inspired by the theory of special relativity (Einstein, 1905), several attempts to create models of motion perception according to the Lorentzian transforms with a transform matrix in the form of,

$$\mathbf{T} = \gamma \begin{bmatrix} 1 & -v_{ref} \\ \frac{-v_{ref}}{c^2} & 1 \end{bmatrix}, \tag{7}$$

where $\gamma = 1 \Big/ \sqrt{1 - \frac{v_{ref}^2}{c^2}}$, have been made (Caelli, Hoffman, & Lindman, 1978; Drösler, 1979; Müsseler, 1999). In this formulation, time depends on speed. For small speeds ($v_{ref} << c$), the Galilean and Lorentzian approaches yield the same coordinate transformation matrix. On the other hand, as the speed approaches the speed of light, speed dependent space and time distortions occur. A critical difference between physics and perception is that, however, the constraint that led physicists to use a velocity dependent coordinate transform was the constant speed of light. On the other hand, a similar constraint for perception has not been formulated in a satisfactory manner. An analogous constraint as a *finite propagation rate of signals in the human visual system* has been proposed (Caelli, 1981). In contrast to the concept of constant speed of light in physics, this constraint (often referred to as c*) depends on many stimulus conditions and parameters (Drösler, 1979; Dzhafarov, 1992; Müsseler, 1999) which weakens the basis for using Lorentzian type of formulation in perception. Nevertheless, it has been shown to be successful in explaining some experimental data (Caelli et al., 1978; Müsseler, 1999).

It is important to note that the Eqns. 3, 4, and 7 assume that the coordinate transform is linear. However, this may not be necessarily the case. If the transform is nonlinear, then the Eqn. (2) can be rewritten as

$$c_{nr} = f(c_r), \tag{8}$$

where f(.) represents a nonlinear function.

7.1. Future Experiments

In this section, we will outline a series of experiments by which one can determine the transform matrix, **T**, given in Eqn. (3) or the transform function, f(.), given in Eqn. (8). The proposed experiments make use of a paradigm known as *slit viewing (or anorthoscopic perception),* which we reviewed in Chapter 1 and used as a tool in Chapter 2. Anorthoscopic perception occurs when an object moves behind a narrow slit. Although each fragment of the object falls on the same retinotopic strip over time, which creates retinotopic conflict of information about object's form, observers often see the object as a whole. This indicates that integration of each fragment into a whole, i.e., construction of the spatial layout of the object, takes place. In Chapter 2, we showed that this construction is done via the perceptual dimension of motion.

As an object moves behind a narrow slit, its features overlap both spatially and temporally. But the visual system can still construct the spatially extended form of the moving object and attribute features correctly. As we showed in Chapter 2, the attribution of features of a moving object is carried out based on object's perceived motion vector, which serves as a non-retinotopic reference frame, in slit viewing. Retinotopic processing and integration predicts that features of the moving object should blend into each other. According to our theory, this does not happen because features are processed on a moving reference frame rather than on a static retinotopic reference frame.

A direct testable prediction of this theory (non-retinotopic reference frames formed by perceived motion vectors) is that if the perceived motion is different for different parts of a moving object, then it should undergo deformations. Indeed, there is an agreement on that the moving objects in anorthoscopic viewing are often perceived compressed along the axis of motion (Anstis & Atkinson, 1967; Haber & Nathanson, 1968; McCloskey & Watkins, 1978; Morgan et al., 1982; Parks, 1965; Rock, 1981; Zöllner, 1862). In order to provide an account for the apparent compression, Aydin, Herzog, and Ogmen (Aydin et al., 2008) measured the perceived speeds of trailing and leading parts and the compression of an ellipse moving behind a slit. While the speeds of both parts are overestimated, trailing parts are perceived as moving faster than leading parts. According to the RFMF theory, each fragment of an object is registered in a non-retinotopic space with its corresponding motion vectors. This process works in real-time to build up the non-retinotopic representation of the moving object in anorthoscopic perception. The leading part of the object becomes visible first, and registered with its own velocity vector (Fig. 7.1). Then the middle and trailing parts follow and complete the non-retinotopic representation of the whole object. Note that local motion vectors gradually increase from the leading to the trailing edge. Fig. 7.1 shows how different perceived speeds of different parts of an object can explain perceived compressions in anorthoscopic perception.

In order to fully characterize the coordinate transform, one needs to measure the temporal as well as the spatial aspects of perception. In other words, in addition to the measurements of perceived speed and size as a function of physical speed (i.e., speed on a spatiotopic reference frame), one needs to measure also the perceived durations of events as a function of speed. This will allow us to understand how the visual system constructs representations of dynamic objects in real time.

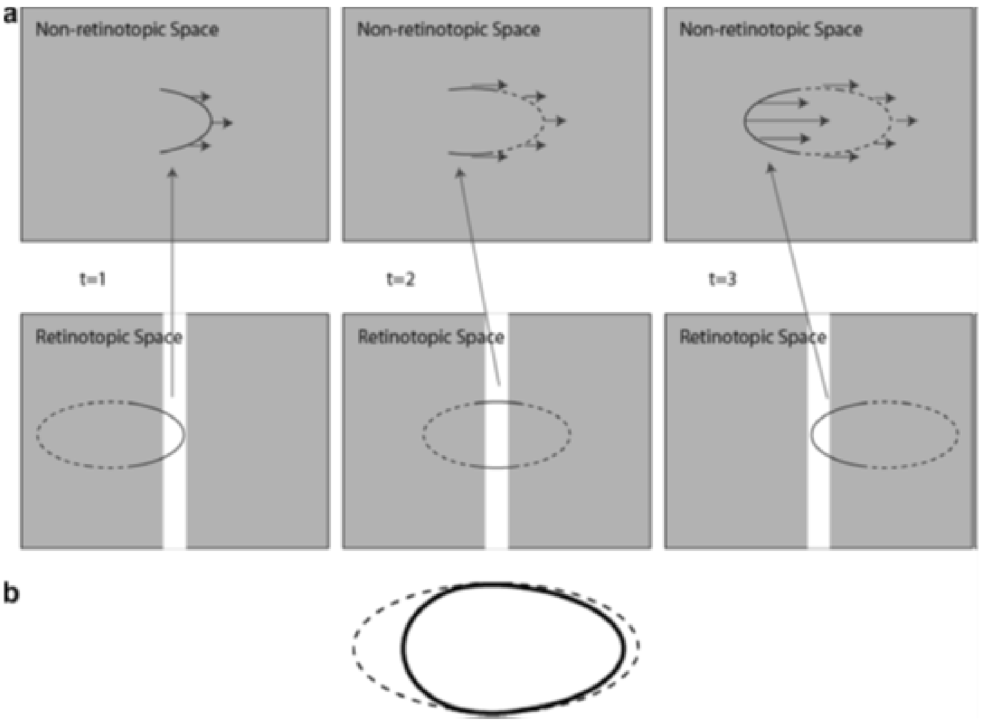

Figure 7.1. Apparent compressions in anorthoscopic perception can be explained by differential perceived speeds of different parts of the object according to the RFMF theory. Adapted from Aydin, Herzog, and Ogmen (2008).

Building upon Aydin, Herzog and Ogmen (2008)'s study, one can measure the perceived duration of different parts of an object through the slit. It has long been known that moving objects are perceived to have longer durations than stationary objects (Baro et al., 1992; Hughes et al., 1992; Kanai et al., 2006; Kaneko & Murakami, 2009; Tse, Intriligator, Rivest, & Cavanagh, 2004). Another common finding in these studies is that the faster the motion of the object, the longer the perceived duration. It has also been shown that perceived duration measurements are prone to vary with slight changes in the stimulus parameters such as size, luminance contrast, and color. Many methods and procedures have been used to measure perceived durations such as duration matching (Kaneko and

Murakami, 2009), reproduction (Kanei et al., 2006), comparison of onset and offset reaction times (Baro et al., 1992), method of single stimuli (comparison of current stimulus with all other stimuli presented in previous trials) (Tse et al., 2004), pairwise comparison (Hughes et al., 1992), and the method of constant stimuli (Tse et al., 2004).

Perceived durations of different parts of an object could be measured by using the method of constant stimuli. In order to avoid possible effects of attention, one can present a test object moving behind the slit and a stationary comparison object (e.g., a rectangular bar) separately. A certain part of the test diamond, whose perceived duration within the slit is to be measured, can be highlighted during its motion through the slit. The task of the observers can be to report whether the travel time of the highlighted part of the test object is longer than the perceived duration of the static comparison object. Percentage of responses in which observers reported longer durations for the test object (% longer) can be plotted against the physical durations of the comparison objects, and psychometric functions can be mapped for each physical speed of the test diamond. Durations which yield 50% longer responses can be obtained as Point of Subjective Equality (PSE).

Another prediction of the RFMF theory is that a “formless” motion stimulus should be able to generate reference frame fields. In other words, a complete perception of form is not required to obtain the field effect that we found in our experiments. In the future, this prediction can be tested by using random dot cinematograms (RDC) with minimal form structure and looking at whether and how they can influence the perceived motion of a another “formless” motion stimuli or a moving object with a complete and salient form structure.

7.2. Motion in Depth

All of our experiments contained motion in the fronto-parallel plane and hence, were restricted to two-dimensional space. However, our environment is three-dimensional. In fact, dynamic perception involves four dimensions; three-dimensional space and time. The reference fields we have reported here can also be extended to third and even to fourth dimension. For instance, if an object moves in depth (a common situation in normal viewing conditions), it might affect the perceived motion of nearby objects in a field-like manner. Induced motion in depth has been shown by several studies (Farné, 1970; Gogel & MacCracken, 1979; Gogel, 1974; Nefs & Harris, 2008). However, there has been conflicting results as to whether depth affects induced motion in the fronto-parallel plane or the reports of positive effect are due to methodological artifacts related to target velocity (DiVita & Rock, 1997; Gogel & MacCracken, 1979; Léveillé & Yazdanbakhsh, 2010). Further experimentation is needed to resolve whether reference fields extend in depth.

The span of a reference field might also extend in time. In other words, interactions between reference fields might depend on temporal variation in the reference field strength as well. Noory, Herzog, and Ogmen (2013) recently demonstrated the influence of temporal phase difference between two apparent motions on the interactions between reference fields generated by them.

7.3. Potential Neural Substrates

There are several studies supporting partial contributions of different exogenous (space-based, object-based, motion-based etc.) reference frames depending on the spatial structure and geometrical organization (Farrell-Whelan & Brooks, 2013; Magnussen et al., 2013; Shum & Wolford, 1983; Tadin et al., 2002), depth (Gogel & Koslow, 1972; Gogel, 1974), belongingness (DiVita & Rock, 1997), speed (Hisakata et al., 2013; Léveillé & Yazdanbakhsh, 2010; Mori, 1984), lighting conditions (Shum & Wolford, 1983), eccentricity (Thurman & Lu, 2013), and even interactions among different modalities (Avillac, Denève, Olivier, Pouget, & Duhamel, 2005). Furthermore, studies on perception during voluntary movement of eyes, head, or body indicate varying contributions of endogenous (retinocentric, headcentric, etc.) and exogenous reference frames (Agaoglu, Ogmen, & Herzog, 2012; Becklen et al., 1984; Brenner & van den Berg, 1994; Durgin, Gigone, & Scott, 2005; Hisakata et al., 2013; Johansson, 1976; Souman et al., 2006a; Turano & Heidenreich, 1999).

The reference fields mentioned here might be simply a result of lateral inhibition from stimulated motion detectors in MT/MST or an inevitable consequence of interaction between processing in the dorsal and the ventral streams (due to reciprocal projections between the two streams). It is well known that early visual areas V1, V2, V3, V4/V8, and V3a are retinotopically organized (Brewer, Press, Logothetis, & Wandell, 2002; Sereno et al., 1995; Tootell et al., 1982). However, mounting evidence shows that there are non-retinotopic representations in the brain, and that a multitude of representations based on various reference frames coexist. Supplementary Eye Field (SEF) of the monkey has been implicated for object-centered representations of targets selected for eye movements and attention (Batista & Newsome, 2000; Olson & Tremblay, 2000; Olson, 2003; Tremblay, Gettner, & Olson, 2002). EEG and fMRI studies on humans revealed overlapping neural circuitry in occipito-temporal and medial-superior posterior networks (Galati et al., 2000; Gramann, Müller, Schönebeck, & Debus, 2006; Zaehle et al., 2007). Studies on hemispatial neglect also demonstrated that in some patients with damaged right parietal cortex, the neglect is object-centered or object-aligned indicating object-based representations in these regions (Gainotti, Messerli, & Tissot, 1972; Halligan & Marshall, 1993; Walker & Findlay, 1996). In addition, the Lateral Occipital Complex (LOC) shows object selectivity and hence is considered as “object areas” (Grill-Spector, Kushnir, Hendler, & Malach, 2000; Kourtzi & Kanwisher, 2000; Malach et al., 1995). Moreover, LOC has direct projections to human motion area MT+ and V3A, an area implicated in dynamic form processing (Zeki, 1990), and that MT+ and V3A have reciprocal connections (Kim & Kim, 2005). Although speculative at this point, these reciprocal connections might be the source of reference fields generated around motion-based reference frames. Single-cell recordings or neuroimaging techniques over networks of neurons can be utilized to understand how reference-fields emerge within or across neural circuitry in the brain.

References

Agaoglu, M. N., Herzog, M. H., & Ogmen, H. (in press). Field-like interactions between motion-based reference frames. *Attention, Perception & Psychophysics.*

Agaoglu, M. N., Herzog, M. H., & Ogmen, H. (2015). Motion-based nearest-vector metric for reference-frame selection in the perception of motion. Manuscript submitted for publication.

Agaoglu, M. N., Herzog, M. H., & Ogmen, H. (2012). Non-retinotopic feature processing in the absence of retinotopic spatial layout and the construction of perceptual space from motion. *Vision Research*, *71*, 10–7.

Agaoglu, M. N., Herzog, M., & Öğmen, H. (2015). The effective reference frame in perceptual judgments of motion direction. *Vision Research*, *107*, 101–112.

Anstis, S. M., & Atkinson, J. (1967). Distortions in moving figures viewed through a stationary slit. *The American Journal of Psychology*, *80*(4), 572–85.

Anstis, S. M., & Casco, C. (2006). Induced movement: the flying bluebottle illusion. *Journal of Vision*, *6*(10), 1087–92.

Arnoldussen, D. M., Goossens, J., & van den Berg, A. V. (2011). Adjacent visual representations of self-motion in different reference frames. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 11668–11673.

Aubert, H. (1886). Die Bewegungsempfindung. *Pflügers Archiv*, *39*, 347–370.

Avillac, M., Denève, S., Olivier, E., Pouget, A., & Duhamel, J.-R. (2005). Reference frames for representing visual and tactile locations in parietal cortex. *Nature Neuroscience*, *8*(7), 941–9.

Aydin, M., Herzog, M. H., & Ogmen, H. (2008). Perceived speed differences explain apparent compression in slit viewing. *Vision Research*, *48*(15), 1603–12.

Aydin, M., Herzog, M. H., & Ogmen, H. (2009). Shape distortions and Gestalt grouping in anorthoscopic perception. *Journal of Vision*, *9*, 1–8.

Aydin, M., Herzog, M. H., & Ogmen, H. (2011). Attention modulates spatio-temporal grouping. *Vision Research*, *51*(4), 435–46.

Bachmann, T. (1994). *Psychophysiology of visual masking: The fine structure of conscious experience*. Commack, NY: Nova Science.

Baker, C. J., & Braddick, O. (1982). Does segregation of differently moving areas depend on relative or absolute displacement? *Vision Research*, *22*(7), 851–856.

Bardi, L., Regolin, L., & Simion, F. (2011). Biological motion preference in humans at birth: role of dynamic and configural properties. *Developmental Science*, *14*(2), 353–359.

Baro, J., Brzezicki, L., Lehmkuhle, S., & Hughes, H. (1992). The perceived duration of gratings. *Perception*, *21*(2), 161–166.

Batista, A. P., & Newsome, W. T. (2000). Visuo-motor control: giving the brain a hand. *Current Biology : CB*, *10*(4), R145–8.

Becklen, R., Wallach, H., & Nitzberg, D. (1984). A limitation of position constancy. *Journal of Experimental Psychology. Human Perception and Performance*, *10*(5), 713–23.

Bertamini, M., & Proffitt, D. R. (2000). Hierarchical motion organization in random dot configurations. *Journal of Experimental Psychology. Human Perception and Performance*, *26*(4), 1371–86.

Bex, P. J., Edgar, G. K., & Smith, A. T. (1995). Sharpening of drifting, blurred images. *Vision Research*, *35*(18), 2539–2546.

Bhatia, B., & Verghese, C. (1964). Threshold size of moving object as function of its speed. *Journal of the Optical Society of America*, *54*(7), 948–&.

Bischof, N., & Kramer, E. (1968). Investigations and considerations of directional perception during voluntary saccadic eye movements. *Psychologische Forschung*, *32*(3), 185–218.

Boi, M., Ogmen, H., & Herzog, M. H. (2011). Motion and tilt aftereffects occur largely in retinal, not in object, coordinates in the Ternus–Pikler display. *Journal of Vision*, *11*, 1–11.

Boi, M., Ogmen, H., Krummenacher, J., Otto, T. U., & Herzog, M. H. (2009). A (fascinating) litmus test for human retino-vs. non-retinotopic processing. *Journal of Vision*, *9*, 1–11.

Boi, M., Vergeer, M., Ogmen, H., & Herzog, M. H. (2011). Nonretinotopic exogenous attention. *Current Biology : CB*, *21*(20), 1732–7.

Borjesson, E., & von Hofsten, C. (1972). Spatial determinants of depth perception in two-dot motion patterns. *Perception & Psychophysics*, *11*(4), 263–268.

Borjesson, E., & von Hofsten, C. (1975). A Vector Model for Perceived Object Rotation and Translation in Space. *Psychological Research*, (38), 209–230.

Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: a comparison, review, and update. *Perception & Psychophysics*, *62*(8), 1572–95.

Bremner, L. R., & Andersen, R. A. (2012). Coding of the reach vector in parietal area 5d. *Neuron*, *75*(2), 342–51.

Bremner, L. R., & Andersen, R. A. (2014). Temporal Analysis of Reference Frames in Parietal Cortex Area 5d during Reach Planning. *Journal of Neuroscience*, *34*(15), 5273–5284.

Brenner, E., & van den Berg, A. V. (1994). Judging object velocity during smooth pursuit eye movements. *Experimental Brain Research*, *99*(2), 316–24.

Brewer, A. A., Press, W. A., Logothetis, N. K., & Wandell, B. A. (2002). Visual areas in macaque cortex measured using functional magnetic resonance imaging. *The Journal of Neuroscience*, *22*(23), 10416–26.

Bridgeman, B., Van der Heijden, A. H. C., & Velichkovsky, B. M. (1994). A theory of visual stability across saccadic eye movements. *Behavioral and Brain Sciences*, *17*(2), 247–293.

Brown, J. S. (1931). The visual perception of velocity. *Psychologische Forschung*, *14*(1), 247–258.

Burr, D. (1980). Motion smear. *Nature*, *284*(5752), 164–165.

Burr, D., & Morgan, M. (1997). Motion deblurring in human vision. *Proceedings of the Royal Society B: Biological Sciences*, *264*(1380), 431–436.

Burr, D., Ross, J., & Morrone, M. C. (1986). Seeing objects in motion. *Proceedings of the Royal Society B: Biological Sciences*, *227*(1247), 249–265.

Caelli, T. (1981). *Visual perception, theory and practice*. Oxford: Pergamon.

Caelli, T., Hoffman, W., & Lindman, H. (1978). Subjective Lorentz transformations and the perception of motion. *Journal of the Optical Society of America*, *68*(3), 402–11.

Cavanagh, P., Hunt, A. R., Afraz, A., & Rolfs, M. (2010). Visual stability based on remapping of attention pointers. *Trends in Cognitive Sciences*, *14*(4), 147–53.

Chen, S., Bedell, H. E., & Ogmen, H. (1995). A target in real motion appears blurred in the absence of other proximal moving targets. *Vision Research*, *35*(16), 2315–2328.

Churan, J., Guitton, D., & Pack, C. C. (2011). Context dependence of receptive field remapping in superior colliculus. *Journal of Neurophysiology*, *106*(4), 1862–74.

Churan, J., Guitton, D., & Pack, C. C. (2012). Perisaccadic remapping and rescaling of visual responses in macaque superior colliculus. *PloS One*, *7*(12), e52195.

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, *22*, 319–349.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics*, *27*(3), 183–228.

Crawford, J. D., Henriques, D. Y. P., & Medendorp, W. P. (2011). Three-dimensional transformations for goal-directed action. *Annual Review of Neuroscience*, *34,* 309–31.

Currie, C. B., McConkie, G. W., Carlson-Radvansky, L. A., & Irwin, D. E. (2000). The role of the saccade target object in the perception of a visually stable world. *Perception & Psychophysics, 62*(4), 673–83.

Cutting, J. E., & Proffitt, D. R. (1982). The minimum principle and the perception of absolute, common, and relative motions. *Cognitive Psychology*, *14*(2), 211–46.

Day, R. (1978). Induced visual movement as nonveridical resolution of displacement ambiguity. *Perception & Psychophysics*, *23*(3), 205–209.

Deubel, H. (2004). *Localization of targets across saccades: Role of landmark objects*. *Visual Cognition* (Vol. 11, pp. 173–202).

Di Lollo, V., & Hogben, J. (1985). Suppression of visible persistence. *Journal of Experimental Psychology: Human Perception and Performance*, *11,* 304–316.

DiVita, J. C., & Rock, I. (1997). A belongingness principle of motion perception. *Journal of Experimental Psychology. Human Perception and Performance*, *23*(5), 1343–52.

Dixon, N., & Hammond, E. (1972). Attenuation of visual persistence. *British Journal of Psychology*, *63*(May), 243–254.

Dodge, R. (1904). The participation of the eye movements in the visual perception of motion. *Psychological Review*, *11*(1), 1–14.

Drösler, J. (1979). Relativistic effects in visual perception of real and apparent motion. *Archiv Für Psychologie*, *131*(3), 249.

Duffy, C. J., & Wurtz, R. (1995). Response of Monkey MST Neurons to Optic Flow Stimuli with Shifted Centers of Motion. *The Journal of Neuroscience, 15*(July), 5192–5208.

Duffy, C. J., & Wurtz, R. H. (1991a). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology*, *65*(6), 1329–45.

Duffy, C. J., & Wurtz, R. H. (1991b). Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli. *Journal of Neurophysiology*, *65*(6), 1346–59.

Duhamel, J. R., Bremmer, F., Ben Hamed, S., & Graf, W. (1997). Spatial invariance of visual receptive fields in parietal cortex neurons. *Nature, 389*(6653), 845–8.

Duncker, K. (1929). Über induzierte Bewegung. *Psychologische Forschung*, *12*(1), 180–259.

Durgin, F. H., Gigone, K., & Scott, R. (2005). Perception of Visual Speed While Moving. *Journal of Experimental Psychology. Human Perception and Performance, 31*(2), 339–353.

Dzhafarov, E. N. (1992). Visual Kinematics I . Visual Space Metric in Visual Motion. *Journal of Mathematical Psychology*, (32), 471–497.

Eifuku, S., & Wurtz, R. (1998). Response to Motion in Extrastriate Area MSTl : Center-Surround Interactions. *Journal of Neurophysiology*, 282–296.

Einstein, A. (1905). On the electrodynamics of moving bodies. *Annalen Der Physik*, *17*(891), 50.

Ellis, W. D. (1938). *A Source Book of Gestalt Psychology*. London.

Farné, M. (1970). Induced motion in three dimensions. *Perceptual and Motor Skills, 30*(2), 426.

Farrel, J. (1984). Visible persistence of moving objects. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 502–511.

Farrell-Whelan, M., & Brooks, K. R. (2013). Differential processing: towards a unified model of direction and speed perception. *Vision Research*, *92*, 10–8.

Fendrich, R., Rieger, J., & Heinze, H. (2005). The effect of retinal stabilization on anorthoscopic percepts under free-viewing conditions. *Vision Research*, *45*, 567–582.

Festinger, L., Sedgwick, H. A., & Holtzman, J. D. (1976). Visual perception during smooth pursuit eye movements. *Vision Research*, *16*(12), 1377–1386.

Fields, C. (2011). Trajectory recognition as the basis for object individuation: a functional model of object file instantiation and object-token encoding. *Frontiers in Psychology*, *2*(March), 49.

Filehne, W. (1922). Uber das optische Wahrnehmen von Bewegungen. *Zeitschrift Für Sinnephysiologie*, *53*.

Fischer, J., Spotswood, N., & Whitney, D. (2011). The emergence of perceived position in the visual system. *Journal of Cognitive Neuroscience*, *23*, 119–136.

Fleischl, E. V. (1882). Physiologisch-optische Notizen, 2. Mitteilung. *Sitzung Wiener Bereich Der Akademie Der Wissenschaften*, *3*, 7–25.

Freeman, T. C. A. (2001). Transducer models of head-centred motion perception. *Vision Research*, *41*(21), 2741–55.

Freeman, T. C. A., & Banks, M. S. (1998). Perceived head-centric speed is affected by both extra-retinal and retinal errors. *Vision Research*, *38*(7), 941–945.

Freeman, T. C. A., Champion, R. A., Sumnall, J. H., & Snowden, R. J. (2009). Do we have direct access to retinal image motion during smooth pursuit eye movements? *Journal of Vision*, *9*(33), 1–11.

Freeman, T. C. A., Champion, R. A., & Warren, P. A. (2010). A Bayesian model of perceived head-centered velocity during smooth pursuit eye movement. *Current Biology : CB*, *20*(8), 757–62.

Furman, M., & Gur, M. (2003). Self-organizing neural network model of motion processing in the visual cortex during smooth pursuit. *Vision Research*, *43*(20), 2155–2171.

Furman, M., & Gur, M. (2005). Alteration of the perceived path of a non-pursued target during smooth pursuit: analysis by a neural network model. *Vision Research*, *45*(13), 1755–68.

Gainotti, G., Messerli, P., & Tissot, R. (1972). Qualitative analysis of unilateral spatial neglect in relation to laterality of cerebral lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, *35*(4), 545–50.

Galati, G., Lobel, E., Vallar, G., Berthoz, A., Pizzamiglio, L., & Le Bihan, D. (2000). The neural basis of egocentric and allocentric coding of space in humans: a functional magnetic resonance study. *Experimental Brain Research*, *133*(2), 156–64.

Galati, G., Pelle, G., Berthoz, A., & Committeri, G. (2010). Multiple reference frames used by the human brain for spatial perception and memory. *Experimental Brain Research*, *206*(2), 109–120.

Galletti, C., Battaglini, P. P., & Fattori, P. (1995). Eye position influence on the parieto-occipital area PO (V6) of the macaque monkey. *The European Journal of Neuroscience*, *7*(12), 2486–501.

Gershman, S., Jäkel, F., & Tenenbaum, J. (2013). Bayesian Vector Analysis and the Perception of Hierarchical Motion. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (Vol. 1).

Gibson, J. J. (1979). *The Ecological Approach To Visual Perception*. Boston: Houghton Mifflin.

Gogel, W. C. (1974). Relative motion and the adjacency principle. *Quarterly Journal of Experimental Psychology*, *26*(3), 425–437.

Gogel, W. C. (1977). The metric of visual space. In W. Epstein (Ed.), *Stability and constancy in visual perception, mechanisms, and processes* (pp. 129–181). New York: John Wiley.

Gogel, W. C., & Koslow, M. (1972). The adjacency principle and induced movement. *Perception & Psychophysics*, *11*(4), 309–314.

Gogel, W. C., & MacCracken, P. (1979). Depth adjacency and induced motion. *Perceptual and Motor Skills*, *48*(2), 343–50.

Gogel, W. C., & Sharkey, T. (1989). Measuring attention using induced motion. *Perception*, *18*(3), 303–20.

Gramann, K., Müller, H. J., Schönebeck, B., & Debus, G. (2006). The neural basis of ego- and allocentric reference frames in spatial navigation: evidence from spatio-temporal coupled current density reconstruction. *Brain Research*, *1118*(1), 116–29.

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience*, *3*(8), 837–43.

Grossberg, S., Léveillé, J., & Versace, M. (2011). How do object reference frames and motion vector decomposition emerge in laminar cortical circuits? *Attention, Perception & Psychophysics*, *73*(4), 1147–70.

Haber, R., & Nathanson, L. (1968). Post-retinal storage. Some further observations on parks camel as seen through eye of a needle. *Perception & Psychophysics*, *3*(5A), 349–355.

Haber, R., & Standing, L. (1970). Direct estimates of the apparent duration of a flash. *Canadian Journal of Psychology*, *24*(4), 216–229.

Halligan, P. W., & Marshall, J. C. (1993). When two is one: a case study of spatial parsing in visual neglect. *Perception*, *22*(3), 309–12.

Hamker, F. H., Zirnsak, M., Calow, D., & Lappe, M. (2008). The peri-saccadic perception of objects and space. *PLoS Computational Biology*, *4*(2), e31.

Hammett, S. T. (1997). Motion blur and motion sharpening in the human visual system. *Vision Research*, *37*(18), 2505–2510.

Herzog, M. H., & Ogmen, H. (2014). Apparent motion and reference frames. In J. Wagemans (Ed.), *Oxford Handbook of Perceptual Organization*. Oxford University Press.

Hisakata, R., Terao, M., & Murakami, I. (2013). Illusory position shift induced by motion within a moving envelope during smooth-pursuit eye movements. *Journal of Vision*, *13*(21), 1–12.

Hochberg, J. E., & Fallon, P. (1976). Perceptual analysis of moving patterns. *Science*, *194*(4269), 1081–1083.

Hochberg, J. E., & McAlister, E. (1953). A quantitative approach, to figural“ goodness”. *Journal of Experimental Psychology*, *46*(5), 361.

Hughes, J., Lishman, J., & Parker, D. (1992). Apparent duration and spatial structure. *Perception & Psychophysics*, *52*(2), 222–230.

Inaba, N., Shinomoto, S., Yamane, S., Takemura, A., & Kawano, K. (2007). MST Neurons Code for Visual Motion in Space Independent of Pursuit Eye Movements. *Journal of Neurophysiology*, *97*, 3473–3483.

Jansson, G., Bergstrom, S., & Epstein, W. (1994). *Perceiving events and objects*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Johansson, G. (1950). *Configurations in event perception : an experimental study*. Stockholm: Almqvist & Wiksell.

Johansson, G. (1958). Rigidity, stability, and motion in perceptual space. *Acta Psychologica*, *14*, 359–370.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*(2), 201–211.

Johansson, G. (1974). Vector analysis in visual perception of rolling motion. *Psychologische Forschung*, *36*(4), 311–319.

Johansson, G. (1975). Visual motion perception. *Scientific American*, *232*(6), 76–88.

Johansson, G. (1976). Spatio-temporal differentiation and integration in visual motion perception. *Psychological Research*, *38*(4), 379–393.

Johansson, G. (1985). Vector analysis and process combinations in motion perception: A reply to Wallach, Becklen, and Nitzberg (1985). *Journal of Experimental Psychology: Human Perception and Performance*, *11*(3), 367–371.

Johansson, G. (1986). Relational invariance and visual space perception: On perceptual vector analysis of the optic flow. *Acta Psychologica*, *63*(2), 89–101.

Johansson, G., & Jansson, G. (1968). Perceived rotary motion from changes in a straight line. *Perception & Psychophysics*, *4*(3), 165–170.

Johansson, G., von Hofsten, C., & Jansson, G. (1980). Event Perception. *Annual Review of Psychology*, (31), 27–63.

Joiner, W. M., Cavanaugh, J., & Wurtz, R. H. (2011). Modulation of shifting receptive field activity in frontal eye field by visual salience. *Journal of Neurophysiology*, *106*(3), 1179–90.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cognitive Psychology*, *24*(2), 175–219.

Kanai, R., Paffen, C., Hogendoorn, H., & Verstraten, F. (2006). Time dilation in dynamic visual display. *Journal of Vision*, *6*(12), 1421–1430.

Kaneko, S., & Murakami, I. (2009). Perceived duration of visual motion increases with speed. *Journal of Vision*, *9*(7), 1–12.

Kano, C., & Hayashi, K. (1981). The apparent path of a stationary and a circularly moving spot during the smooth pursuit of another circularly moving spot. *Acta Psychologica*, *48*(1-3), 151–60.

Kawabe, T. (2008). Spatiotemporal feature attribution for the perception of visual size. *Journal of Vision*, *8*(8), 7.1–9.

Kim, D.-S., & Kim, M. (2005). Combining functional and diffusion tensor MRI. *Annals of the New York Academy of Sciences*, *1064*, 1–15.

Koffka, K. (1935). *Principle of Gestalt Psychology*. New york: Harcourt.

Kourtzi, Z., & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *20*(9), 3310–8.

Léveillé, J., Myers, E., & Yazdanbakhsh, A. (2014). Object-centered reference frames in depth as revealed by induced motion. *Journal of Vision*, *14*(3), 1–11.

Léveillé, J., & Yazdanbakhsh, A. (2010). Speed, more than depth, determines the strength of induced motion. *Journal of Vision*, *10*, 1–9.

Löffler, G., & Orbach, H. S. (1999). Computing feature motion without feature detectors: a model for terminator motion without end-stopped cells. *Vision Research*, *39*(4), 859–71.

Löffler, G., & Orbach, H. S. (2001). Anisotropy in judging the absolute direction of motion. *Vision Research*, *41*(27), 3677–92.

Mack, A., Fisher, C., & Fendrich, R. (1975). A reexamination of two-point induced movement. *Perception & Psychophysics*, *17*(3), 273–276.

Mack, A., & Herman, E. (1972). A new illusion: The underestimation of distance during pursuit eye movements. *Perception & Psychophysics*, *12*(6), 471–473.

Mack, A., & Herman, E. (1973). Position constancy during pursuit eye movement: an investigation of the Filehne illusion. *The Quarterly Journal of Experimental Psychology*, *25*(1), 71–84.

Mack, A., & Herman, E. (1978). The loss of position constancy during pursuit eye movements. *Vision Research*, *18*(1), 55–62.

Magnussen, C. M., Orbach, H. S., & Loffler, G. (2013). Motion trajectories and object properties influence perceived direction of motion. *Vision Research*, *91*, 21–35.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., … Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(18), 8135–9.

Malkinson, T., McKyton, A., & Zohary, E. (2012). Motion adaptation reveals that the motion vector is represented in multiple coordinate frames. *Journal of Vision*, *12*, 1–11.

Marinovic, W., & Arnold, D. H. (2013). An illusory distortion of moving form driven by motion deblurring. *Vision Research*, *88*, 47–54.

Marr, D., & Ullman, S. (1981). Directional Selectivity and its Use in Early Visual Processing. *Proceedings of the Royal Society B: Biological Sciences*, *211*(1183), 151–180.

Maruya, K., Holcombe, A., & Nishida, S. (2013). Rapid encoding of relationships between spatially remote motion signals. *Journal of Vision*, *13*, 1–20.

Mateeff, S., & Hohnsbein, J. (1989). The role of the adjacency between background cues and objects in visual localization during ocular pursuit. *Perception*, *18*(1), 93–104.

Mateeff, S., Hohnsbein, J., & Ehrenstein, W. H. (1990). Visual localization and estimation of extent of target motion during ocular pursuit: a common mechanism? *Perception*, *19*(4), 459–69.

Maus, G. W., Fischer, J., & Whitney, D. (2013). Motion-dependent representation of space in area MT+. *Neuron*, *78*(3), 554–62.

McCloskey, M., & Watkins, M. (1978). Seeing-more-than-is-there phenomenon - implications for locus of iconic storage. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(4), 553–564.

Melcher, D., & Colby, C. L. (2008). Trans-saccadic perception. *Trends in Cognitive Sciences*, *12*(12), 466–473.

Melcher, D., & Morrone, M. C. (2003). Spatiotopic temporal integration of visual motion across saccadic eye movements. *Nature Neuroscience*, *6*(8), 877–81.

Michotte, A., Thinès, G., & Crabbé, G. (1964). *Les complements amodaux des structures perceptives*. Louvain: Studia Psychologica, Publications Universitaires de Louvain.

Moorman, D. E., & Olson, C. R. (2007). Combination of neuronal signals representing object-centered location and saccade direction in macaque supplementary eye field. *Journal of Neurophysiology*, *97*(5), 3554–66.

Morgan, M., Findlay, J., & Watt, R. (1982). Aperture viewing: A review and a synthesis. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *34*(2), 211–233.

Mori, T. (1979). Relative locations among moving spots and visual vector analysis. *Perceptual and Motor Skills*, *48*(2), 587–92.

Mori, T. (1984). Change of a frame of reference with velocity in visual motion perception. *Perception & Psychophysics*, *35*(6), 515–8.

Morrone, M. C., Tosetti, M., Montanaro, D., Fiorentini, A., Cioni, G., & Burr, D. (2000). A cortical area that responds specifically to optic flow, revealed by fMRI. *Nature Neuroscience*, *3*(12), 1322–8.

Müsseler, J. (1999). Perceiving and measuring of spatiotemporal events. In *Modeling consciousness across the disciplines* (pp. 95–112).

Nefs, H. T., & Harris, J. M. (2008). Induced motion in depth and the effects of vergence eye movements. *Journal of Vision*, *8*(3), 8.1–16.

Neggers, S. F. W., Van der Lubbe, R. H. J., Ramsey, N. F., & Postma, A. (2006). Interactions between ego- and allocentric neuronal representations of space. *NeuroImage*, *31*(1), 320–31.

Nishida, S. (2004). Motion-based analysis of spatial patterns by the human visual system. *Current Biology*, *14*(10), 830–839.

Nishida, S. (2011). Advancement of motion psychophysics: Review 2001–2010. *Journal of Vision*, *11*, 1–53.

Nishida, S., Watanabe, J., Kuriki, I., & Tokimoto, T. (2007). Human visual system integrates color signals along a motion trajectory. *Current Biology*, *17*(4), 366–72.

Noory, B., Herzog, M. H., & Ogmen, H. (n.d.). *The non-retinotopic reference-frame field effect in human vision*.

Noory, B., Herzog, M. H., & Ogmen, H. (2013). Temporal characteristics of non-retinotopic reference frames in human vision. *Journal of Vision*, *13*(15), T13–T13.

Ogmen, H. (2007). A theory of moving form perception: Synergy between masking, perceptual grouping, and motion computation in retinotopic and non-retinotopic representations. *Advances in Cognitive Psychology*, *3*(1-2), 67–84.

Ogmen, H., & Herzog, M. H. (2010). The Geometry of Visual Perception: Retinotopic and Non-retinotopic Representations in the Human Visual System. *Proceedings of the IEEE. Institute of Electrical and Electronics Engineers*, *98*(3), 479–492.

Ogmen, H., Herzog, M. H., & Noory, B. (2013a). Non-Retinotopic Perception: Predictions and Empirical Tests of a Reference-Frame Metric Field Theory. *Journal of Vision*, *13*(9), 681–681.

Ogmen, H., Herzog, M. H., & Noory, B. (2013b). Reference-Frame Metric Field (RFMF) Theory of Non-retinotopic Visual Perception. *Perception*, *ECVP Abstr*(42), 160.

Ogmen, H., Otto, T. U., & Herzog, M. H. (2006). Perceptual grouping induces non-retinotopic feature attribution in human vision. *Vision Research*, *46*(19), 3234–42.

Olson, C. R. (2003). Brain representation of object-centered space in monkeys and humans. *Annual Review of Neuroscience*, *26*, 331–54.

Olson, C. R., & Tremblay, L. (2000). Macaque supplementary eye field neurons encode object-centered locations relative to both continuous and discontinuous objects. *Journal of Neurophysiology*, *83*(4), 2392–411.

Otto, T. U., Ogmen, H., & Herzog, M. H. (2006). The flight path of the phoenix--the visible trace of invisible elements in human vision. *Journal of Vision*, *6*(10), 1079–86.

Otto, T. U., Ogmen, H., & Herzog, M. H. (2008). Assessing the microstructure of motion correspondences with non-retinotopic feature attribution. *Journal of Vision*, *8*, 1–15.

Otto, T. U., Ogmen, H., & Herzog, M. H. (2009). Feature integration across space, time, and orientation. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(6), 1670–86.

Otto, T. U., Ogmen, H., & Herzog, M. H. (2010a). Attention and non-retinotopic feature integration. *Journal of Vision*, *10*, 1–13.

Otto, T. U., Ogmen, H., & Herzog, M. H. (2010b). Perceptual learning in a nonretinotopic frame of reference. *Psychological Science*, *21*(8), 1058–63.

Pack, C., Grossberg, S., & Mingolla, E. (2001). A neural model of smooth pursuit control and motion perception by cortical area MST. *Journal of Cognitive Neuroscience*, *13*(1), 102–20.

Pantle, A., & Picciano, L. (1976). A multistable movement display: evidence for two separate motion systems in human vision. *Science (New York, N.Y.)*, *193*(4252), 500–2.

Parks, T. (1965). Post-retinal visual storage. *The American Journal of Psychology*, *78*(1), 145–147.

Pavlova, M., Birbaumer, N., & Sokolov, A. (2006). Attentional modulation of cortical neuromagnetic gamma response to biological movement. *Cerebral Cortex (New York, N.Y. : 1991)*, *16*(3), 321–7.

Pavlova, M., & Sokolov, A. (2000). Orientation specificity in biological motion perception. *Perception & Psychophysics*, *62*(5), 889–899.

Piaget, J., & Inhelder, B. (1969). *The Psychology Of The Child* (p. 192). Basic Books.

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

Pola, J., & Wyatt, H. J. (1989). The perception of target motion during smooth pursuit eye movements in the open-loop condition: Characteristics of retinal and extraretinal signals. *Vision Research*, *29*(4), 471–483.

Poljac, E., Verfaillie, K., & Wagemans, J. (2011). Integrating biological motion: the role of grouping in the perception of point-light actions. *PloS One*, *6*(10), e25867.

Proffitt, D. R., & Cutting, J. E. (1980). Perceiving the centroid of curvilinearly bounded rolling shapes. *Perception & Psychophysics*, *28*(5), 484–487.

Proffitt, D. R., Cutting, J. E., & Stier, D. M. (1979). Perception of Wheel-Generated Motions. *Journal of Experimental Psychology. Human Perception and Performance*, *5*(2), 289–302.

Purushothaman, G., Patel, S. S., Bedell, H. E., & Ogmen, H. (1998). Moving ahead through differential visual latency. *Nature*, *396*(6710), 424.

Ramachandran, V. S., Rao, V. M., & Vidyasagar, T. R. (1974). Sharpness constancy during movement perception (short note). *Perception*, *3*(1), 97–8.

Restle, F. (1979). Coding theory of the perception of motion configurations. *Psychological Review*, *86*(1), 1–24.

Rieger, J., & Grüschow, M. (2007). The appearance of figures seen through a narrow aperture under free viewing conditions: Effects of spontaneous eye motions. *Journal of Vision*, *7*(6), 10.

Roberts, V. (1957). The Solar and Lunar Theory of Ibn ash-Shatir: A Pre-Copernican Copernican Model. *Isis*, 428–432.

Rock, I. (1981). Anorthoscopic perception. *Scientific American*, *244*(3), 145–153.

Rock, I., Auster, M., Schiffman, M., & Wheeler, D. (1980). Induced movement based on subtraction of motion from the inducing object. *Journal of Experimental Psychology-Human Perception and Performance*, *6*(3), 391–403.

Rock, I., & Sigman, E. (1973). Intelligence factors in the perception of form through a moving slit. *Perception*, *2*(3), 357–369.

Rogers, B., & Graham, M. (1979). Motion parallax as an independent cue for depth-perception. *Perception*, *8*(2), 125–134.

Rubin, J., & Richards, W. A. (1988). Visual perception of moving parts. *Journal of the Optical Society of America A*, *5*(12), 2045.

Rushton, S. K., Bradshaw, M. F., & Warren, P. A. (2007). The pop out of scene-relative object movement against retinal motion due to self-movement. *Cognition*, *105*(1), 237–45.

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), 1421–31.

Scharnowski, F., Hermens, F., & Herzog, M. (2007). Bloch's law and the dynamics of feature fusion. *Vision Research*, *47*(18), 2444–2452.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R., Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*(5212), 889–93.

Shimojo, S., & Richards, W. (1986). "Seeing" shapes that are almost totally occluded: a new look at Parks's camel. *Perception & Psychophysics*, *39*(6), 418–26.

Shimozaki, S. S., Eckstein, M., & Thomas, J. P. (1999). The maintenance of apparent luminance of an object. *Journal of Experimental Psychology. Human Perception and Performance*, *25*(5), 1433–53.

Shum, K., & Wolford, G. (1983). A quantitative study of perceptual vector analysis. *Perception & Psychophysics*, *34*(1), 17–24.

Snowden, R. J. (1992). Sensitivity to relative and absolute motion. *Perception, 21*(5), 563–8.

Sohmiya, T., & Sohmiya, K. (1992). Where does an anorthoscopic image appear. *Perceptual and Motor Skills*, *75*(3), 707–714.

Sohmiya, T., & Sohmiya, K. (1994). What is a crucial determinant in anorthoscopic perception. *Perceptual and Motor Skills*, (78), 987–998.

Sokolov, A., & Pavlova, M. (2006). Visual motion detection in hierarchical spatial frames of reference. *Experimental Brain Research*, *174*(3), 477–86.

Sommer, M. A., & Wurtz, R. H. (2004). What the brain stem tells the frontal cortex. II. Role of the SC-MD-FEF pathway in corollary discharge. *Journal of Neurophysiology*, *91*(3), 1403–23.

Souman, J. L., Hooge, I. T. C., & Wertheim, A. (2005). Vertical object motion during horizontal ocular pursuit: compensation for eye movements increases with presentation duration. *Vision Research*, *45*, 845–853.

Souman, J. L., Hooge, I. T. C., & Wertheim, A. H. (2006a). Frame of reference transformations in motion perception during smooth pursuit eye movements. *Journal of Computational Neuroscience*, *20*, 61–76.

Souman, J. L., Hooge, I. T. C., & Wertheim, A. H. (2006b). Localization and motion perception during smooth pursuit eye movements. *Experimental Brain Research*, *171*(4), 448–58.

Sperry, R. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *J Comp Physiol Psychol*, *43*(6), 482–489.

Stoper, A. E. (1973). Apparent motion of stimuli presented stroboscopically during pursuit movement of the eye. *Perception & Psychophysics*, *13*(2), 201–211.

Swanston, M., Wade, N., & Day, R. (1987). The representation of uniform motion in vision. *Perception*, *16*(2), 143–159.

Tadin, D., Lappin, J. S., Blake, R., & Grossman, E. D. (2002). What constitutes an efficient reference frame for vision? *Nature Neuroscience*, *5*, 1010–1015.

Takemura, H., Ashida, H., Amano, K., Kitaoka, A., & Murakami, I. (2012). Neural correlates of induced motion perception in the human brain. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(41), 14344–54.

Ternus, J. (1926). Experimentelle Untersuchung über phänomenale Identität. *Psychologische Forschung*, *7*, 81–136.

Thurman, S., & Lu, H. (2013). Complex interactions between spatial, orientation, and motion cues for biological motion perception across visual space. *Journal of Vision*, *13*, 1–18.

Tootell, R. B., Silverman, M., Switkes, E., & Valois, R. De. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science*, *218*(4575), 902–904.

Tremblay, L., Gettner, S. N., & Olson, C. R. (2002). Neurons with object-centered spatial selectivity in macaque SEF: do they represent locations or rules? *Journal of Neurophysiology*, *87*(1), 333–50.

Tse, P. U., Intriligator, J., Rivest, J., & Cavanagh, P. (2004). Attention and the subjective expansion of time. *Perception & Psychophysics*, *66*(7), 1171–89.

Turano, K. A., & Heidenreich, S. M. (1996). Speed discrimination of distal stimuli during smooth pursuit eye motion. *Vision Research*, *36*(21), 3507–17.

Turano, K. A., & Heidenreich, S. M. (1999). Eye movements affect the perceived speed of visual motion. *Vision Research*, *39*(6), 1177–87.

Turano, K. A., & Massof, R. W. (2001). Nonlinear contribution of eye velocity to motion perception. *Vision Research*, *41*(3), 385–95.

Veselovsky, I. N. (1973). Copernicus and Nasir al-Din al-Tusi. *Journal for the History of Astronomy*, *4*, 128.

Von Helmholtz, H. (1867). *Handbook of Phsyiological Optics* (1963rd ed.). New York: Dover Reprint.

Von Holst, E. (1950). Das reafferenzprinzip. *Naturwissenschaften*, *37*(20), 464–476.

Von Holst, E., & Mittelstaedt, H. (1971). The principle of reafference: Interactions between the central nervous system and the peripheral organs. In *Perceptual processing: Stimulus equivalence and pattern recognition* (pp. 41–71). PC Dodwell (Ed. and Trans.).

Wade, N., & Swanston, M. (1987). The representation of nonuniform motion in vision. *Perception*, *16*(5), 555–571.

Walker, R., & Findlay, J. M. (1996). Saccadic eye movement programming in unilateral neglect. *Neuropsychologia*, *34*(6), 493–508.

Wallach, H. (1959). The perception of motion. *Scientific American*, *201*(1), 56–60.

Wallach, H., & Becklen, R. (1985). Response to Gunnar Johansson's critical commentary. *Journal of Experimental Psychology. Human Perception and Performance*, *11*(3), 372–373.

Wallach, H., Becklen, R., & Nitzberg, D. (1985). Vector analysis and process combination in motion perception. *Journal of Experimental Psychology. Human Perception and Performance*, *11*(1), 93–102.

Wallach, H., O'Leary, A., & McMahon, M. L. (1982). Three stimuli for visual motion perception compared. *Perception & Psychophysics*, *32*(1), 1–6.

Warren, P. A., & Rushton, S. K. (2009). Optic flow processing for the assessment of object movement during ego movement. *Current Biology : CB*, *19*(18), 1555–60.

Watanabe, T. (1997). Velocity decomposition and surface decomposition - reciprocal interactions between motion and form processing. *Vision Research*, *37*, 2879–2889.

Wertheim, A. (1987). Retinal and extraretinal information in movement perception: how to invert the Filehne illusion. *Perception*, *16*(3), 299–308.

Wertheim, A. (1994). Motion perception during selfmotion: The direct versus inferential controversy revisited. *Behavioral and Brain Sciences*, *17*(2), 293–355.

Wurtz, R. H. (2008). Neuronal mechanisms of visual stability. *Vision Research*, *48*(20), 2070–89.

Yamada, Y., & Kawabe, T. (2013). Localizing non-retinotopically moving objects. *PloS One*, *8*(1), e53815.

Yin, C., Shimojo, S., Moore, C., & Engel, S. A. (2002). Dynamic shape integration in extrastriate cortex. *Current Biology : CB*, *12*(16), 1379–85.

Yoshimoto, S., Uchida-Ota, M., & Takeuchi, T. (2014). Effect of light level on the reference frames of visual motion processing. *Journal of Vision*, *14*(13).

Zaehle, T., Jordan, K., Wüstenberg, T., Baudewig, J., Dechent, P., & Mast, F. W. (2007). The neural basis of the egocentric and allocentric spatial frame of reference. *Brain Research*, *1137*(1), 92–103.

Zeki, S. (1990). Parallelism and functional specialization in human visual cortex. *Cold Spring Harbor Symposia on Quantitative Biology*, *55*, 651–61.

Zirnsak, M., Gerhards, R. G. K., Kiani, R., Lappe, M., & Hamker, F. H. (2011). Anticipatory saccade target processing and the presaccadic transfer of visual features. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(49), 17887–91.

Zirnsak, M., Lappe, M., & Hamker, F. H. (2010). The spatial distribution of receptive field changes in a model of peri-saccadic perception: predictive remapping and shifts towards the saccade target. *Vision Research*, *50*(14), 1328–37.

Zirnsak, M., Steinmetz, N. A., Noudoost, B., Xu, K. Z., & Moore, T. (2014). Visual space is compressed in prefrontal cortex before eye movements. *Nature*, *507*(7493), 504–507.

Zöllner, F. (1862). Über eine neue art anorthoskopischer zerrbilder. *Annalen Der Physik Und Chemie*, *117*, 477–484.