Winter 12-15-2014

Joint Representation of Translational and Rotational Components of Self-Motion in the Parietal Cortex

Adhira Sunkara
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

School of Engineering & Applied Science
Department of Biomedical Engineering

Dissertation Examination Committee:
Dora Angelaki, Co-Chair
Lawrence Snyder, Co-Chair
Pablo Blazquez
Gregory DeAngelis
Daniel Moran
Camillo Padoa-Schioppa
Barani Raman

Joint Representation of Translational and Rotational Components of Self-Motion in the Parietal Cortex

by
Adhira Sunkara

A dissertation presented to the
Graduate School of Arts & Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2014

St. Louis, Missouri

Table of Contents

List of Figures

Acknowledgements

First and foremost, I thank my advisor and mentor, Dora Angelaki, for always pushing me to be my best and providing immeasurable support during my PhD. I was fortunate enough to have had not one, but two advisors through this process – thank you Greg DeAngelis for all your help and advice on my projects.

I would like to thank Larry Snyder for being the co-chair of my committee and helping me navigate my way to graduation. I also truly appreciate how accommodating and encouraging Pablo Blazquez, Camillo Padoa-Schioppa, Daniel Moran, and Barani Raman were as part of my thesis committee.

The biggest thank you of all goes to my family, who deserve much of the credit for this accomplishment. I owe so much to my parents, for always believing in me; making me appreciate the importance of hard work and perseverance; most of all, for loving me unconditionally.

The award for having the toughest job while I was finishing up my PhD goes to my husband, Ari Rosenberg, who also happens to be an amazing neuroscientist. Thank you for proof-reading my papers, bringing me take-out, easing my worries and putting up with me through all the ups and downs. Last, but certainly not least, I am thankful to Lena for the invaluable moral support – sitting by me as I worked late nights/early mornings and always bringing a smile to my face.

Adhira Sunkara

Washington University in St. Louis

December 2014

To Mummy & Papa.

ABSTRACT OF THE DISSERTATION

Joint Representation of Translational and Rotational

Components of Self-Motion in the Parietal Cortex

by
Adhira Sunkara

Doctor of Philosophy in Biomedical Engineering
Washington University in St. Louis, 2014
Professor Dora E. Angelaki, Co-Chair
Professor Lawrence H. Snyder, Co-chair

Navigating through the world involves processing complex visual inputs to extract information about self-motion relative to one's surroundings. When translations (T) and rotations (R) are present together, the velocity patterns projected onto the retina (optic flow) are a combination of the two. Since navigational tasks can be extremely varied, such as deciphering heading or tracking moving prey or estimating one's motion trajectory, it is imperative that the visual system represent both the T and R components. Despite the importance of such joint representations, most previous studies have only focused on the representation of translations. Moreover, these studies emphasized the role of extra-retinal cues (efference copies of self-generated rotations) rather than visual cues for decomposing the optic flow. We recorded single units in the macaque ventral intraparietal area (VIP) to understand the role of visual cues in decomposing optic flow and jointly representing both the T and R components. Through the following studies, we establish that the visual system can rely on purely visual cues to derive the translational and rotational components of self-motion. We also show for the first time, joint representation of T and R at the level of single neurons.

Chapter 1– Introduction

Navigating through the world involves processing complex visual inputs to extract information about self-motion relative to one's surroundings. Self-motion consists of translations and rotations, which result in movement and distortion of the images projected onto the retina. Gibson (1950) termed these retinal velocity patters as 'optic flow'. When translations (T) and rotations (R) occur simultaneously, as is often the case during navigation, the resultant optic flow is a sum of the corresponding T and R flow patterns. However, depending on the navigational task at hand, an estimate of the translational component, the rotational component, or both may be required. For instance, while walking down a sidewalk and simultaneously looking at a passing car using eye or head rotations, the brain must discount the visual consequences of rotations to estimate and maintain one's direction of translation (i.e., heading). Similarly, rotation estimates are also crucial for tasks such as estimating the angular velocity of a tracked object (e.g. stalking moving prey) or spatial constancy. Travelling along a curved path, on the other hand, requires both the translational and rotational estimate to correctly predict complex trajectories and navigate around obstacles (Royden et al., 2006; Cheng and Li, 2012). Hence, successful navigation requires the visual system to decipher both the translational and rotational components of self-motion. Despite the significance of visual navigation, little is known about the joint representation of T and R during self-motion.

When rotations are self-generated (such as eye or head rotations), they are accompanied by efference copies of the motor command signal (extra-retinal signals). For over a century, studies have emphasized the role of such extra-retinal cues in estimating rotations and

consequently deriving the translation direction. However, this solution for decomposing self-motion has certain limitations (Bridgeman, 2007). Firstly, there are several different sources of rotation that can occur simultaneously, such as eye-in-head (R_{EH}), head-on-body (R_{HB}), and body-in-world (R_{BW}) rotations. Therefore, in general, multiple efference copies would need to be added to achieve an estimate of the total rotation. However, these signals are inherently noisy and such a strategy would be inefficient due to the potential compounding of the noise that is associated with each signal (Gellman and Fletcher, 1992; Li and Matin, 1992; Crowell et al., 1998). In addition, rotations are not always accompanied by efference copy signals; for instance, during curvilinear motion, the angular velocity is a consequence of one's trajectory and not self-generated eye/head rotations. Hence, it is likely that the brain uses alternative cues for decomposing optic flow.

Several theoretical studies have shown that the difference in properties of translational and rotational flow optic flow can be used to decompose translations and rotation based purely on the optic flow patterns. This potential solution to resolving self-motion components has certain advantages over a purely extra-retinal solution. Since retinal image motion is determined by the total translation and rotation of the eye relative to the world (T_{EW} and R_{EW}), all the relevant information required to decompose the components of self-motion are readily available to the visual system irrespective of the sources of rotation. Thus, if the brain can use optic flow to directly estimate translations and rotations, such mechanisms may provide a complementary and potentially more efficient way to decompose rotations and translations.

In this chapter, I will discuss in detail the properties of optic flow that make such computations feasible, as well as psychophysical and electrophysiological literature summarizing our current understanding of the relative role of visual and non-visual cues in

separating translations from rotations. As will be evident from this chapter, several studies have focused on the encoding of translations (Royden et al., 1992; Royden, 1994; Bradley et al., 1996; Crowell et al., 1998; Page and Duffy, 1999; Zhang et al., 2004), but none have attempted to understand if and how a joint representation of the two is achieved. In order to truly understand how the visual system contributes to navigation, it is imperative to study if and how translations and rotations are represented jointly when both are present simultaneously. The research presented in the subsequent chapters will address this fundamental question.

1.1 Characteristics of Optic Flow

Quantitative characterization of the retinal flow field is necessary to build a theoretical framework for separating the translational and rotational components of the visual stimuli. During simultaneously translations and rotations, the resulting optic flow is simply a vector sum of the translation and rotation velocity vector fields. Longuet-Higgins & Prazdny (1980) showed that these two components are separable based on the fundamental differences in the two vector fields. They consider a simplified framework, where the eye can be thought of as a pinhole camera that is in motion through a static

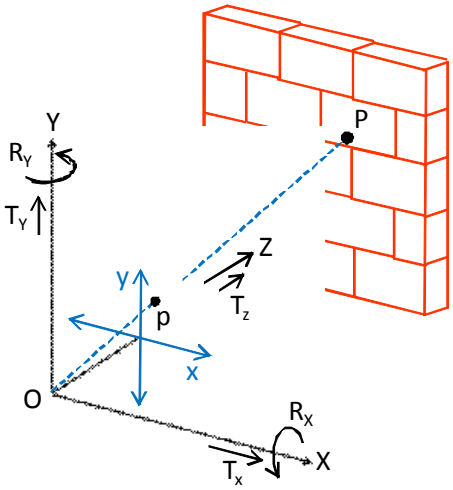

Figure 1-1. Pin-hole representation of the retina. Coordinate system based on Longuet-Higgins & Prazdny (1980) showing the center of the eye at O and the plane xy as an approximation of the retina. Any point P projects on the retinal coordinate plane xy.

environment. The retina can then be idealized as a plane at a distance of the focal length from the pinhole. As shown in Figure 1-1, let O be the instantaneous position of the pinhole camera and OXYZ be the Cartesian coordinate system fixed with respect to the eye. Let (T_X, T_Y, T_Z) be the translation velocity of the eye and (R_X, R_Y, R_Z) be its angular velocities. (X, Y, Z) are the instantaneous coordinates of any world-fixed point P in the scene in the eye coordinate system. p denotes the planar projections of any point P on to the retina and can be measured in the retinal planar coordinates of (x,y). Using this coordinate system, we can apply the resulting equations to a variety of combinations of translations and rotations in different visual scenes.

1.1.1 Pure Translation

Based on the coordinate system described above, let us first consider the simple condition of translation towards a fronto-parallel plane. The velocity components of P in the moving eye reference frame will be (-T_X, -T_Y, -T_Z) and its retinal position p, can be represented in retinal coordinates as:

$$(x, y) = \left(\frac{X}{Z}, \frac{Y}{Z}\right) \qquad (1)$$

Substituting from Eq.1, the velocity components of the point moving across the retina are:

$$\dot{x_T} = \frac{\dot{X}}{Z} - \frac{X\dot{Z}}{Z^2} = -\frac{T_X}{Z} + \frac{XT_Z}{Z^2} = -\frac{T_X}{Z} + \frac{xT_Z}{Z} \qquad (2)$$

$$\dot{y_T} = \frac{\dot{Y}}{Z} - \frac{Y\dot{Z}}{Z^2} = -\frac{T_Y}{Z} + \frac{YT_Z}{Z^2} = -\frac{T_Y}{Z} + \frac{yT_Z}{Z} \qquad (3)$$

The optic flow components of translation are dependent on the structure of the scene and change with the depth of point P as well as the (x,y) coordinate of projection onto the

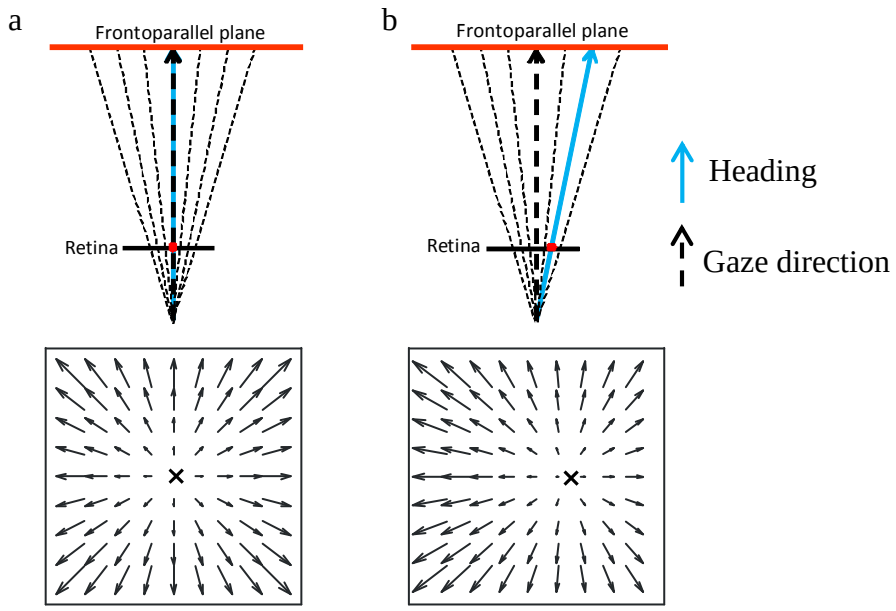

Figure 1-2. Optic flow patterns on the retina during translation towards a fronto-parallel plane. The optic flow pattern is radially symmetric about the focus of expansion (FOE), marked with an 'X'. (a) Heading and gaze angle are straight ahead. (b) Heading to the right with gaze angle still perpendicular to the FP plane.

retina. The result is a radial vector field expanding out from the projection of the translation vector (T_X, T_Y, T_Z) on the retina. Figure I-2a is an example of translational optic flow where an observer is facing the wall such that the retinal plane is parallel to the wall and heading straight towards the wall (0, 0, T_Z). As shown in the figure, the velocity vectors are symmetric about the FOE (indicated by a red dot) and lie at (0, 0) on the retinal plane in this example. The retinal location of the FOE can be defined in this coordinate system as (x_0,y_0), where:

$$x_0 = \frac{T_X}{T_Z}, \qquad y_0 = \frac{T_Y}{T_Z} \tag{4}$$

If we consider the same gaze position of the observer with respect to the wall, but a translation to the right (T_X, 0, T_Z), (x_0,y_0) is now shifted to the right, indicating a rightward heading direction (Figure 1-2b). The velocity profile is still symmetric about (x_0,y_0) because

the Z component of the distance between any point on the plane and the origin is still constant.

In the case of an infinite 3D cloud stimulus, there is no world-fixed frontal plane (like a wall). In such a situation, the velocity vector patterns are similar to Figure I-2b at any given plane depth. However, the magnitude of the vectors decreases with the distance of a point from the retina. In the following sections we will discuss how this feature (motion parallax) distinguishes translations from rotations.

1.1.2 Pure Rotation

Eye or head rotations also result in specific optic flow patterns projected onto the retina. For instance, a leftward eye movement adds rightward velocity patterns to retinal images (Figure 1-3). Using the same coordinate system as above, a simple case of only horizontal eye rotation (R_Y) yields the following equation defining optic flow resulting from pure rotations:

$$\dot{x}_R = -R_Y + R_Y x^2 \qquad (5)$$

$$\dot{y}_R = -R_Y xy \qquad (6)$$

As evidenced in Figure 1-3, rotational optic flow cannot be described as simple laminar flow. It results in a shearing and deformation of the image that is referred to as 'dynamic perspective cues'. Specifically, the velocity vector is dependent on the horizontal distance of a

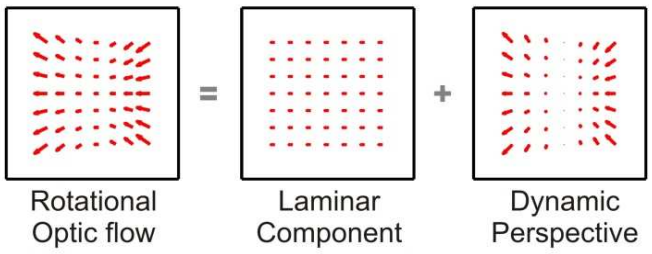

Figure 1-3. Components of rotational optic flow. Rotational optic flow is a sum of laminar flow and dynamic perspective cues.

point from the center of the retina. This is evident from the x^2 term in Equation (5) and the xy term in Equation (6). Hence, rotational optic flow can be characterized as a combination of laminar flow and dynamic perspective cues resulting from the changing orientation of the eye relative to the scene (planar image projection shown in Figure 1-3). Theoretical studies have proposed that the latter may play an important role in estimating and discounting the rotational component of optic flow to estimate heading (Koenderink and van Doorn, 1976, 1981; Grigo and Lappe, 1999). Indeed, a recent electrophysiological study in MT provides evidence that the visual system may be capable of utilizing these dynamic perspective cues to estimate rotations (Kim et al., Under revision).

The second feature of rotational optic flow is that the velocity vectors are not dependent on the depth variable, Z. In comparison, Equations (2) and (3) show an inverse dependence of the translational velocity vectors and distance from the retina (depth). This difference in features of translation and rotation has implications for the local motion parallax in a scene (discussed below).

1.1.3 Translation plus Rotation

In the case of an eye or head rotation in conjunction with translation, the flow field becomes more complex than described above. For instance, a leftward rotation (adding rightward velocity vectors) causes the FOE at a given depth plane to shift to the left. Therefore, the FOE no longer indicates the direction of translation. In order to extract the accurate heading direction from such a flow field, the brain must decompose the flow field into translational and rotational components. There are two broad sources of information that the brain has to extract such information – (1) Extra-retinal signals, (2) Retinal signals.

However, as discussed earlier, the extra-retinal solution has limitations and it may be more efficient for the brain to use retinal signals instead. Here, I will discuss the significant characteristics of retinal flow fields in different scenes and the information they contain that allow us to solve the optic flow problem without using extra-retinal signals.

Since rotational optic flow is depth-invariant, but translational optic flow isn't, one retinal-based solution that has been suggested is the use of local motion parallax cues (Helmholtz, 1924; Longuet-Higgins and Prazdny, 1980; Koenderink and van Doorn, 1987; Tomasi and Shi, 1996). These motion parallax cues can be defined as the difference in the velocity of two points projecting close to each other on the retina, but differing in depth. Since translation velocity vectors are dependent on the depth, such a spatial derivative in depth of velocities at any given location on the retina result in a motion parallax field. The point of zero motion parallax then indicates the direction of translation. The addition of rotations does not disrupt the motion parallax field, since rotations are depth-independent. As a result, motion parallax is an important cue for estimating translations in the presence of rotations. Once an estimate of the translation is determined, the visual system can attribute the remaining flow field to rotations.

The dependence or independence of velocity vectors to depth has another important consequence. In the presence of a deep 3D structure in the scene, the closest depth planes carry largely translational information while the farther planes are highly influenced by the rotational vectors. Therefore, the rotation rate can also be estimated based purely on the distant planes, whereas the translation can be estimated based on the flow fields at near depth planes (Li and Warren, 2000; Saunders and Niehorster, 2010). This is a simple alternative, but highly susceptible to noise as a large part of the stimulus is essentially being ignored in the

estimation of each component. While, the use of local motion parallax cues provides a more robust method, it is also limited by the structure of the scene. The solution is most effective in the presence of a rich 3D depth structure, such as walking through a forest. In the absence of such structure, the brain must rely on cues that are not contingent upon the density of the visual scene, such as the dynamic perspective cues resulting from rotations.

In addition to the FOE being dependent on the distance of a given plane from the retina, the position of the FOE on the retina also changes over time as the direction of gaze changes during rotations. The combination of these spatial and temporal changes in the FOE has different consequences on the optic flow patterns based on the structure of the scene. Let's consider two examples of different visual scenes to examine the effect of this feature on how optic flow evolves over time – moving towards a fronto-parallel (FP) wall and moving through a dense 3D cloud of dots (Figure 1-4). Figure 1-4a, b show the result of translations and eye/head rotations in the two visual scenarios. In both scenes, the heading direction (blue cross) is fixed both spatially and temporally in world coordinates.

In a 3D environment, optic flow fields from different depth planes are simultaneously projected onto the retina, resulting in the absence of a coherent FOE. If the visual system completely compensates for the added rotations and represents only the translational component, the 'corrected' FOE would still change its position on the retina over time. In other words, the translational estimate in retinal coordinates would change in relation to the changing gaze angle with time (Figure 1-4a, dashed line). Therefore, in order to ascertain heading in world/body coordinates, the brain must also perform a coordinate transform and encode headings in non-retinal coordinates. Compensating for rotations and reference frame transformation between retina and body/world are two separate concepts that have previously

been erroneously considered interchangeable.

Figure 1-4b depicts the realistic situation where a fronto-parallel wall moves closer to the observer during translation, resulting in a temporally varying FOE shift based on the distance of the wall at a given time. In contrast, in the 3D scene, the range of planar depths projected onto the retina remains constant over time. The second row shows the position of the FOE in retinal coordinates over time. Due to the time-varying depth structure of the FP plane, the FOE location changes on the retina non-linearly. However, after subtracting the rotation

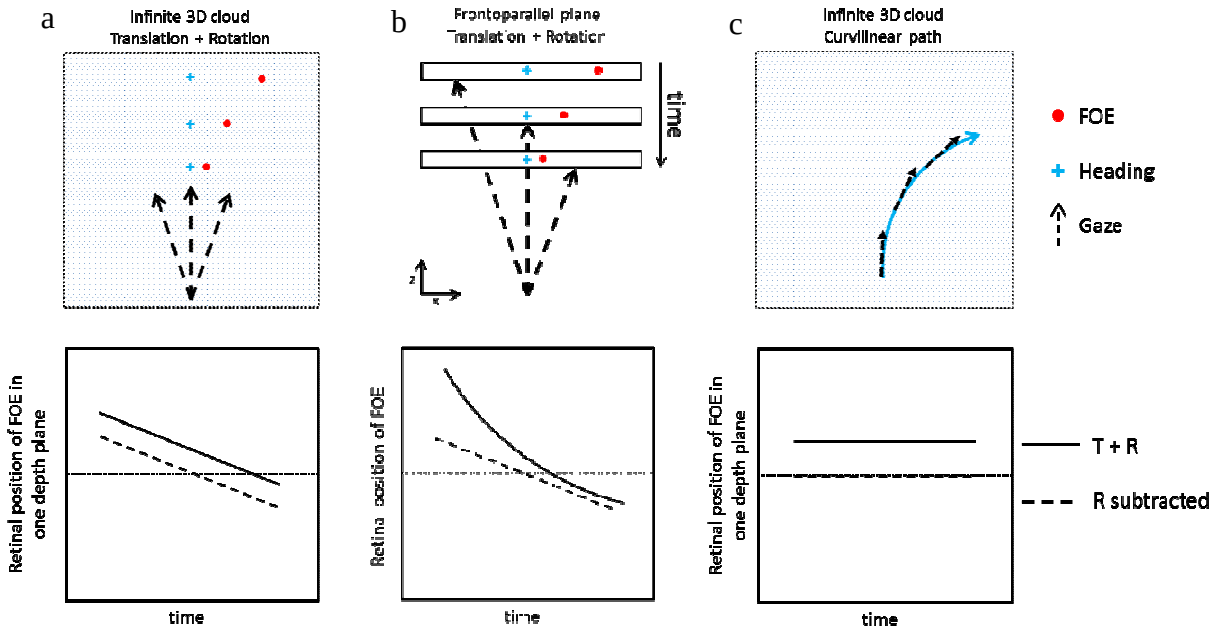

Figure 1-4. Temporal evolution of optic flow during different viewing conditions in the present of both T and R. Columns a and b are examples of linear translation plus eye/head rotations through a 3D cloud and towards a FP plane respectively. Column c depicts curvilinear motion through a 3D cloud. The bottom row shows the FOE position with respect to time before and after subtracting out the rotational component of the optic flow (solid and dotted lines respectively). In the case of linear translation, the position of the FOE on the retina changes over time; whereas it remains constant during the CL motion.

vectors, the shifts of the FOE are linear (assuming constant rotation velocity). An interesting feature between the two visual scenes is that the information available spatially in the 3D cloud scene (at different depth planes) is available as temporally in the FP plane scenario. Therefore, theoretically, if the brain can integrate this velocity information over time, it is possible to extract motion parallax cues even in a FP plane stimulus based on the time-varying estimates. However, it is unclear if such second-order processing is used in the visual system for optic flow solutions.

Even though the 3D cloud has enough information to extract heading from purely retinal signals, studies have not been able to conclusively show evidence of extracting translations in the absence of extra-retinal signals (see sections below). While there are several experimental factors that seem to affect the ability to use purely retinal signals, it is often the case that subjects mistake simulated rotation plus translation (T+R) as travelling on a curvilinear (CL) path with gaze position fixed with respect to the instantaneous heading (Figure 1-4c). Such a situation would be easily distinguishable in the presence of extra-retinal signals because the T+R condition would have efferent eye velocity signals present, but not in the case of CL motion. In the case of small rotation amplitudes (gaze angles), the instantaneous velocity profiles for the two conditions are very similar (see Royden, 1994 for derivation). A T+R stimulus could therefore be interpreted as CL path of radius T/R due to the similarity in optic flow patterns (especially in the absence of second order processing).

However, there is one major difference between the two conditions. From Figure 1-4c we can see that during CL motion, heading with respect to the retina remains constant throughout the motion. In the case of the 3D cloud (and even FP plane), the FOE is constantly changing position on the retina even if it is fixed in world coordinates. In other words, the translation

estimate is constant in retinal coordinates, but changes temporally in world coordinates during CL motion. The inverse is true for eye or head rotations during straight translations. This is an important distinction that becomes more prominent for larger rotations. Therefore, the temporal evolution of optic flow can hold key information about self-motion trajectories and needs to be studied further in addition to the joint estimation of translational and rotational components.

1.2 Psychophysics

During everyday life we are constantly moving and making eye/head movements, making pursuit compensation an important and practical part of accurately perceiving self-motion. Consider a situation where a person is driving a car and simultaneously making a head movement to the left to check his blind spot; if humans were not capable of compensating for such eye and head movements, the driver might perceive self-motion to the left and could cause an accident by trying to correct his perceived direction. We know this does not happen and people in most real world situations are capable of accounting for the rotations resulting from their head and eye movements. However, the mechanism by which such compensation happens is unclear and a highly debated topic.

As described in the previous section, there are several cues, both retinal and extra-retinal that can theoretically be used in the process of pursuit compensation. One way to understand the relative importance of each of these cues is to systematically study the changes to perception of self-motion during the absence of each of the cues. Several psychophysical studies have attempted to understand the role of the different cues but due to a lack of constraints on other factors, these studies are difficult to compare and have seemingly

conflicting results. Also note that the psychophysics literature on navigation in the presence of both translations and rotations focuses primarily on estimating translations by discarding rotations. Therefore, this section will delve into some of the results and points of debate regarding this question of 'pursuit compensation'.

In order to study the role of retinal signals in the process of pursuit compensation, many studies design a 'simulated pursuit' condition in which the retinal flow field corresponding to an eye pursuit is presented to the subject while the subject is fixating. The initial findings indicated that subjects were incapable of compensating during such simulated conditions despite the similarity in the retinal image during the real and simulated pursuit conditions (Royden et al., 1992; Royden, 1994; Banks et al., 1996; Crowell et al., 1998). These studies point to the theory that the extra-retinal signal is an indispensable cue, without which, humans are incapable of accurately compensating for rotations. However, as discussed in the above, it is possible to compensate for pursuit based on purely retinal mechanisms if the scene has sufficient visual information (Koenderink and van Doorn, 1987). More recent psychophysical studies have shown that in the presence of richer visual stimuli, subjects' performance during simulated pursuit improves, but depends significantly on the instructions given to subjects (Li and Warren, 2002, 2004). We will discuss some of the stimuli features that contribute to the visual processing of optic flow.

1.2.1 Rotation Velocities

Some of the first studies on heading judgments during pursuit reported that subjects were able to accurately determine their heading during real as well as simulated rotations (Warren and Hannon, 1988; Warren and Hannon, 1990). They showed that heading thresholds were below 1.5° for a random-dot as well as a 3-D cloud stimulus in both the real and

simulated pursuit conditions, indicating that the direction of motion could be perceived even in the absence of extra-retinal signals. Subjects also reported that they experienced an illusion that their eyes were actually rotating, supporting the hypothesis that purely retinal cues can be used to extract direction of translation even in the presence of rotation. However, the mean rotation rate used in the studies was less than 1°/s with total rotation amplitude ranging from 0.5° to 3.5°.

In contrast, Royden et al (1992; 1994) tested the same hypothesis but with larger rotation rates of up to 5°/s and reported large errors in heading during the simulated rotation conditions. The errors made by subjects were in the direction of the simulated rotation corresponding to the shifts in FOE; the errors also increased as a function of the rotation rate. These findings indicated that extra-retinal signals might be necessary at higher rotation rates. They reported higher thresholds during the simulated pursuit condition for ground plane, fronto-parallel plane, 2 transparent FP planes and the 3-D cloud stimulus. We will discuss the differences between each of these stimuli subsequently.

It is important to note that while the conflicting results of these two studies reports may indicate the need for extra-retinal signals in pursuit compensation during larger rotations, there are experimental differences in the studies that may have influenced the quantitative differences between the reports. Both the Warren et al. and the Royden et al. studies used high translation speeds of 190cm/s and 250cm/s respectively during the 3-D cloud stimulus. For a given depth plane, higher translational speeds result in smaller shifts in the FOE, which means that even though Royden et al. used larger rotations, it was somewhat offset by the larger translational velocity. Another key difference was that the Royden experiments used a denser 3-D cloud that extended from depths of 0 – 3730cm, whereas the Warren study used a volume

with depths from 690cm to 3730cm. Hence, the shift in FOE was larger in the Royden study with the shift at the closest plane being 14.7° and only 2.5° in the Warren study. In the Royden stimulus, beyond the depth plane of 1400cm, the velocity profile was dominated by rotations such that there was no FOE on the retinal plane whereas in the Warren stimulus, the shift in FOE was 14.5° at the farthest depth plane (still on the visible screen). Since the velocity profiles of the closer depth planes are dominated by translation and the farther ones are dominated by rotation, there is a difference in the information present in the two stimuli. Royden et al used their higher density and larger range 3-D cloud and also found that for rotation rates less than 1°, the simulated pursuit responses were indistinguishable from the real pursuit condition. However, they did not calculate the exact thresholds because the task was a 7AFC task, so it is possible that the errors were smaller than the sensitivity with which the 7AFC task could measure them. There is an important distinction to be made between absolute rotation velocities and the relative rotational velocity with respect to translation, which determines the location of FOE on a plane. Based on these two studies, it is not clear if the difference in responses is due to larger rotations or if it corresponds to the location of the FOE at various depth planes. If the responses depend on the latter, factors such as cloud depth and distance would bias the results. Therefore, there is still the need for a systematic study that can disambiguate FOE positions from rotation rates.

Due to the extra 0-690cm depth planes added in the Royden stimulus, it can be argued that the Royden study had more information about the translational direction, further supporting the hypothesis that extra-retinal signals become more important with larger rotation rates. However, it is important to keep in mind that during the simulated pursuit condition, the presented stimulus results in a conflict between retinal and extra-retinal signals

– optic flow may signal the presence or rotations, but the lack of efference copies from real eye/head rotations, indicates zero rotation. This conflict only increases with larger rotation velocities. Therefore these results could also be interpreted as a reweighting of the potentially more reliable cue (extra-retinal signals) and may not be indicative of the mechanism used to determine heading during real pursuit. Furthermore, the simulated pursuit also suffers from the disadvantage that the optic flow resulting from pursuit during translation is very similar to the optic flow from travelling on a curved path.

1.2.2 Path ambiguity

In the simulated pursuit condition of many studies, subjects reported that they perceived motion along a curvilinear path rather than a straight line with path independent rotation (Royden et al., 1992; Royden, 1994; Crowell et al., 1998; Ehrlich et al., 1998). Hence, the poor perceptual responses during simulated pursuit described above could be due to this misinterpretation of the scene as translation along a curved path. The instantaneous flow fields during the two conditions are very similar, especially for smaller rotational displacements (Royden, 1994), but the implications for the path and heading are different (Figure 1-4). Therefore, resolving the ambiguity between path-independent rotations and curvilinear motion is very important for successful navigation.

The same instantaneous velocity profiles can be generated with a curved path where the radius of curvature is the ratio of the translation and rotation velocities. Furthermore, there are infinite combinations of path-independent rotation and curvilinear motion that can generate a wide range of instantaneous flow fields that may be indistinguishable from each other. However, there are time-varying differences in the two stimuli that could theoretically be used to distinguish the two paths. In the case of eye/head rotation, the direction of

translation is changing constantly with respect to the retina, but is constant in world coordinates (Figure 1-4a). In the case of curvilinear motion, the gaze angle with respect to the trajectory is constant resulting in the velocity profile projected on to the retina to be constant over time (Figure 1-4c). A deformation of the velocity profiles could also be caused by movement along a curved path with a changing radius of curvature; however the heading direction (FOE) will still remain constant in retinal coordinates. (Rieger and Lawton, 1985) showed that the ambiguity in the two flow fields can theoretically be resolved using differences in the acceleration profiles for the two conditions. But in the visual system, such second-order acceleration components might not be effective for distinguishing the two types of trajectories, since human sensitivity to visual acceleration is poor (Schmerler, 1976).

Another potential solution is for the visual system is to extract path information by tracking elements, since the positions of points diverge over time during the two trajectories. This is a computationally intensive solution that would require tracking the elements individually over time and integrating the path lines. However, this solution becomes more feasible in the presence of reference objects whose positions can be easily tracked over time. Therefore, theoretically, there is enough information in the retinal flow alone to distinguish the two motion paths but there are conflicting results about whether this information is indeed used by the visual system.

Earlier experiments such as the ones presented in Banks et al. (Banks et al.), asked the subjects to place a cursor in their perceived heading direction, while many other studies used probes with no clear depth position for reporting perceived heading. However, in order to determine the perceived path, it is necessary to specify the depth of the probes. In the absence of probe depth information, the subjects' responses are ambiguous because it is possible that

they perceived a curved path and responded according to some instantaneously perceived direction of self-motion. In later experiments (Ehrlich et al., 1998), the subjects were presented posts at different depths to measure the error in their judgment with respect to depth. If the subjects' errors are larger at larger post distances, it is indicative of the subject perceiving a curved path rather than translation with rotation. The results from Ehrlich et al seemed to suggest that subjects perceive simulated rotation as a curved path. The data fell between the correct heading and the curved path, which the authors attributed to screen/retinal center bias. However, once again, the rotation rates used were much smaller than the translational speeds (R = 5°/s, T = 200cm/s) and the cloud started 300 cm from the observer, which has the same problem as described earlier – less translational information from the close depth planes. The paper also did not present any real pursuit data to allow for a comparison of the two conditions. Curiously, the errors associated with the simulated path stimuli was also larger than most of the previously reported studies.

The experiments conducted by Li and Warren (2004) provide some further insights into how the path ambiguity is perceived by subjects. Like in previous studies, they showed subjects simulated pursuit stimuli, but gave specific instructions to the subjects that either informed them of the type of path (straight or curved) or gave no path-related information. Interestingly, the errors during simulated pursuit were rather small in the straight path directive and closer to the straight path in the absence of any directive. Even though the same stimuli were used in all conditions, the subjects perceived the stimuli to be different based on the instructions given. One major difference in their stimuli was the use of either a textured ground plane or dense 3D posts. However, it is interesting to note that the perceptual differences were based simply on the instructional biases about the type of stimulus. Royden

et al. (2006) repeated similar experiments and also found that error in perceived heading decreased when the subjects were explicitly told the stimulus corresponded to a straight path. While these results are very interesting, they still leave room for ambiguity in the interpretation. For instance, it is possible that during the straight path directive, subjects responded based only on initial instantaneous translation even if they perceived a curved path. Despite the quantitative differences in the results of the two groups, both the results seem to indicate that the visual system is capable of extracting heading information from purely retinal signals even in the presence of rotation. These studies point to the hypothesis that retinal cues can contribute to pursuit compensation, but only if the cue conflict between retinal and extra-retinal signals can somehow be resolved.

1.2.3 Trigger Model

In order to at least partially resolve the retinal/extra-retinal cue conflict, studies were conducted by Crowell and Andersen (2001) and Banks et al. (1996), testing a 'trigger model'. The idea of the trigger model was that the brain utilizes retinal cues for pursuit compensation, but only in the presence of extra-retinal signals. Therefore, the efference copies were not utilized as an actual measure of the rotation velocity, but were rather a gate or trigger for resolving the path ambiguity. In order to study this, both studies used stimuli where the rotation was a sum of both real and simulated pursuit. Like much of the psychophysical literature on pursuit compensation, there results also appear contradictory. Banks et al argued that in the case of mixed simulated and real pursuit, the subjects' behavior most closely resembled an extra-retinal model, whereas Crowell and Andersen argued that the responses are closer to their proposed trigger model.

Once again the results are confounded by the difference in experimental methods and

stimuli. Most importantly, the Crowell and Andersen study used conditions where simulated pursuit was in the direction opposite to the real pursuit, whereas Banks et al. used different ratios of the two rotation components. The part of the Crowell and Andersen data that supports the trigger model is based on these opposite combinations of real and simulated pursuit. The pursuit speeds used were higher in the Crowell experiments and the stimulus used was a ground and sky plane rather than a 3D cloud. The other issue is that though both studies use a post to estimate subject perception of heading, the Banks experiment did not specify post position in depth which, as discussed earlier, could result in ambiguity in the understanding the strategy used by subjects. The Crowell experiments also used shorter duration stimuli (300ms) based on the hypothesis that extra-retinal signals have a longer latency (Grigo and Lappe, 1999). Therefore the results from these two studies may not necessarily be inconsistent and provide some evidence for the trigger model.

1.2.4 Summary

The wide range of parameters in these experiments and the quantitative differences in results make the understanding of the mechanism of pursuit compensation based on psychophysics literature difficult. However, there are some key take-away points as well as gaps in the knowledge which need to be addressed.

1. Most of the experiments used very large translational velocities (upwards of 100cm/s) and relatively small rotational velocities and amplitudes (< 9°/s). This results in the dynamic perspective cue being relatively small and weak.

2. There is no disambiguation between absolute eye velocities and the relative ratio of translation and rotation velocities. Therefore, earlier studies claiming that larger eye velocities resulted in larger errors during simulated pursuit were more likely due to a larger shift in the

FOE (smaller T/R ratio). Studies need to be performed where the T/R ratio is kept constant and the eye rotation velocity is varied and vice versa. This could also give us insight into the role of extra-retinal signals and the noise associated with such signals.

3. There is no clear consensus on the role of stereo in solving the rotation problem. Van den Berg and Brenner (1994), and Ehrlich et al. (1998) have conflicting results regarding the importance of binocular disparity. The role of stereo has also not been studied during the situation of approach to a FP plane. In the absence of motion parallax cues, this would be a great model to study the role of stereo information (since the wall gets closer and the depth information could inform the relative roles of T and R).

4. The ambiguity of the T+R and curved path needs to be resolved in a more rigorous way rather than changing verbal instructions. For instance, an external perceptual cue (such as vestibular motion) would provide a more quantitative model for studying the role of cue conflict between retinal and extra-retinal signals.

5. The focus of all these psychophysical studies was the encoding of translations by discarding or subtracting the distortions resulting from rotations. However, none of these studies attempted to evaluate how good the subjects were at estimating rotations in the presence of translations. In order to thoroughly understand visual navigation, the joint estimation of both the translation and rotation needs to be studied.

1.3 Electrophysiology

Despite the mixed results in the psychophysics literature, it is clear that during ‘real pursuit’, most subjects compensate for their eye movements and there is some evidence for the role of both retinal and extra-retinal cues. As with the psychophysics literature,

electrophysiological studies have also largely focused on the encoding of translations in the presence of rotations. Here, I will detail the current understanding of the neural basis of such pursuit compensation. This section will focus mainly on electrophysiological studies from two macaque cortical areas that encode heading based on optic flow – Medial Superior Temporal area (MST) and Ventral Intraparietal area (VIP).

1.3.1 Pursuit Compensation in Medial Superior Temporal Area (MST)

MST has been extensively studied for its responses to complex non-uniform visual stimuli. The cells in MST usually have large receptive fields that are necessary for the analysis of complex motion such as optic flows. Tanaka and Saito (1989) showed that MST has stronger responses to larger stimulus fields pointing towards evidence of spatial pooling of information. MST cells are also selective to spiral motion (Tanaka and Saito, 1989; Graziano et al., 1994), which are often part of the retinal information during gaze rotation with translation. While a lot of MST cells show preferences towards leftward or rightward translations, this is ideal for making fine discriminations of forward heading (Gu et al., 2006). All these feature point towards MST playing a vital role in the perception of self-motion. Furthermore, some MST neurons respond to smooth pursuit eye movements even in the dark indicating that it has direct extra-retinal inputs (Newsome et al., 1988).

The initial study that tested pursuit compensation in MST was by Bradley et al. (1996). The study simulated translation towards a fronto-parallel plane and test MST response tuning curves for fixation only, real pursuit and simulated pursuit conditions. Based on the rotation speed of 15.7°/s, the effective shift in the FOE for their stimulus was 30°. The results of the experiment showed that the mean shift during real pursuit was 15°, but only 58% had

shift larger than 10° and only 27% of the cells compensated completely (>30° shift). In comparison, during simulated pursuit, the optimal shifts were around 0o. Similar experiments performed by Shenoy et al. (1999) reported different proportions of the population as compensating. The Shenoy study expected a mean shift of 24° based on their rotation and translation parameters. The results indicated a mean shift of 21.2° in the real pursuit condition and 12.5° during simulated pursuit. However, the distribution of shifts was broader and skewed in the positive direction for real pursuit (75th percentile at 41°) and sharp for simulated pursuit (75th percentile at 15.5°). The study also studied the effect of VORC, which will not be discussed in this review. The Shenoy results showed larger compensation for both real and simulated pursuit. One of the differences in the two studies was that Bradley et al used a 50x50° stimulus aperture whereas Shenoy et al used only 18°. Furthermore, the rotation rate used in the Bradley stimulus was larger, 15.7°/s as compared to 9.2°/s in the Shenoy experiments.

Around the same time, Page and Duffy (1999) published results for MST tuning shifts during pursuit. Instead of just studying tuning properties for heading along the horizontal plane, they used a total of 9 heading directions – straight ahead and 8 directions distributed 45° intervals, 30° eccentric from the center. This provided a more complete tuning curve of the MST cells, but they did not have a simulated pursuit condition in their protocol. There experiments found that only 18% of MST cells maintained the response amplitude of the fixation-preferred FOE during pursuit (i.e. 82% showed gain modulation) and only 5% maintained the fixation-preferred FOE during pursuit (i.e. 95% showed shifts corresponding to no compensation) and only 3% maintained both the amplitude and the preferred FOE. These reported numbers seem to indicate that a much smaller percentage of MST neurons

compensated for pursuit. However, it must be noted that due to the circular data set, the authors did not calculate a tuning curve for their data and the above percentages are based on significant changes between fixation and pursuit conditions (F statistics). Therefore, if the cells did not compensate completely (like most cells reported in the previous studies mentioned), they might get included as part of the 97% of non-compensatory cells. The small percentage of compensatory cells can also be attributed to the sampling of FOE tuning in more than just the horizontal plane. If MST cells have 3D optic flow tuning, recording in the horizontal plane would just be a slice through the full tuning curve. This would be sufficient to characterize pursuit compensation only if we assume a separate mechanism for selecting neurons based on pursuit axis. Although only 3% of the cells in the study showed perfect compensation, a population vector analysis showed no significant differences between the vectors recorded during fixation and pursuit. The contrast between single cell and population vector results is very fascinating and warrants further investigation with different models of population pooling.

1.3.2 Pursuit Compensation in Ventral Intraparietal Area

Much like MST, VIP has characteristics that support its role in heading perception and both areas have many similarities in their responses to visual stimuli. VIP is a multi-modal area that has been shown to have visual, vestibular, somatosensory and auditory responses. The presence of multi-modal responses could make VIP an ideal candidate for combining various stimuli in order to determine heading directions more accurately. While VIP has slightly smaller receptive fields compared to MST, most neurons are still well-tuned to optic flow stimuli and a wide range of headings is represented in the population (Bremmer et al., 2010; Chen et al., 2011). Another important feature of VIP is that RFs in VIP shift in order to

partially or completely transform the RF into head-centered coordinates (Duhamel et al., 1997; Avillac et al., 2005; Avillac et al., 2007). This is an important feature as it implicates the presence of strong extra-retinal inputs. Zhang and Britten (2011) used microstimulation and found significant biases in choice of perceived heading direction indicating that VIP plays a perceptually causal role for heading discrimination.

Compared to MST studies, only a few studies have been performed in VIP with the purpose of understanding the neural mechanisms of pursuit compensation. We will focus specifically on two studies from the Britten lab that are of particular interest to this topic. Zhang et al. (2004) recorded VIP neuronal responses while simulating motion in a 3D cloud in the presence or absence of smooth eye pursuit movements. On average they found that the cells maintained a stable representation of heading even in the presence of eye pursuit movements. A following study (Zhang and Britten, 2011) evaluated the causal role of VIP in making heading judgments during micro-stimulation. The psychometric curves were generated by asking a trained monkey to judge its heading direction at the end of the visual stimulus. The monkey performed this task after either fixating on a dot during stimulus presentation or performing a smooth eye pursuit movement and during half those trials, VIP was stimulated (interleaved with the control). The first point to notice is that the psychometric curves during fixation and pursuit show very little difference in bias, though there is a slight increase in the thresholds especially for left pursuit. The results of this study showed that the bias of the psychometric functions increased in all three conditions, but the change was larger during the pursuit condition. There were also a larger percentage of cells with significant shifts during pursuit as compared to fixation only (Left pursuit: 77%; Right pursuit: 82%; Fixation: 59%). These results provide further evidence that VIP plays a causal role in

determining heading direction and to some extent in pursuit compensation.

While these studies establish VIP as an important area in its properties of heading discrimination and pursuit compensation, there are some experimental features that raise further questions. Like with the MST studies, these stimuli did not have any stereoscopic information about dot depths. The stimulus represented a 3D cloud, but as a 2D projection at screen depth. Moreover, since the stimulus was viewed binocularly, there is a cue conflict between the depth indicated by the velocity profiles of the dots and the disparity which places all the dots at screen distance. VIP is known to prefer depth planes that are close (Yang et al., 2011), which makes the need for binocular disparity important in presented stimuli. While similar stimuli have been used in psychophysics studies, the stimuli were usually viewed monocularly and some studies have reported better behavioral compensation in the presence of binocular (and to some extent even monocular) depth cues (van den Berg and Brenner, 1994). These properties of VIP and psychophysical perceptions raise the need for using true 3D clouds with binocular disparity to accurately depict a range of depth planes. The psychophysical results obtained from monkeys can also be interpreted as ambiguous when comparing just the different no stimulation conditions. During training, the monkeys receive feedback regarding their choice in the form of a juice reward. Therefore, it is possible that the almost perfect compensation observed during pursuit could be a learned behavior and based not on perception but an artificial shift in behavior in order to maximize the reward. Nonetheless, the difference in the effects due to microstimulation during fixation and pursuit indicate that VIP may play an important role in pursuit compensation.

1.3.3 Experimental Limitations

While these results hint towards MST being involved in pursuit compensation, there

are several experimental factors that could influence the data. The most significant aspect of all three experiments was that they simulated approach towards a FP plane. However, there were no depth cues and even though the simulation was that of an approach, the velocity profile of the stimulus remained constant. This is equivalent to the variable Z from the coordinate system described in Figure 1-1 being held constant. Therefore, the resultant optic flow would simulate approaching a FP wall without actually getting any closer to it. It can be argued that the visual stimulus still accurately represented an instantaneous point during translation + rotation, but then a temporal conflict between eye movement and translation would arise because the FOE is shifting on the retina due to the dynamic perspective cue and is not representative of an instantaneous point in time. This issue can be circumvented based on the assumption that cell responses to a certain depth are constant and the stimulus being presented is not the real-world simulation of an approach towards a FP wall, but instead a plane sliced from a 3D cloud at a fixed distance from the retina.

The larger issue is with regards to the comparison between real and simulated pursuit responses. Both the Bradley et al. (1996) and Shenoy et al. (1999) papers used similar simulated pursuit stimuli. The simulated pursuit condition was accomplished by drifting the entire stimulus across the screen at the same rate as the real pursuit eye movements. However, as described earlier, there are several other features of optic flow that are caused due to eye rotations during translation. So, while this approach would have resulted in very similar FOE shifts compared to real pursuit, the remainder of the optic flow stimulus would not have been accurate – especially farther away from the center. Shenoy et al. argue that this method simulated real pursuit accurately for the range of their small aperture. However, this is precisely the reason a larger stimulus is necessary.

The psychophysics experiments discussed in the previous section hint towards a reweighting of retinal and extra-retinal cues based on the reliability of the cues. If we consider a model for pursuit compensation in MST that uses both retinal and extra-retinal cues, using a small screen and inaccurate flow fields would increase the reliance on extra-retinal cues. While there is no concrete evidence of this yet, the larger compensation seen in Shenoy et al. could also be a result of this cue reweighting. If we assume that in both the experiments, the neurons depended almost exclusively on the extra-retinal signals, then excess visual information in the form of larger screen size could have been added noise. While this may not be the reason for the discrepancies in data, we point this out as a possibility to draw attention to the importance of using real-world accurate stimuli and the present lack of information about the relative roles of retinal and extra-retinal signals in MST. Some support for this theory is evident based on the results of Upadhyay et al. (2000) where they showed that using stimuli that contained motion parallax cues resulted in larger compensation observed in MST. There were limitations to their study as well since they did not have stereoscopic information in the stimulus, but the variability in the responses based on visual stimulus hints towards some influence of the retinal mechanism in solving the rotation problem.

Another point of interest is the use of screen and retinal coordinates to present the data. As shown in Figure 1-4, the FOE is constantly shifting on the retina even after the rotational vectors are subtracted. Therefore, if the data were to be plotted in truly retinal coordinates, the neural data would have to be plotted based on the time course of the pursuit for each corresponding retinal position of the FOE with respect to gaze position. The papers also classify and refer to cells as encoding in retinal versus head/world coordinates. Cells that compensate completely are said to be encoding in head coordinates and cells that don’t, as

encoding in retinal coordinates. However, a lack of compensation or incomplete compensation could simply imply that the heading estimates were erroneous, possibly due to misestimating the rotation velocity and not as a result of using the wrong reference frame. Once again, if the cells truly encoded in retinal space, the responses of the cells during the course of the stimulus would vary temporally based on relative gaze positions. Since there is no data printed showing the time course of MST firing rates during pursuit trials, it is not possible to deduce if some of the cells do in fact encode the heading direction in retinal coordinates. This is an important distinction to make because if it can be shown that both the compensatory and the non-compensatory cells encode in head coordinates, the errors and shifts in the responses can be attributed directly the cells' ability to accurately subtract rotational velocities and will bring us a step closer to addressing the question about the relative roles of retinal and extra-retinal signals.

Furthermore, the neurophysiological data recorded in most of the pursuit studies use stimuli that are limited to ~30-40° around straight ahead. This raises some problems in understanding the exact mechanism used by the cells to compensate for pursuit. With a limited range of the tuning curve available, it is difficult to distinguish gain changes from lateral shifts in the curves, which could be resolved by sampling across the whole horizontal plane. The differences in reported compensations in MST when only the horizontal plane is sampled versus a larger part of the 3D space, further emphasizes the need for analyzing the effects of pursuit on the full 3D tuning curve.

The results from the MST studies showing poor compensation during simulated pursuit seem to support the initial psychophysical results that showed very poor performance during simulated pursuit with a FP wall. These results clearly show that in the absence of

reliable visual stimuli, both neurometric and psychometric predictions rely heavily on extra-retinal signals. However, later psychophysical studies have shown that increasing the information present in the visual stimulus as well as reducing the cue conflict between extra-retinal and retinal signals during simulated pursuit improves the accuracy of heading perception in humans. The only evidence of such an improvement is from the larger compensations seen in MST due to the addition of motion parallax cues (Upadhyay et al., 2000). However, electrophysiological studies of pursuit compensation have yet to use more complex visual stimuli to study the possible reweighting of retinal and extra-retinal cues. Another significant difference between the psychophysical and electrophysiological studies is the pursuit speeds used in the two sets of studies. While the largest pursuit speeds used in most psychophysics experiments is around 5°/s, the electrophysiological studies range from 10-18°/s. This, in addition to the fact that the translation speeds used in psychophysics are generally much larger than the translation speeds in electrophysiology, makes the direct comparison of the two results difficult.

1.4 Summary of Aims

Evaluation of both psychophysical and electrophysiological literature on navigation based on optic flow reveals shortcomings in our knowledge that need to be addressed to understand how the brain deciphers self-motion. First, there is no conclusive evidence that the visual system can use purely retinal cues to estimate translations and rotations from optic flow. However, based on the limitations of extra-retinal cues and the evidence from psychophysical results, we hypothesize that the visual system can indeed use retinal cues to decompose self-motion into translations and rotations. Second, both psychophysical and

electrophysiological studies have neglected the importance of estimating and rotations jointly representing both translation and rotation during self-motion. Thus, in the following studies, we will evaluate the roles of retinal and extra-retinal cues in estimating and jointly representing T and R.

1.4.1 Aim 1: Rotation-invariant heading representation based on visual and non-visual cues.

In the first study, we evaluated how translations are represented in the presence of rotations in the macaque VIP based on visual stimuli. We presented head-fixed monkeys with various optic flow stimuli simulating translations in the horizontal plane. During the 'real pursuit' (RP) condition, the translations were accompanied by the monkey performing smooth pursuit of a moving target, thus resulting in both retinal and extra-retinal cues. During the 'simulated pursuit' (SP) condition, both translations and rotations were simulated, resulting in only retinal cues.

We found the first conclusive evidence of neurons representing rotation-invariant translations based purely on optic flow (retinal cues). Specifically, we found that 24% of neurons during SP (30% during RP) were invariant to added rotations. However, the mean compensation during RP was larger than SP, indicating that while retinal cues play a significant role in extracting translations, extra-retinal cues also contribute to the computation. We further evaluated the role of 3D and 2D visual cues using both a 3D cloud stimulus and a FP plane. The results showed for the first time that 2D visual information such as dynamic perspective cues, can also be used to separate translations from rotations.

This rotation-invariant representation of translations based on retinal cues provides evidence for the hypothesis that retinal cues are sufficient to decompose self-motion into its

translational and rotational components and allowed us to further evaluate the joint representation of T and R components.

1.4.2 Aim 2: VIP neurons jointly encode translations and rotations.

Effective navigation requires estimates of both translations and rotations. Hence, in this study we evaluated the representation of rotations in the presence of translations as well as the joint representation of both at the single neuron level in area VIP. We used visual stimuli similar to that in Aim 1, but included multiple rotation velocities to probe rotation representations.

We found that about 43% of VIP neurons encoded both translations and rotations. The rotation representation in these neurons was in the form of gain fields of translation tuning curves. We also found that the level of rotation compensation observed in a neuron did not influence the presence of rotation tuning, implying that rotations are represented in area VIP regardless of the accuracy of the optic flow decomposition. Importantly, this is the first study to show that single neurons in VIP jointly represent both translations and rotations. This supports our hypothesis that the visual system encodes translations and rotations in a flexible manner such that either component or a combination of both can be easily decoded.

Chapter 2 – Role of visual and non-visual cues in constructing a rotation-invariant representation of heading in parietal cortex

2.1 Summary

As we navigate through the world, eye and head movements add rotational velocity patterns to the retinal image. When eye/head rotations accompany observer translation, the rotational velocity patterns must be discounted to accurately perceive heading. The conventional view holds that this computation requires efference copies of self-generated eye/head movements. Here we demonstrate that the brain implements an alternative solution in which retinal velocity patterns are themselves used to dissociate translations from rotations. These results reveal a novel role for visual cues in achieving a rotation-invariant representation of heading in the macaque ventral intraparietal area. Specifically, we show that the visual system utilizes both 3D cues (motion parallax) and 2D perspective distortions to estimate heading in the presence of rotations. These findings further suggest that the brain is capable of performing complex computations to infer eye movements and discount their sensory consequences based solely on visual cues.

2.2 Introduction

Retinal images of the environment are altered by self-generated rotations such as eye or head movements. In order to perceive the world accurately, the component of retinal patterns resulting from such rotations needs to be discounted by the visual system. How the brain achieves such a rotation-invariant visual representation of the world remains unclear. Visually guided navigation is an important context in which achieving rotation-invariance is critical for accurate behavior (Gibson, 1950; Warren and Saunders, 1995; Grigo and Lappe, 1999). For example, while walking down a sidewalk and simultaneously looking at a passing car using eye or head rotations, the brain must discount the visual consequences of self-generated rotations to estimate and maintain one's direction of translation (i.e., heading).

Self-motion results in retinal velocity patterns known as 'optic flow' (Gibson, 1950). During translations, the resulting retinal pattern is generally an expansionary or contractionary radial flow field from which the point of zero velocity (Focus of Expansion, FOE) can be used to estimate heading (Tanaka et al., 1986; Warren et al., 1988; Duffy and Wurtz, 1995; Britten, 2008). However, eye or head rotations alter this flow pattern such that deciphering heading requires decomposing the resultant optic flow into translational and rotational components (Figure II-1a). Psychophysical (Royden et al., 1992; Royden, 1994; Crowell et al., 1998) and electrophysiological (Bradley et al., 1996; Page and Duffy, 1999; Zhang et al., 2004) studies have often emphasized the role of non-visual signals, such as efference copies of self-generated eye/head movements, in discounting rotations to estimate heading. Such non-visual signals can represent several different sources of rotation, including eye-in-head (R_{EH}), head-on-body (R_{HB}), and body-in-world (R_{BW}) movements (Figure 2-1b). Critically, retinal image

motion is determined by the translation and rotation of the eye relative to the world (T_{EW} and R_{EW}, Figure 2-1b), such that extracting heading from optic flow requires compensating for the total rotation of the eye-in-world (where, $R_{EW} = R_{EH} + R_{HB} + R_{BW}$). Therefore, in general, multiple non-visual signals would need to be added to achieve a rotation-invariant estimate of heading, potentially compounding the noise that is associated with each signal (Gellman and Fletcher, 1992; Li and Matin, 1992; Crowell et al., 1998).

Alternatively, rotation-invariance can also theoretically be achieved exclusively through visual processing (Longuet-Higgins and Prazdny, 1980; Rieger and Lawton, 1985). If the brain can use optic flow to directly estimate and discount rotations of the eye-in-world (R_{EW}), such mechanisms may provide a complementary and potentially more efficient way to decompose rotations and translations to achieve invariant heading perception. Psychophysical studies have provided evidence that visual cues may play a role in estimating heading in the presence of rotations (Grigo and Lappe, 1999; Li and Warren, 2000; Crowell and Andersen, 2001; Li and Warren, 2002, 2004; Royden et al., 2006). However, electrophysiological evidence for the role of visual cues is ambiguous, in part because previous neurophysiological studies either did not include visual controls for eye rotation (Zhang et al., 2004), used incorrect visual stimuli to simulate rotations (Bradley et al., 1996; Shenoy et al., 1999; Shenoy et al., 2002) or employed inappropriate analysis methods (Bradley et al., 1996; Shenoy et al., 1999; Shenoy et al., 2002; Bremmer et al., 2010 see Discussion; Kaminiarz et al., 2014).

We recorded neural activity from the macaque ventral intraparietal area (VIP) to evaluate the relative roles of visual and non-visual cues in computing heading in the presence of rotations. To elucidate the role of visual cues, we accurately simulated combinations of

translations and rotations using visual stimuli containing a variety of cues present during natural self-motion. Our results provide novel evidence that (1) a subpopulation of VIP neurons utilizes visual cues to signal heading in a rotation-invariant fashion and (2) both 3D visual cues (motion parallax) and 2D cues (perspective distortions) present in optic flow contribute to these computations. In addition, we find that visual and non-visual sources of rotation elicit similar responses in VIP, suggesting multi-sensory combination of both cues in representing rotations. We further show that rotation-invariance is distinct from the reference frame used to represent visual heading, and provide additional support for an eye-centered representation in VIP (Chen et al., 2013).

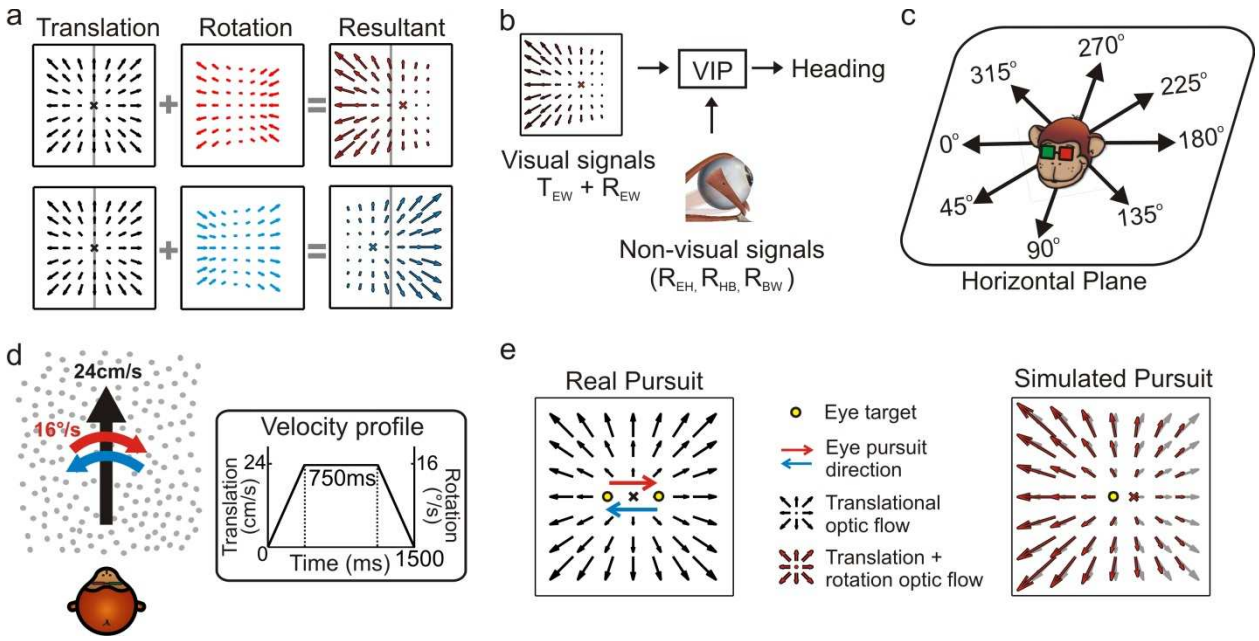

Figure 2-1. Figure 1. Dissociating translations and rotations, and experimental approaches. (a) Optic flow patterns during self-motion. Forward translations result in symmetric flow patterns (black vector fields) with a focus of expansion (FOE) indicating heading. When rotations are added to translations, the resultant optic flow pattern has an FOE shift in the direction of the added rotation (rightward rotation: red, leftward rotation: blue). (b) VIP receives both visual and non-visual signals that may be used to achieve rotation-invariant heading estimates. Visual optic flow signals contain information about translation and rotation of the eye in the world (T_{EW}, R_{EW}) whereas non-visual signals (efference copies) may contain information about rotation of eye-in-head (R_{EH}), rotation of head-on-body (R_{HB}), or rotation of body-in-world (R_{BW}). (c) Visual stimuli simulating translations in 8 directions spanning the entire horizontal plane were presented to the monkey. (d) Schematic showing the translation and rotation parameters in the simulated 3D cloud. Inset shows the trapezoidal velocity profile of translation and rotation during the course of a trial (1500ms). (e) During the 'Real pursuit (RP)' condition, the optic flow stimulus simulated translation, while rotation was added by having the monkey smoothly pursue a visual target that moved leftward or rightward across the screen. During the 'Simulated pursuit (SP)' condition, the monkey fixated at the center of the display while optic flow simulated combinations of translation and eye rotation. During real and simulated pursuit, the optic flow patterns projected onto the monkey's retina were nearly identical.

2.3 Results

To investigate the effect of rotations on the visual heading tuning of VIP neurons, we presented visual stimuli simulating 8 directions of translation in the horizontal plane (Figure 2-1c) and two directions of rotation (Figure 2-1d). To evaluate the relative roles of visual and non-visual cues, rotations were introduced in the form of either 'real' or 'simulated' pursuit eye movements. During real pursuit (RP, Figure 2-1e, left), the monkey smoothly tracked a target moving across the screen such that both visual and non-visual rotation cues were present. During simulated pursuit (SP, Figure 2-1e, right), the visual motion stimulus accurately simulated a combination of translation and eye rotation while the monkey fixated a stationary target at the center of the display (non-visual cues were absent). In order to provide a rich visual environment, the first experiment simulated self-motion through a 3D cloud of dots, a stimulus that contains both motion parallax (Helmholtz, 1924; Longuet-Higgins and Prazdny, 1980; Koenderink and van Doorn, 1987; Tomasi and Shi, 1996) and 2D image deformation cues (Koenderink and van Doorn, 1976; Grigo and Lappe, 1999). To further explore the underpinnings of a retinal solution in achieving rotation-invariance, a second experiment used a fronto-parallel plane (FP) of dots, which contains only 2D image deformation cues.

2.3.1 Analysis of the effects of rotation on optic flow

When rotation and translation occur simultaneously, the resulting pattern of retinal velocity vectors can differ substantially from the typical radial optic flow patterns observed during pure translation. This change is often conceptualized as a shift in the focus of expansion (FOE) (Warren and Hannon, 1990; Bradley et al., 1996; Shenoy et al., 1999;

Shenoy et al., 2002). However, in a visual scene with depth structure, adding rotation results in different FOE shifts at different depths (Zhang et al., 2004). This is due to a key difference in the properties of optic flow resulting from translations and rotations – the magnitude of translational optic flow vectors decrease with distance (depth), whereas rotational optic flow vectors are independent of depth (Longuet-Higgins and Prazdny, 1980). Hence, for more distal points in a scene, rotations produce a larger FOE shift (Figure 2-2). For the translation

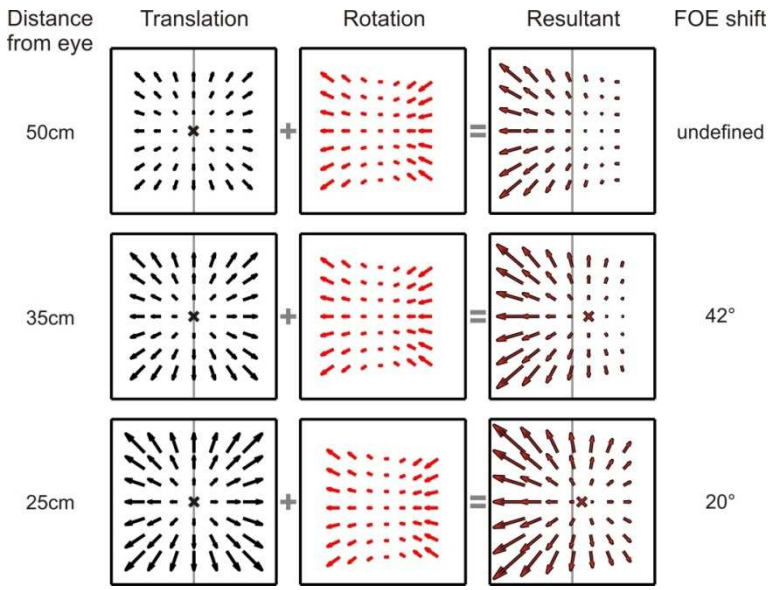

Figure 2-2. Dependence of translational and rotational optic flow properties on viewing distance. Translational optic flow vectors (left column) decrease in magnitude as the distance of the plane being viewed increases. Rotational optic flow (middle column), however, remains constant irrespective of the viewing distance. When these translation and rotation flow fields are added, the resultant FOE shift varies with distance to the plane (right column). Therefore, in a 3D environment where objects are present at varying distance from the observer, no single FOE exists. For the stimulus parameters used in this study, the nearest depth plane of the simulated 3D cloud (25cm) results in a 20° shift in FOE; at the screen depth of 35cm, the shift is 42° and for any plane beyond 50cm (50-125cm), the FOE is undefined as the optic flow is dominated by rotations.

and rotation parameters used in this study, the nearest plane in the 3D cloud (25cm) results in a 20° FOE shift. However, for any plane farther than 50cm, the resultant optic flow has an undefined FOE (Figure 2-2, top row). The simulated 3D cloud ranged from 25cm to 125cm, resulting in a large volume of the stimulus space having undefined FOE shifts. Since FOE shift is an ill-defined measure of the visual consequence of rotations, we simply refer to the net visual stimulation associated with simultaneous translation and rotation as the 'resultant optic flow'.

Forward translations result in an expansionary flow field, for which adding a rightward rotation causes a rightward shift of the focus of expansion (for any given plane). On the other hand, backward translations produce a contractionary flow field and adding a rightward rotation results in a leftward shift in the focus of contraction (Figure 2-3a). If a neuron signals heading regardless of the presence of rotations, then its tuning curves during real and simulated pursuit should be identical to the heading tuning curve measured during fixation (Figure 2-3b). For a neuron that instead represents the resultant optic flow rather than the translation component (heading), a transformation of the tuning curve is expected due to the added rotations. As a result of the opposite shifts expected for forward (expansionary flow field) and backward translations (contractionary flow field), the heading tuning curve of a neuron preferring forward headings would have a peak that shifts to the right and a trough that shifts to the left during rightward eye rotation; together, these effects cause a skewing of the tuning curve (Figure 2-3c, red curve). For the same neuron, leftward eye rotation would cause the peak to shift to the left and the trough to shift to the right, thus having an opposite effect on the shape of the tuning curve (Figure 2-3c, blue curve). Neurons that prefer lateral headings, which are common in VIP (Chen et al., 2011), may in fact, show no shift in the

peak. But, since opposite shifts are expected for forward and backward headings, the resulting tuning curve may exhibit substantial bandwidth changes (Figure 2-3d).

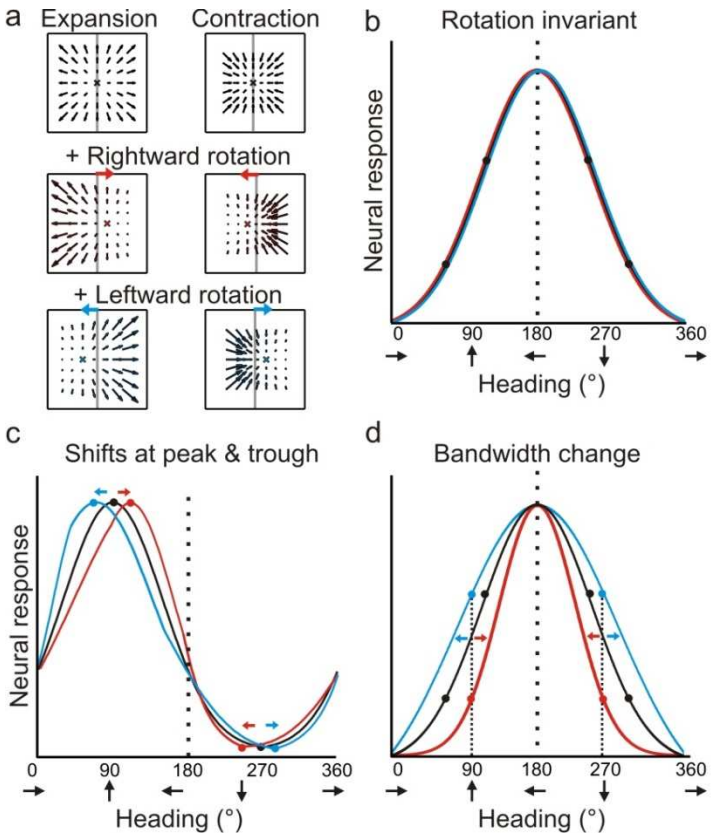

Figure 2-3. Predicted transformations of heading tuning curves due to rotations. (a) Forward and backward translations result in expansion and contraction flow fields, respectively. Adding rotation causes the FOE to shift in opposite directions for forward and backward translations (rows 2, 3). (b, c, d) Hypothetical heading tuning curves show the expected transformations due to rotations (rightward, red; leftward, blue). (b) Schematic illustration of rotation-invariant heading tuning curves. (c) Schematic representing a cell that responds to resultant optic flow (no rotation tolerance) with a heading preference of straight ahead (90°). Rightward rotation causes a rightward shift of the tuning curve for forward headings (0-180º), and a leftward shift for backward headings (180-360º). The opposite pattern holds for leftward rotations. The net result of rotation is a skewing of the tuning curve. (d) Schematic tuning of a cell with a leftward heading preference (180°) and no rotation tolerance. Here, the tuning bandwidth increases for leftward rotations and decreases for rightward rotations.

Under the null hypothesis that neural responses are simply determined by the resultant optic flow, the expected effect of rotation on heading tuning is not simply a global shift of the tuning curve, as was assumed previously (Bradley et al., 1996; Page and Duffy, 1999; Shenoy et al., 1999; Shenoy et al., 2002; Bremmer et al., 2010; Kaminiarz et al., 2014). Further illustrations of the expected effects of rotation for hypothetical neurons with different heading preferences are shown in Figure 2-4. We designed our quantitative analysis of heading tuning curves specifically to account for these previously unrecognized complexities (*Methods*).

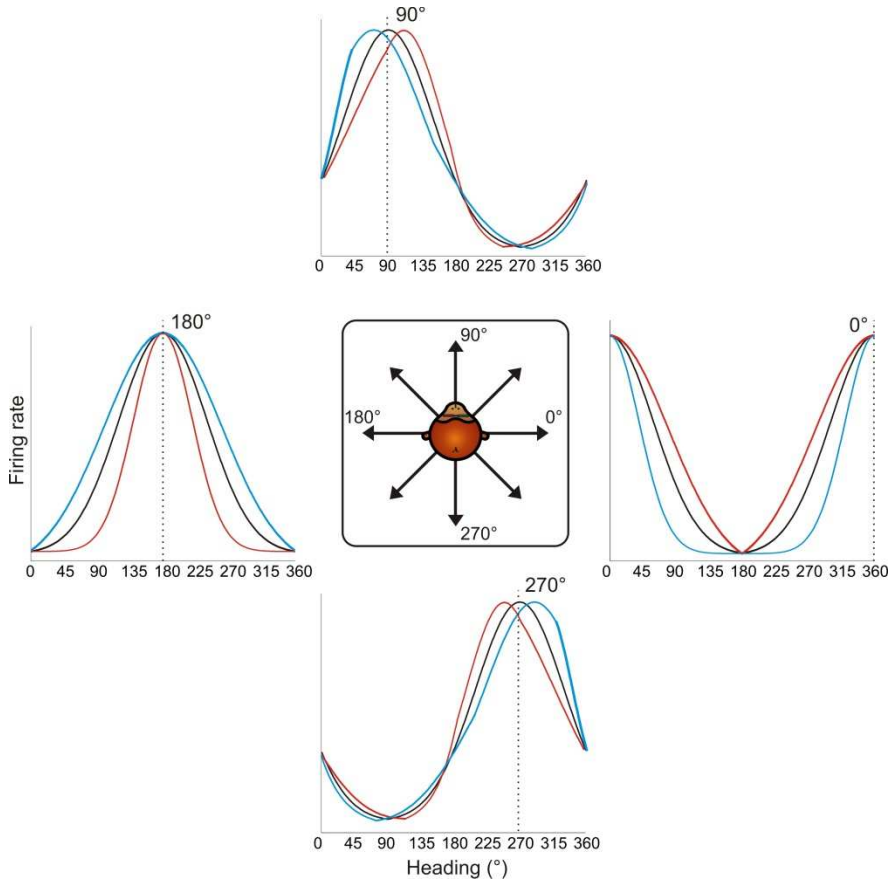

Figure 2-4. Schematic showing tuning curve transformations for hypothetical neurons with different heading preferences. For cells that prefer lateral headings (0°, 180°), rotations cause changes in tuning bandwidth. The expected change in bandwidth is opposite for cells preferring 0° and 180°. For cells preferring forward or backward motion, rotations cause opposite directions of shifts in the peak and trough of the tuning curve, thus changing the shape of the tuning curve.

2.3.2 Influence of visual and non-visual cues on heading representation in VIP

Heading tuning curves (translation only) can be compared to real pursuit (RP) and simulated pursuit (SP) tuning curves (translation + rotation) to evaluate whether a VIP neuron signals heading invariant to rotations (Figure 2-3b), or whether it simply responds to the resultant optic flow (Figure 2-3c, d). Figure 2-5a shows heading tuning curves for an example neuron during pure translation (black curve), as well as during rightward (red) and leftward (blue) rotations added using RP and SP conditions. The tuning curves in this example show only minor changes during RP indicating that the cell signals heading in a manner that is largely invariant to eye rotation, consistent with previous findings for real eye rotation (Zhang et al., 2004). Interestingly, the tuning curves of the same neuron during SP also change very little, showcasing the role of visual signals in compensating for rotation. Thus, effects in VIP that were previously attributed to non-visual signals (Zhang et al., 2004) might also be driven by visual cues.

Data for another example VIP neuron (Figure 2-5b) reveal RP tuning curves that are also largely consistent in shape with the pure translation curve, but have larger response amplitudes during leftward pursuit. During simulated pursuit, however, the tuning curves of this neuron show clear bandwidth changes. Thus, this second example neuron appears to rely more on non-visual cues to discount rotations. Note that this example neuron preferred lateral headings (leftward) and showed large bandwidth changes during SP, as predicted in the schematic illustration of Figure 2-3d. Such bandwidth changes were observed consistently among VIP neurons that preferred lateral translations; specifically, rightward rotations increased bandwidth for cells preferring rightward headings (~0°) and decreased bandwidth for cells preferring leftward headings (~180°), with the opposite pattern holding for leftward

rotations (Figure 2-6).

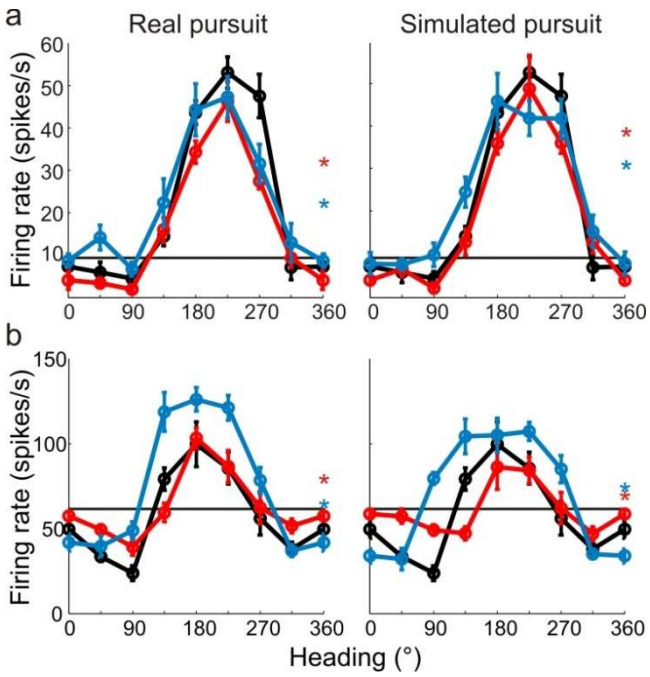

Figure 2-5. Heading tuning curves from two example VIP neurons. Five tuning curves were obtained per cell: 1 pure translation curve (black), 2 real pursuit (RP, left column) curves, and 2 simulated pursuit (SP, right column) curves (rightward rotation: red, leftward rotation: blue). Black horizontal line indicates baseline activity. Red and blue stars in the left column indicate responses during pursuit in darkness, and in the right column indicate responses to simulated eye rotation. (a) This neuron has largely rotation-invariant tuning curves in both RP and SP conditions (shifts not significantly different from 0, CI from bootstrap), and has significant rotation responses during both pursuit in darkness and simulated rotation (compared to baseline; Wilcoxon signed rank test $p<0.05$). (b) This example neuron shows significant bandwidth changes during SP (shifts > 0°, CI from bootstrap), similar to the prediction of Figure 2-3d. Of the rotation-only conditions, the cell only responds significantly during rightward pursuit in darkness (Wilcoxon signed-rank test $p = 0.01$).

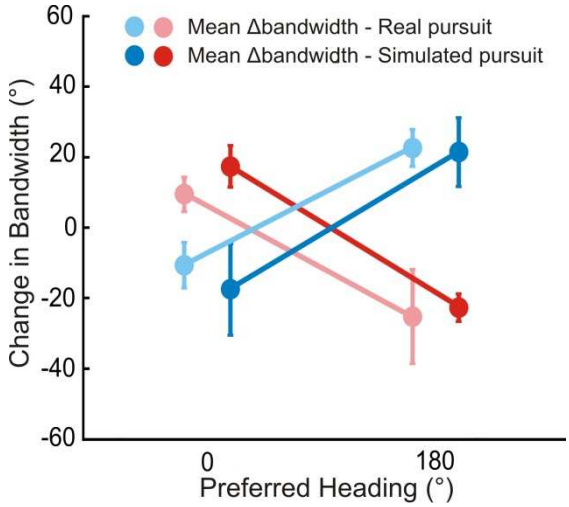

Figure 2-6. Bandwidth changes observed in data. Bandwidths of linearly interpolated tuning curves were calculated as the full width at half height (FWHH). The difference in FWHH between the pure translation and rotation-added tuning curves (reds: rightward, blues: leftward rotation) are plotted for cells with lateral heading preferences (since the largest bandwidth changes would occur for cells preferring lateral motion). The bandwidth changes observed are in the directions predicted by Figure 2-3, 2-4. The change in bandwidths at 0° and 180° headings are significantly different for both RP (Wilcoxon rank sum test; leftward, rightward: $p < 0.001$) and SP (Wilcoxon rank sum test; leftward, rightward: $p < 0.001$) conditions.

Because of these changes in tuning curve bandwidth or shape, analysis of the effects of rotation on heading tuning requires more complex and rigorous approaches (Figure 2-7) than the cross-correlation or rank-order methods used in previous studies (Bradley et al., 1996; Shenoy et al., 1999; Shenoy et al., 2002; Bremmer et al., 2010; Kaminiarz et al., 2014). It is also critical to distinguish between changes in response gain and changes in the shape (Figure 2-8) of tuning curves, which our analysis allows because we sample the entire heading tuning curve (Mullette-Gillman et al., 2009; Chang and Snyder, 2010; Rosenberg and Angelaki, 2014). As shown in Figure 2-7, the first step in the analysis involves normalizing each RP and SP tuning curve to match the dynamic range of the pure translation tuning curve. Following this transformation, the change in the shape of the RP and SP tuning curves can be measured without ambiguity. To account for the expected changes in bandwidth and skew, partial shifts of the tuning curve were measured separately for forward (0°:180°) and backward (180°:360°) headings. Thus, 4 shift values were obtained from each neuron for both real and simulated pursuit, corresponding to forward/backward headings and left/right rotation directions. These 4 values were averaged for each neuron to quantify the transformation in shape and obtain one shift metric for RP tuning curves and one for SP tuning curves (see *Methods*, Figure 2-7).

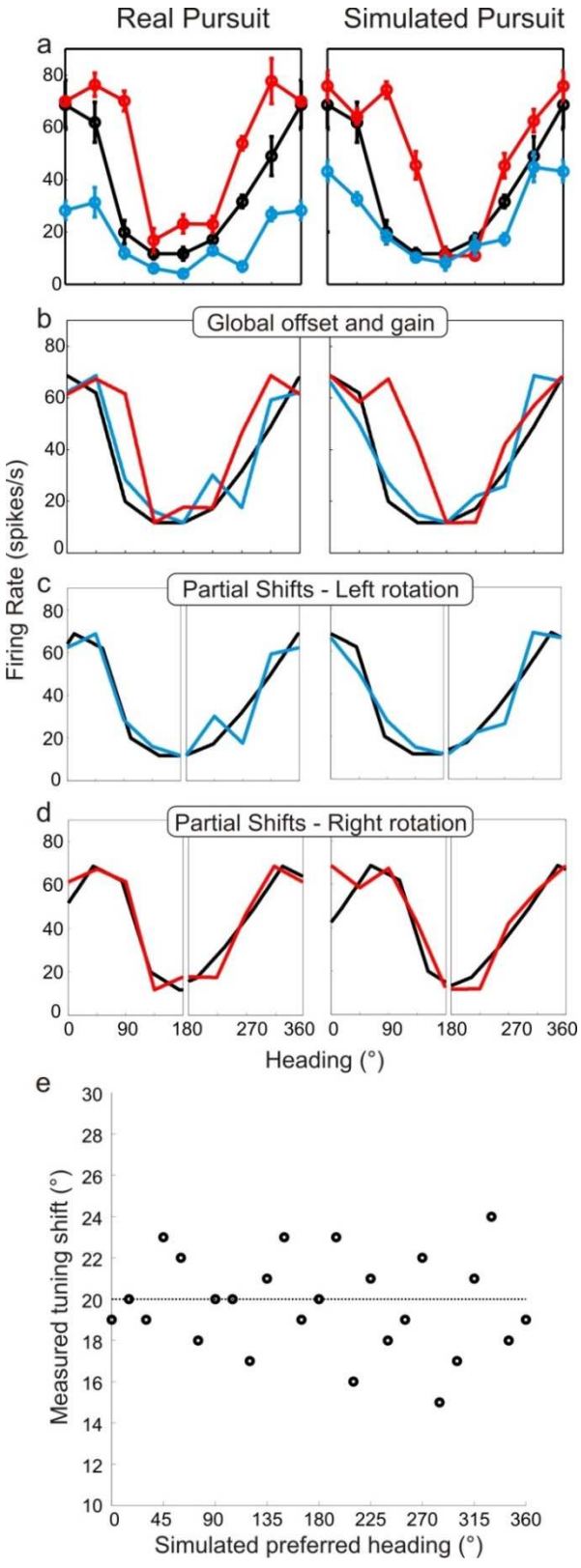

Figure 2-7. Method for analyzing tuning curve shifts. (a) Example tuning curves of a single neuron for RP and SP conditions. Black - pure translation; red - rightward rotation; blue - leftward rotation. (b) First, the offset and gain of the RP/SP tuning curves are corrected to match the offset and gain values of the pure translation tuning curve. (c,d) To account for bandwidth changes in the shift calculations, the RP/SP tuning curves are split into halves corresponding to forward (0:180°) and backward (180:360°) headings. The pure translation tuning curve is then circularly shifted to minimize the sum squared error with both halves of the leftward rotation (c) and rightward rotation tuning curves (d). This yields four shift values (shown in c, d) each, for SP and RP, which are averaged. (e) We simulated noisy neuronal tuning curves that have different response amplitudes, offsets and bandwidth/shape changes similar to real data (see Methods). The expected shifts for the simulated data should correspond to 20°. The mean shift is not significantly different from 20° (t-test; p=1), indicating that our analysis method correctly extracts tuning curve changes despite variations in shape, response amplitude, or baseline response.

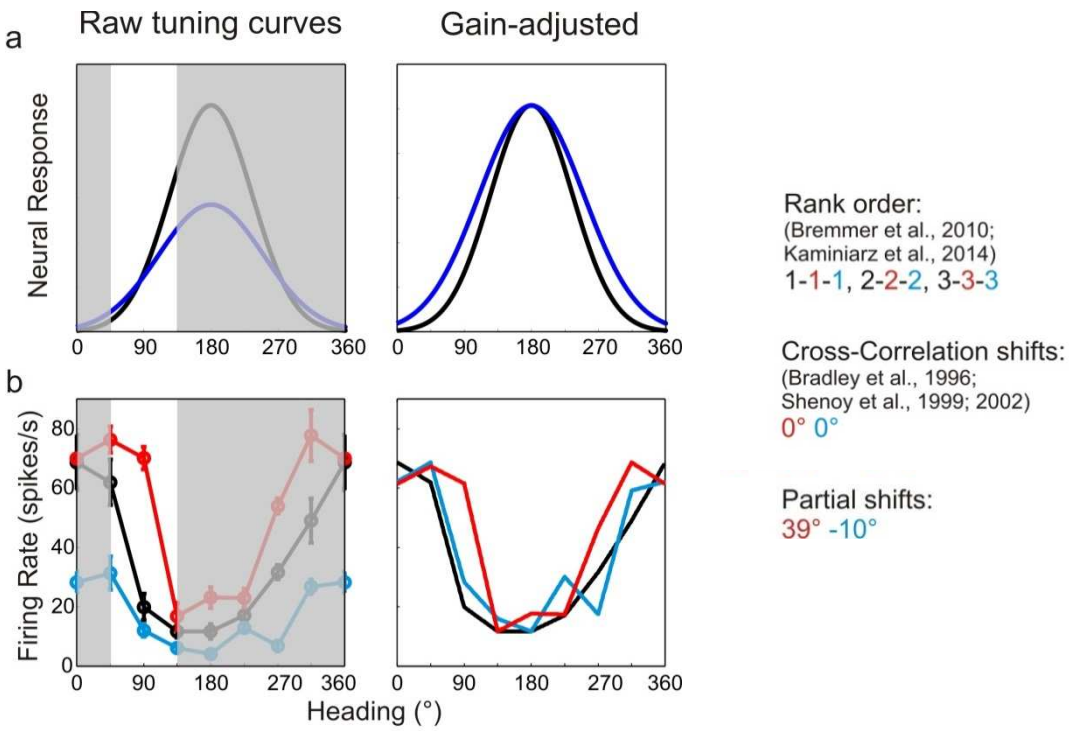

Figure 2-8. Ambiguity between shifts and gain fields in the absence of full tuning curve. Previous studies (Bradley et al., 1996; Shenoy et al., 1999; Shenoy et al., 2002; Bremmer et al., 2010; Kaminiarz et al., 2014) evaluated heading tuning in a narrow range around straight ahead (white region around 90° in the left column). (a) Two hypothetical tuning curves with different bandwidths and amplitudes (left column). Correcting for the difference in response amplitudes (right column) reveals a difference in bandwidths. But, a cross-correlation analysis between the portions of the tuning curves around straight ahead (white region), would erroneously report no shift (Bradley et al., 1996; Shenoy et al., 1999; Shenoy et al., 2002). The rank order of the responses would be identical for the two curves (largest for leftward and smallest for rightward headings), which can be erroneously interpreted as rotation-invariance (Bremmer et al., 2010; Kaminiarz et al., 2014). (b) Tuning curves from an example VIP neuron (same cell as Figure 2-7). The changes in tuning bandwidth are missed by the cross-correlation method and misinterpreted as gain changes resulting in 0° shifts. Rank ordering the responses reveals that all three tuning curves prefer a heading of 45°, resulting in rank orders [1-1-1, 2-2-2, 3-3-3] which indicates no shift. Our partial shifts method reveals a large significant shift during rightward rotation (red curve).

Results are summarized for the population of recorded neurons (n=72; from two monkeys) in Figure 2-9. A shift of 0° implies that the neuronal representation of translation is invariant to rotation (i.e., the shape of heading tuning curves are highly similar, as in Figure 2-5a). A positive shift indicates under-compensation for rotation, such that responses change in a manner consistent with the resultant optic flow. Negative shifts indicate that the tuning curve

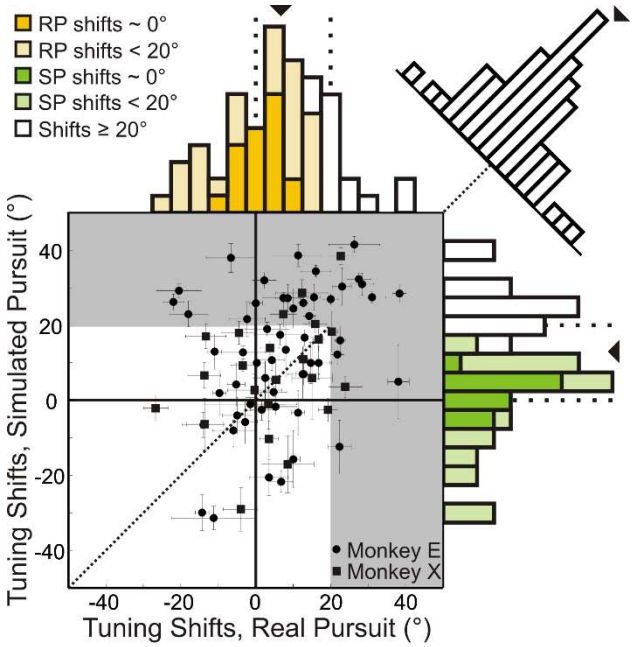

Figure 2-9. Scatterplot and marginal distributions of shifts measured during real pursuit (RP) and simulated pursuit (SP) using 3D cloud stimuli (n = 72 cells). A shift of 0° indicates rotation-invariance. Positive and negative shifts indicate under-compensation and over-compensation for rotation, respectively. Grey shaded area corresponds to shifts ≥ 20° (conservative estimate of no tolerance to rotations). Error bars depict bootstrapped 95% confidence intervals (CI). Colored regions of marginal distributions indicate shifts significantly < 20°. Darker colors indicate shifts not significantly different from 0°. Uncolored histograms indicate shifts ≥ 20°. Diagonal histogram shows difference in RP and SP shifts for each neuron with a median of -6.0° indicating that for most cells SP shifts tended to be larger than RP shifts (significantly < 0°; Wilcoxon signed-rank test p = 0.02).

transformation was in the direction opposite to that expected based on the resultant optic flow. This can be interpreted as an over-compensation for rotation. As noted earlier, though the FOE shift for the nearest depth plane (25cm) in our stimuli is 20°, a majority of the cloud volume (50-125cm deep) is dominated by rotations, such that the resultant optic flow has undefined FOEs. This implies that neurons should show shifts that are generally much larger than 20° if they do not discount the rotations and merely represent the resultant optic flow.

In the RP condition, 22/72 (30.6%) neurons showed shifts that were not significantly different from zero (bootstrap 95% CI); these cells can be considered to represent heading in a rotation-invariant fashion. For SP, 17/72 (23.6%) neurons had shifts that were not significantly different from zero, indicating that purely visual cues were sufficient to achieve rotation-invariance in these neurons. Only 12/72 (16.7%) neurons during RP and 19/72 (23.4%) neurons during SP showed shifts that were significantly greater than 20°, suggesting that only a minority of VIP neurons simply represent the resultant optic flow.

The median shift of the population during RP is 8.5°, which is significantly less than the 13.8° median shift observed during SP (Wilcoxon signed-rank test; $p = 0.02$), indicating greater tolerance to rotations in the presence of both non-visual and visual cues. However, both median shifts are significantly greater than 0° (Wilcoxon signed-rank test; $p < 0.001$), and less than 20° (Wilcoxon signed-rank test; RP: $p<0.001$, SP: $p = 0.005$) suggesting that, on average, VIP neurons do not simply represent the resultant optic flow, but rather signal heading in a manner that is at least partially tolerant to rotations. Together, these findings indicate that VIP can signal heading in the presence of rotations using both visual and non-visual cues. Importantly, this tolerance to rotations is observed even when only visual cues are present (SP).

2.3.3 Visual and non-visual rotation signals in VIP

The previous section shows that VIP neurons can use visual cues to signal heading in the presence of rotations, but it is unclear if the rotational component is also represented. During real pursuit, the rotation arises from a movement of the eye relative to the head. In this case, both non-visual and visual sources of information about the rotation are available. These two sources of information differ in that the non-visual source signals the rotation of the eye relative to the head (R_{EH}) and the visual source signals the rotation of the eye relative to the world (R_{EW}). Previous studies have shown that VIP receives efference copies of pursuit eye movements (Colby et al., 1993; Duhamel et al., 1997), reflecting an R_{EH} signal. However, no previous studies have tested if VIP also carries an R_{EW} signal based on visual rotation information present in optic flow.

To test whether neurons in VIP signal rotations based on both non-visual and visual cues, we analyzed data from interleaved rotation-only trials (leftward and rightward rotations) in which the monkey either pursued a target in darkness (non-visual R_{EH} signal) or fixated centrally while the visual stimulus simulated a rotation (visual R_{EW} signal) with the same velocity profile as pursuit in darkness. We found that about half of the rotation responses were significantly different from baseline activity during both real and simulated rotations (144 responses from 72 cells; 73/144, 50.7% during pursuit in darkness and 78/144, 54.2% during simulated rotation).

In our experiments, the R_{EW} signal is equivalent to the R_{EH} signal since only eye rotations are considered. Therefore, similarity between the efference copy signal (R_{EH}) and the neural responses to purely visual rotation stimuli (R_{EW}) would suggest the presence of an integrated (visual and non-visual) R_{EW} signal in VIP. We find that the baseline-subtracted responses to

these two types of rotation stimuli are significantly correlated (rightward rotation: Spearman r = 0.50, $p < 0.001$; leftward rotation: Spearman r = 0.39; p = 0.001), supporting the presence of a rotation signal derived from purely visual cues (R_{EW}) in area VIP (Figure 2-10a). Furthermore, the difference in response between rightward and leftward rotations (Figure 2-10b) shows that many VIP neurons exhibit direction-selective responses to rotation. We also find significant correlation between the differential responses (left − right rotation) during real and simulated rotation (Spearman r = 0.59; $p < 0.001$). These results support the hypothesis of multi-sensory convergence of visual and non-visual cues to provide consistent rotation

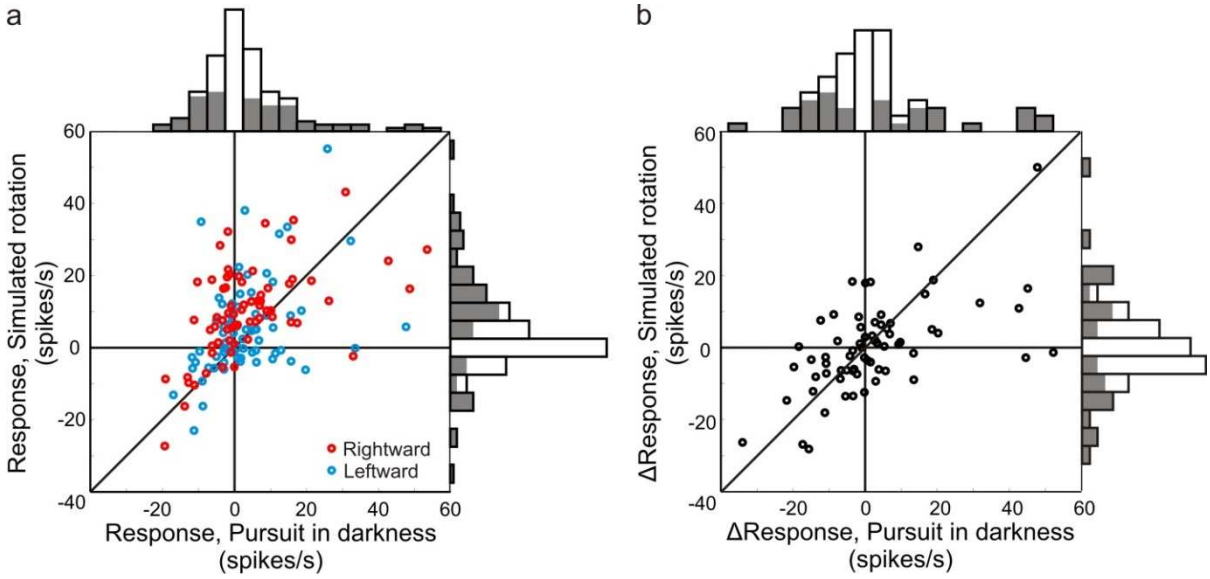

Figure 2-10. Neural responses to pure rotation stimuli. (a) Scatterplot and marginal distributions of baseline-subtracted rotation responses. The monkey either pursued a target across a dark screen (pursuit in darkness) or fixated centrally as rotation was simulated in the 3D dot cloud (simulated rotation). Filled marginal distributions indicate significant rotation responses compared to baseline (t-test, $p \leq 0.05$). Red and blue symbols denote rightward and leftward rotations, respectively. (b) Scatterplot of differences between leftward and rightward rotation responses. Filled marginal distributions indicate significant differences between leftward and rightward rotation responses (t-test, $p \leq 0.05$).

information, which may be critical for encoding rotations, in addition to achieving a rotation-invariant representation.

It is important to note that, in general, retinal motion corresponding to R_{EW} is a combination of R_{EH}, rotation of the head-on-body (R_{HB}), and body-in-world (R_{BW}). And each of these different rotations will be accompanied by different efference copy (non-visual) signals. If VIP neurons represent R_{EW} based on non-visual signals, then they would have to represent a combination of all efference copy signals: $R_{EW} = R_{EH} + R_{HB} + R_{BW}$. Although we cannot test this directly with our data, the correlations observed in Figure 2-10 allow for the possibility that VIP neurons represent R_{EW} based on both visual and non-visual cues.

2.3.4 Role of perspective distortions in achieving rotation-invariance

Results from the 3D cloud experiment (Figure 2-9) demonstrate, for the first time at the neural level, a clear contribution of visual cues in achieving a rotation-tolerant representation of heading. To gain a deeper understanding of the visual mechanisms involved in dissociating translations and rotations, we investigated which optic flow properties are used by the visual system to infer self-motion from visual cues. Gibson (1950) suggested that motion parallax information can be used to dissociate translations from rotations in a 3D scene. Motion parallax refers to the difference in retinal velocities of points that have different depths but similar retinal locations. As illustrated in Figure 2-2, translational optic flow vectors are dependent on depth, whereas rotational vectors are not. As a result, the difference between optic flow vectors at different depths for a rich 3D environment results in a motion parallax field for which the rotational component has been subtracted away (Longuet-Higgins and Prazdny, 1980; Rieger and Lawton, 1985; Warren and Hannon, 1990), and the point of zero motion parallax corresponds to the heading (Figure 2-11a). However, this solution requires

rich depth structure in the scene, which is not always present. For instance, walking through a dense forest provides robust motion parallax cues, but walking towards a wall or through an open field, does not.

In the absence of rich depth structure, retinal flow patterns may still provide rotational information in the form of 2D shearing and image deformation effects that we term 'dynamic perspective cues' (also see Chapter I). Rotational optic flow can be characterized as a combination of laminar flow and dynamic perspective cues resulting from the changing orientation of the eye relative to the scene (planar image projection shown in Figure 2-11b). Theoretical studies have proposed that the latter may play an important role in estimating and discounting the rotational component of optic flow to estimate heading (Koenderink and van Doorn, 1976, 1981; Grigo and Lappe, 1999). A recent electrophysiological study in MT provides evidence that the visual system may be capable of using these dynamic perspective cues (Kim et al., Under revision).

To examine the role of dynamic perspective cues, we conducted a second set of experiments using a fronto-parallel (FP) plane of dots with zero disparity. These visual stimuli contain 2D rotation cues, but lack motion parallax information. For 11/34 neurons recorded, the stimulus was viewed binocularly; the remaining cells were recorded while the monkey viewed the stimulus monocularly with the eye contralateral to the recording hemisphere. During the middle 750ms duration of each trial, the simulated distance of the FP plane from the monkey changed and resulted in an average FOE shift of 37°. Hence, heading tuning shifts significantly smaller than 37° would provide evidence for the hypothesis that the visual system can use dynamic perspective cues to discount rotations.

Figure 2-11c summarizes the shifts in heading tuning measured during presentation of the

FP plane stimulus. The median shifts across the population for real pursuit (14.3°) and simulated pursuit (21.5°) were both significantly less than the 37° expected if there were no tolerance for rotations (Wilcoxon signed-rank test; $p < 0.005$). The median values were also significantly different from each other (Wilcoxon signed-rank test; $p = 0.03$) and greater than

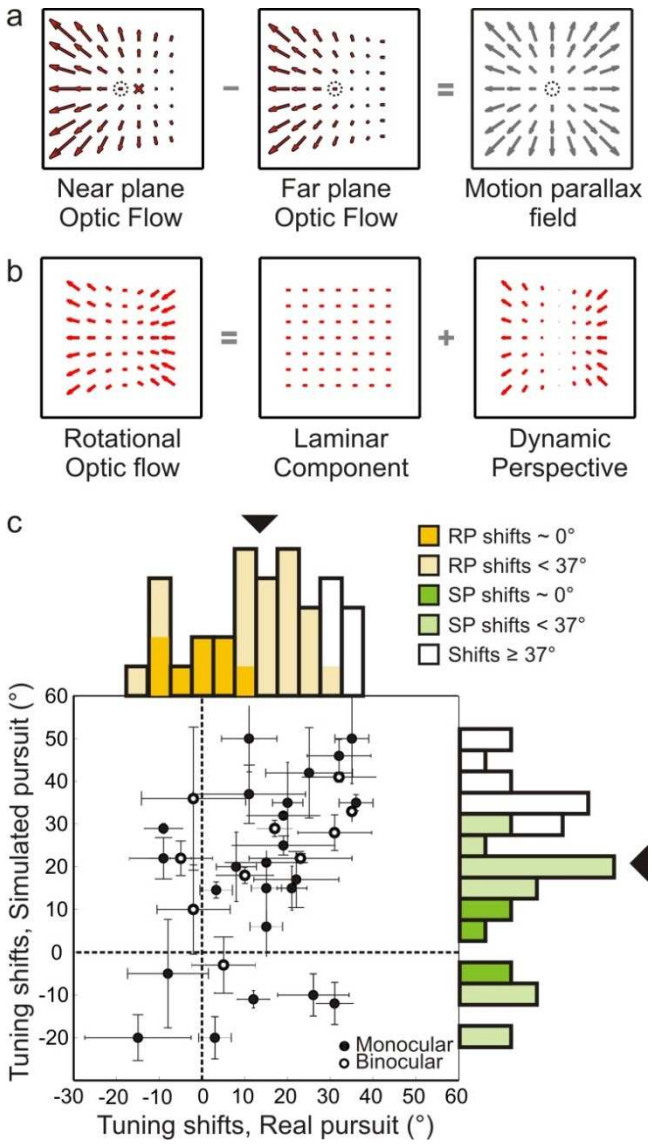

Figure 2-11. Role of 2D dynamic perspective cues in signaling rotation-invariant heading. (a) Optic flow fields during combined T and R, at two different depth planes have different FOE shifts. Subtracting these flow fields yields a motion parallax field that eliminates the rotational component. The point of zero motion parallax corresponds to the true heading (dotted circle). (b) Rotational optic flow can be decomposed into laminar flow and dynamic perspective cues. Dynamic perspective cues may signal eye rotations even in the absence of depth structure. (c) Scatterplot and marginal distributions of shifts measured using the fronto-parallel plane stimulus during RP and SP (n = 34 cells). Open and filled symbols denote data collected during binocular and monocular viewing, respectively. Errorbars denote bootstrapped 95% CIs. All filled histograms indicate shifts significantly < 37°. Dark colored histogram bins indicate cells with shifts not significantly different from 0°. Uncolored bars indicate shifts ≥ 37°.

0° (Wilcoxon signed-rank test; $p < 0.001$). Furthermore, 8/34 (23.5%) neurons during RP and 5/34 (14.7%) neurons during SP had shifts that were not significantly different from 0° (darker colors in Figure 2-11c), implying rotation-invariant heading responses. Only 6/34 (17.6%) neurons during RP and 12/34 (35.3%) neurons during SP showed shifts that were statistically greater than or not different from 37° (bootstrap; see *Methods*). These results indicate that, even in the absence of non-visual signals and 3D visual cues such as motion parallax, a large sub-population of VIP neurons can use 2D perspective cues to at least partially mitigate the effect of rotations on heading tuning. Shifts measured during simulated pursuit in the 3D cloud experiments were significantly less than shifts measured using the FP plane (Wilcoxon rank sum test; $p = 0.02$). This implies that both 3D motion parallax cues and 2D features such as dynamic perspective cues play important roles in visually dissociating translations and rotations. It is also interesting to note that real pursuit shifts were smaller for the 3D cloud stimulus than the FP plane (Wilcoxon rank sum test; $p = 0.007$). From this we can infer that having additional visual cues enhances heading estimation even in the presence of non-visual signals.

2.3.5 Reference frames for representing heading

Since the eyes physically rotate during real pursuit, but the head does not, previous studies interpreted rotation-invariant heading tuning as evidence that VIP neurons represent self-motion in a head-centered reference frame (Zhang et al., 2004). In contrast, studies that measured heading tuning with the eye and head at different static positions have revealed an eye-centered reference frame for visual heading tuning in VIP (Chen et al., 2013; Chen et al., 2014). On the surface, these results appear to be incompatible with each other. However, we

posit that the issues of rotation-invariant heading tuning and reference frames are not necessarily linked. Indeed, we show below that VIP neurons can discount rotations and still signal heading in an eye-centered reference frame.

The key to reconciling these issues is appreciating that, during eye pursuit, the eye-centered reference frame rotates relative to a subject's heading (Figure 2-12a). As the eye rotates, the direction of translation remains constant in head-centered coordinates (Figure 2-12a, dashed green lines). However, in the rotating eye-centered reference frame, the translation direction relative to the eye changes over time, such that the focus of expansion moves across the retina (Figure 2-12b). In our experimental protocol, as well as that of previous studies (Bradley et al., 1996; Shenoy et al., 1999; Shenoy et al., 2002; Zhang et al., 2004), the average eye position during the translation-only, real pursuit and simulated pursuit conditions is the same (centered on the screen) over the duration of a trial. Therefore, the average eye position is the same as the average head position. As a result, time-averaged neural responses may provide insight into what signal is represented (heading or resultant optic flow), but not about whether these signals are represented in an eye- or head-centered reference frame. To evaluate reference frames, responses must be examined with the eye at different positions relative to the head. In our case, we can examine the temporal responses of neurons to study reference frames since the eye changes position over time. An eye-centered representation of heading would result in temporal response variations due to the rotating reference frame, but a head-centered representation would result in responses that are constant over time.

We analyzed the time course of VIP responses (during the 3D cloud protocol) over the same 750ms epoch used in the rest of the analyses. If neurons signal heading in an eye-

centered reference frame, the largest temporal variations in firing rate will occur at headings along the steepest portion of the tuning curve. Therefore, we identified the heading corresponding to the largest positive gradient for each tuning curve, and examined the temporal dynamics of responses for that direction. For a neuron with an eye-centered reference frame, a rightward eye rotation (Figure 2-12b) at the positive slope of the heading tuning curve would result in an upward trend in firing rate. By contrast, a leftward eye rotation would result in a downward trend. It is important to note that these trends are

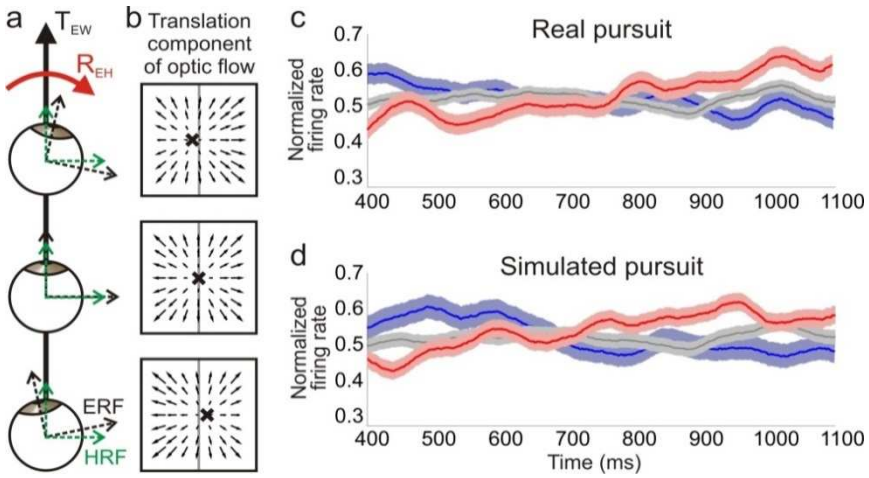

Figure 2-12. Distinguishing reference frames from rotation invariance. (a) Schematic of a rightward eye rotation while translating forward. As eye position changes during pursuit, the eye reference frame (ERF, black axes) rotates relative to the head (R_{EH}) and the direction of translation in the world, T_{EW}. The head reference frame (HRF, green axes) remains constant with respect to the heading. (b) The translation component of optic flow changes with eye position and results in a drifting FOE (x) across the retina. The translation direction represented by the FOE changes from right to left of straight ahead. (c,d) Heading corresponding to the largest firing rate gradient was identified for each neuronal tuning curve and the temporal responses at that heading were evaluated. The population average (n = 72 cells) of the normalized firing rate over time is plotted for – translation only (grey), rightward (red) and leftward rotation (blue) for real pursuit (c) and simulated pursuit (d). Shaded regions indicate standard errors.

determined by the changing eye position and independent of how tolerant the heading representation is to rotations.

Average time courses of normalized responses of the population of VIP neurons showed trends consistent with the hypothesis of an eye-centered or intermediate reference frame (Figure 2-12c, d). Specifically, for real pursuit, average responses increased for rightward eye rotation ($r = 0.8$, $p = 0.002$, Pearson correlation) and decreased for leftward rotation (Pearson $r = -0.6$, $p = 0.009$). Since the temporal response profile was essentially flat during the translation only condition (Pearson $r = 0.11$, $p = 0.8$), these trends cannot be explained by other basic aspects of neural response dynamics, such as adaptation. Interestingly, similar trends are also observed during simulated pursuit (rightward: Pearson $r = 0.6$, $p = 0.006$; leftward: Pearson $r = -0.6$, $p = 0.02$), for which the eye does not physically rotate. Previous studies have demonstrated the role of non-visual signals in estimating the position of the eye or head relative to the body (Squatrito and Maioli, 1997; Lewis et al., 1998; Klier et al., 2005). In contrast, these results suggest that visual signals in VIP carry information about dynamic eye position even in the absence of efference copies. In other words, the temporal dynamics of an eye rotation may be inferred from the rotational components of optic flow and used to modulate neural responses during simulated pursuit. This further strengthens the functional role of visual signals in VIP for estimating rotational information and contributing to a rotation-invariant heading representation.

2.4 Discussion

We evaluated how heading is represented in macaque area VIP in the presence of rotations. We showed that a sub-population of VIP neurons represent heading in a rotation-

invariant fashion while a majority of the population is at least partially tolerant to rotations. Importantly, rotation invariance can be achieved using both non-visual and purely visual cues. Previous neurophysiology literature emphasized the importance of non-visual cues, especially efference copies, but clear evidence for the role of visual cues has been missing, as discussed below. In contrast, our study provides novel evidence for the role of visual cues in discounting rotations and representing heading. Furthermore, we show that both 3D (motion parallax) and 2D (dynamic perspective) visual cues present in optic flow play a significant role in decomposing the components of self-motion. The importance of visual signals is reinforced by our finding that VIP neurons also carry rotation signals derived from purely visual cues. The significant correlation between visual and non-visual rotation responses is consistent with a multi-sensory representation of rotations. In addition, we resolve an important ambiguity in the literature between the concepts of tolerance to rotations and reference frames. Specifically, we examine the effect of a rotating eye reference frame on visual responses to show that rotation tolerance does not necessarily imply a head-centered reference frame. Our findings show conclusively that visual cues play a significant role in achieving rotation-invariant heading representations.

2.4.1 Importance of visual cues

It is important to recognize that the significance of visual cues in discounting rotation extends beyond eye pursuit to head-on-body (R_{HB}) and body-in-world (R_{BW}) rotations as well. The efference copy for each of these sources of rotation depends on the specific motor commands generating the movement. If we consider that eye, head, and body rotations are often generated simultaneously, multiple efference copy signals must be added together and subsequently discounted from the resultant optic flow to signal heading accurately. Each of

these non-visual signals is associated with signal-dependent noise (Gellman and Fletcher, 1992; Li and Matin, 1992; Crowell et al., 1998); thus, combining multiple, potentially independent, efference copy signals to estimate rotations may not always be an efficient solution for the brain. On the other hand, the information contained in visual cues is independent of the source of rotation and represents rotation of the eye relative to the world (R_{EW}). The R_{EW} information present in optic flow inherently reflects the sum of all the different sources of rotation ($R_{EW} = R_{EH} + R_{HB} + R_{BW}$) and thus provides direct information regarding the total rotation of the eyes during self-motion. Therefore, visual signals may have important advantages when the goal is to accurately estimate heading in the presence of self-generated rotations.

However, we also face situations in which visual information may be sparse, such as driving at night on an open road (limited visual range and depth structure), and non-visual signals may be crucial. As expected, given the brain's propensity towards multi-sensory integration, we find that both visual and non-visual signals contribute to discounting rotations to represent heading. Real pursuit shifts are smaller than simulated pursuit shifts, and both types of shifts are smaller for a dense 3D cloud than a fronto-parallel plane.

Given the variety of efference copy signals present in parietal cortex (Andersen, 1997) and the correlation observed between the R_{EH} (pursuit in darkness) and R_{EW} (pure simulated rotation) responses in our data (Figure 2-10), we postulate that VIP contains an integrated representation of rotation that relies on both visual signals and efference copy inputs. However, to conclusively test this theory, experiments with multiple rotation velocities as well as different sources of rotation (e.g., eye vs. head pursuit) need to be conducted. How visual rotation cues are combined with efference copy signals and other non-visual sensory

cues to rotation (e.g., vestibular inputs) warrants further investigation.

2.4.2 Comparison with previous studies

Previous physiological studies emphasized the contribution of efference copy signals to achieving rotation invariance (Bradley et al., 1996; Page and Duffy, 1999; Shenoy et al., 1999; Zhang et al., 2004). However, these studies could not conclusively establish a contribution of visual rotation cues to heading tuning for various reasons. Some studies did not use a simulated pursuit condition and therefore could not disambiguate visual and non-visual contributions to the rotation-invariance of heading tuning they observed (Page and Duffy, 1999; Zhang et al., 2004). On the other hand, Bradley et al. (1996) and Shenoy et al. (1999; 2002) included a simulated pursuit condition in their experiments, but the visual stimulus used to simulate pursuit was incorrect. To mimic pursuit, they simply added laminar flow to their expanding optic flow stimuli by drifting the stimulus across the display, and thus their stimuli lacked the dynamic perspective cues present during real rotations. When rendering visual stimuli, dynamic perspective cues should be incorporated any time the eye changes orientation relative to the scene (Kim et al., Under revision).

If eye rotation is simulated (incorrectly) as laminar flow on a flat screen, then it should not be possible for neurons to exhibit rotation-tolerant heading tuning because the addition of laminar motion simply shifts the focus of expansion in the flow field, and does not provide any rotation cues. Indeed, Bradley et al. (1996) found that MSTd neurons did not compensate for rotations when pursuit was simulated in this manner. In contrast, Shenoy et al. (1999; 2002) reported that MSTd neurons show considerable tolerance to rotation when pursuit was simulated as laminar flow, despite the fact that little or no rotation tolerance was reported psychophysically by the same laboratory for simulated pursuit (Crowell et al., 1998).

Compared to Bradley et al. (1996), Shenoy et al. (2002) used a smaller display size and yet observed larger compensatory effects. This finding contradicts theoretical and psychophysical studies that have established that a larger display size should improve pursuit compensation based on visual cues (Koenderink and van Doorn, 1987; Grigo and Lappe, 1999).

We believe that the counter-intuitive results obtained by Shenoy et al. (1999; 2002) stem from the fact that the boundary of their visual stimuli moved across the retina during real and simulated pursuit (but not during the fixation condition), and thus stimulated different regions of the visual field in and around the receptive field of a neuron over time. Such a moving image boundary defined only by the rotation velocity would not occur under natural conditions as a result of eye rotations. By changing the region of visual space that was stimulated over the course of a trial, Shenoy et al. (1999; 2002) likely invoked changes in the amplitude (response gain) or shape of heading tuning curves. Moreover, since they measured heading tuning over a narrow range (± 30°) around straight ahead, such changes may have confounded their estimates of tuning shifts with gains. As shown in Figure 2-8, cross-correlations are insufficient to deal with the bandwidth changes observed in our data. This is especially of concern since a large percentage of neurons in VIP prefer lateral headings (Chen et al., 2011), resulting in bandwidth changes in the presence of rotations. Thus, we do not consider the findings of Shenoy et al. (1999; 2002) as providing reliable evidence for a role of visual mechanisms in constructing a rotation-tolerant representation of heading.

More recently, Bremmer et al. (2010) and Kaminiarz et al. (2014) reported that neurons in areas MSTd and VIP, respectively, show rotation-invariant heading tuning based solely on visual cues. However, these studies only measured neural responses to three headings (forward, 30º leftward, and 30º rightward), and defined rotation-tolerance based on a rank-

ordering the heading responses across the different eye movement conditions. Since absolute firing rates were not considered, it seems likely that shifts in tuning curves could go un-detected by this method. For instance, their analysis would report identical rank-order for all tuning curves shown in Figure 2-8, and erroneously classify them as rotation-invariant. In addition, the authors did not attempt to compare their results to the tuning shifts that would be expected if neurons do not compensate for rotation. Consider that in their ground-plane stimuli (e.g., Fig. 1 of Kaminiarz et al., 2014), rotation has a large effect on slow-speed optic flow vectors near the horizon, and high-speed foreground vectors are much less altered. For neurons with receptive fields below the horizontal meridian or those with responses dominated by high speeds, one might not expect the rank ordering of heading responses to change even if neurons do not compensate for rotation. Thus, the results of these studies are difficult to interpret.

By comparison with the above studies, we accurately simulated eye rotations such that correct 2D and 3D visual cues are present in the stimuli. We also measured full heading tuning curves and our analysis methods allowed us to disambiguate response gain changes from shifts or shape changes in the tuning curve. By using a large display and maintaining the same area of retinal stimulation for all viewing conditions (see Methods), we eliminated artifacts that likely confounded the results of some previous studies (Shenoy et al., 1999; Shenoy et al., 2002). Therefore, we are confident that our findings in the simulated rotation condition reflect a true contribution of visual cues to the problem of dissociating translations and rotations.

2.4.3 Implications for self-motion and navigation

In order to navigate through the environment and interact successfully with objects, it is

imperative that we distinguish visual motion caused by self-generated movements from that caused by external events in the world (Probst et al., 1984; Wallach, 1987; Warren and Saunders, 1995). For instance, the visual consequences of eye or head rotations need to be discounted in order to accurately perceive whether an object is stationary or moving in the world. The neuroscience literature has extensively studied and emphasized the contribution of efference copy signals to discounting self-generated movements in several sensory systems (Andersen, 1997; Cullen, 2004; Klier et al., 2005). We have presented novel evidence for an alternative solution that is available to the visual system – using large-field motion cues to discount self-generated rotations. The ability of VIP neurons to represent heading during rotations, even in the absence of efference copy signals, suggests that visual mechanisms may make substantial contributions to a variety of neural computations that involve estimating and accounting for self-generated rotations.

The contribution of visual cues may be especially important in situations where efference copy signals are either unreliable or absent. For instance, driving along a winding path and looking in the direction of instantaneous heading does not result in any eye or head movements relative to the body (i.e., no efference copy signals). However, such curvilinear motion still introduces rotational components in the optic flow field and disrupts the FOE. In order to estimate such motion trajectories, the visual system would need to decompose self-motion into both translational and rotational components. This study suggests that such trajectory computations based purely on optic flow may be feasible. How the visual system may implement such computations warrants further research and may provide useful insights to neuroscientists as well as those in the fields of computer vision and robotic navigation.

2.5 Experimental Procedures

2.5.1 Subjects and Surgery

Two adult rhesus monkeys (*Macaca mulatta*), weighing 8-10kg, were chronically implanted with a circular molded, lightweight plastic ring for head restraint and a scleral coil for monitoring eye movements (see Gu et al., 2006; Fetsch et al., 2007; Takahashi et al., 2007 for more detail). Following recovery from surgery, the monkeys were trained to sit head restrained in a primate chair. They were subsequently trained using standard operant conditioning to fixate and pursue a small visual target for liquid rewards, as described below. All surgical and experimental procedures were approved by the Institutional Animal Care and Use Committees at Washington University and Baylor College of Medicine, and were in accordance with NIH guidelines.

The primate chair was affixed inside a field coil frame (CNC Engineering, Seattle, WA, USA) with a flat display screen in front. The sides and top of the coil frame were covered with a black enclosure that restricted the animals' view to the display screen. A three-chip DLP projector (Christie Digital Mirage 2000, Kitchener, Ontario, Canada) was used to rear-project images onto the 60 x 60 cm display screen located ~30cm in front of the monkey (thus subtending 90° x 90° of visual angle). Visual stimuli were generated by an OpenGL accelerator board (nVidia Quadro FX 3000G). The display had a pixel resolution of 1280×1024, 32-bit color depth, and was updated at the same rate as the movement trajectory (60 Hz). Behavioral control and data acquisition were accomplished by custom scripts written for use with the TEMPO system (Reflective Computing, St. Louis, MO, USA).

2.5.2 Stimuli and Task

Visual stimuli were presented for a duration of 1500ms during each trial and consisted of various combinations of eight heading directions in the horizontal plane (Figure 2-1c) and two rotational directions (leftward and rightward). Translation and rotation velocities followed a trapezoidal profile in which the velocity was constant (translation: 24cm/s, rotation: 16°/s) during the middle 750ms (Figure 2-1d) of the stimulus period.

The optic flow stimuli were generated using a 3D rendering engine (OpenGL) to accurately simulate combinations of observer translation and rotation. In the 3D cloud protocol, the virtual scene consisted of a cloud of dots that was 150cm wide, 100cm tall, 160 cm deep and had a density of 0.002 dots/cm^3. In order to maintain a constant volume of dots during the 27cm translation over the duration of a trial, the cloud was clipped in depth to range from 25cm to 125cm in front of the monkey at all times. The stimulus was rendered as a red-green anaglyph that the monkey viewed stereoscopically through red/green filters. In the second experimental protocol, a fronto-parallel plane (FP) of dots was rendered with a density of 0.2 dots/cm^2. The plane of dots was rendered with zero binocular disparity and was viewed by the monkey either binocularly or monocularly. During the course of a trial (1500ms), the 27cm translation resulted in the simulated distance of the wall changing from 45cm at the beginning, to 18cm at the end.

During each session, the monkey's eye position was monitored online using the implanted scleral search coil. Only trials in which the monkey's eye remained within a pre-determined eye window (see below) were rewarded with a drop of juice. Trials were aborted if the eye position constraints set by the eye window were violated.

The experiment consisted of 3 main trial types: pure translation, translation + real eye

pursuit (RP), and translation + simulated pursuit (SP). i) For the pure translation condition, the monkey fixated a visual target at the center of the screen and maintained fixation within a 2° eye window while the optic flow stimuli were presented. Optic flow stimuli simulated 8 headings within the horizontal plane, corresponding to all azimuth angles in 45º steps. The pure translation stimuli were rendered by translating the OpenGL camera along one of the 8 headings with the velocity profile shown in Figure 2-1d. ii) For the real pursuit (RP) condition, the animal actively pursued a moving target while the same translational optic flow stimuli were presented on the display screen. A rightward rotation trial started when the fixation target appeared 9º to the left of center. Once the monkey fixated this target (within 1000ms), it moved to the right following a trapezoidal velocity profile (Figure 2-1d). Analogously, leftward pursuit trials began with the target appearing on the right and moving leftward. The monkey was required to pursue the moving visual target and maintain gaze within a 4° eye window during the acceleration and deceleration periods (0:375ms and 1125:1500ms). During the middle 750ms of the trial (constant velocity phase), the monkey was required to maintain gaze within a 2° window around the visual target. Importantly, the optic flow stimulus was windowed with a software rendered aperture that moved simultaneously with the pursuit target. Thus, the area of the retina being stimulated during the RP trials remained constant over time, eliminating potential confounds from moving the stimulus across the receptive field over time (see Discussion). iii) For the simulated pursuit (SP) condition, optic flow stimuli were presented that accurately simulated combinations of the same 8 headings with leftward or rightward rotations, while the monkey fixated at the center of the screen (2° window). These stimuli were rendered by translating and rotating the OpenGL camera with the same trapezoidal velocity profile of the moving target in the RP

condition. This ensured that the retinal optic flow patterns in the RP and SP conditions were identical (assuming accurate pursuit in the RP condition). The area of retinal stimulation was also identical in the SP and RP conditions.

In addition to these main stimulus conditions, the experimental protocol also included three types of pure rotation conditions for both leftward and rightward directions: i) eye pursuit over a black background (with the projector on), ii) eye pursuit over a static field of dots, and iii) simulated pursuit across a static field of dots. We also included a blank screen during visual fixation and a static field of dots during fixation to measure the spontaneous activity and baseline visual response of the neurons, respectively. Therefore, each block of trials (for both 3D cloud and FP protocols) consisted of 48 unique stimulus conditions: 8 directions * (1 translation only + 2 RP + 2 SP) + 8 controls.

2.5.3 Electrophysiological recordings

To record from single neurons extracellularly, tungsten microelectrodes (FHC; tip diameter, 3 µm; impedance, 1-3 MΩ at 1 kHz) were inserted into the cortex through a transdural guide tube, using a hydraulic microdrive. Neural voltage signals were amplified, filtered (400–5000 Hz), discriminated (Plexon Systems), and displayed on SpikeSort software (Plexon systems). The times of occurrence of action potentials and all behavioral events were digitized and recorded with 1ms resolution. Eye position was monitored online and recorded using the implanted scleral search coil. Raw neural signals were also digitized at a rate of 25 kHz using the Plexon system for off-line spike sorting.

VIP was first identified using MRI scans as described in detail in Chen et al. (2011). Electrode penetrations were then directed to the general area of gray matter around the medial tip of the intraparietal sulcus with the goal of characterizing the entire anterior-posterior

extent of area VIP – typically defined as the intraparietal area with directionally selective visual responses (Colby et al., 1993; Duhamel et al., 1998). To determine direction selectivity, we presented a patch of drifting dots for which the size, position, and velocity could be manipulated manually with a computer mouse. We used this mapping procedure to characterize the presence or absence of strong visual drive as well as the direction and speed selectivity of multi-unit and single-unit activity. At each location along the anterior–posterior axis, we first identified the medial tip of the intraparietal sulcus and then moved laterally until there was no longer a directionally selective visual response in the multi-unit activity.

During each experimental session, we inserted a single microelectrode into the region of cortex identified as VIP. Single unit action potentials were then isolated online using a dual voltage-time window discriminator. Within the region of gray matter identified as VIP, we recorded from any neuron that showed robust visual responses during our search procedure. Once a single unit was isolated, we ran the 3D cloud protocol with all conditions randomly interleaved (72 neurons). Each stimulus was repeated at least 4, and usually 5, times. At the end of the 3D cloud protocol, if isolation of the neuron remained stable, we ran the fronto-parallel plane (FP) protocol for 4-5 repetitions (34 neurons). For the FP protocol, the red/green stereo glasses were either removed during the binocular viewing sessions (11/34), or replaced with an eye patch during the monocular viewing sessions (23/34), such that the eye ipsilateral to the recording hemisphere was occluded.

2.5.4 Analyses

Analysis of spike data and statistical tests were performed using MATLAB (MathWorks). Tuning curves for the different stimulus conditions (translation only, RP, SP) were generated using the average firing rate of the cell (spikes/s) during the middle 750ms of each

successfully completed trial. This analysis window was chosen such that rotation/translation velocities were constant and the monkey was pursuing or fixating the visual target in the small 2^o window. To determine the effect of rotations on neural responses, the translation only tuning curve was compared to the RP/SP tuning curves.

Previous studies (Bradley et al., 1996; Page and Duffy, 1999; Shenoy et al., 1999; Shenoy et al., 2002; Zhang et al., 2004; Kaminiarz et al., 2014) only measured tuning curves over a narrow range of headings around straight ahead. Without measuring the full tuning curve, it is very difficult to distinguish between gain fields and shifts in the tuning curves (Mullette-Gillman et al., 2009; Chang and Snyder, 2010; Rosenberg and Angelaki, 2014). Furthermore, these previous studies assumed that rotations would cause a global shift of the tuning curve in the absence of pursuit compensation. However, as shown in Figure 2-3 and Figure 2-4, rotations can change the shape of the tuning curve, including both skew and bandwidth changes. Therefore, the cross-correlation methods or rank-ordering of responses used in previous studies are insufficient to characterize changes in heading tuning due to rotations.

To account for these more complex changes in heading tuning curves, we developed a novel 3-step analysis procedure, as illustrated for an example cell in Figure 2-7. *Step 1*: We measured the minimum and maximum responses of the pure translation tuning curve. The lowest response (trough) and amplitude (maximum – minimum) of the RP/SP tuning curves were then matched to those of the pure translation curve by vertically shifting and scaling the responses, respectively. *Step 2*: Because the predicted effects of rotation are opposite for forward and backward headings (Figure 2-3a), RP and SP tuning curves were split into heading ranges of 0-180° and 180-360°. We tested whether each half of the tuning curve was significantly tuned using an ANOVA ($p \leq 0.05$). All the tuning curves were then linearly

interpolated to a resolution of 1º. *Step 3*: For half-curves that showed significant tuning, we performed a shift analysis as follows. The pure translation tuning curve was circularly shifted (in steps of 1º) to minimize the sum-squared error with each half of the RP/SP tuning curves. For neurons that were significantly tuned in all conditions and in both direction ranges, this analysis yielded four shift values for real pursuit and four shifts for simulated pursuit. In order to quantify the transformation of heading tuning due to rotations, the four shift values were averaged to arrive at one shift value for real pursuit and one shift for simulated pursuit for each cell.

In order to test the efficacy of our analysis method, we simulated neuronal tuning curves using von Mises functions with preferred direction, bandwidth and skew parameters (Swindale, 1998). To simulate the transformations due to rotation, bandwidth and skew were manipulated to correspond to a 20° shift around straight ahead. Random gain values ranging from 0.5 to 2 were multiplied and offset values (0 to 40 spikes/s) were added to these transformed tuning curves. Poisson random noise was added to both the original and transformed tuning curves and sampled every 45°, similar to the recorded data. Shifts were measured between the original and transformed curves using the partial shift analysis method described above. The results of this simulation (Figure 2-7e) reveal that our method is capable of measuring shifts in the presence of gain, offset and shape changes. To compare our method with other commonly used analyses, data from an example neuron was analyzed using the cross-correlation method (Bradley et al., 1996; Shenoy et al., 1999; Shenoy et al., 2002) and rank ordering of the responses (Bremmer et al., 2010; Kaminiarz et al., 2014) for the narrow range between 45° and 135° (similar to those studies). In contrast with our analysis, these methods failed to identify the changes in bandwidth and erroneously classify responses as

rotation-invariant (Figure 2-8).

The 95% confidence intervals (CIs) for the shifts plotted in Figures 2-9 and 2-11, were calculated using a bootstrap analysis. Bootstrapped tuning curves for translation only, real pursuit, and simulated pursuit were generated by resampling responses with replacement. The same offset, gain and shift calculations were performed on each one of 300 bootstrapped tuning curves to produce a distribution of shifts for each neuron from which the 95% CI was calculated by the percentile method.

To test the rotating reference frame hypothesis (Figure 2-12), the time course of firing rate was measured at the heading for which the tuning curve had its steepest positive slope, for neurons recorded during the 3D cloud protocol. To identify the heading direction to be evaluated, the gradient of firing rate was calculated at each point on the measured tuning curve and the heading associated with the largest positive gradient was selected. For sharply tuned neurons, it is possible that the true largest gradient lay between sampled headings. Hence, the measured largest gradient could be part of the peak or trough of the tuning curve. To account for such instances in the data, we excluded tuning curves for which the mean response at the largest gradient heading was not significantly different (t-test; $p \leq 0.05$) from the responses of its immediate neighboring headings (29/360 total tuning curves from 72 cells). The time course of firing rate during each trial for the selected heading was calculated by convolving the spike events with a Gaussian kernel (σ = 25ms). The temporal responses from all selected tuning curves were averaged by condition and used to calculate the mean and standard errors shown in Figures 2-12c, d.

Chapter 3 – Joint Representation of Translations and Rotations

3.1 Introduction

As we navigate through the world, all our complex movements and trajectories can be described as a combination of translations and rotations (Figure 3-1A). Both translations and rotations results in specific motion patterns onto the retina (optic flow), which get added together during simultaneous translation (T) and rotation (R) (Figure 3-1B). However, based on the navigational task at hand, we may require an estimate of T or R, or most likely both, requiring a decomposition of the optic flow into its components. For instance, travelling along a curved path requires both the translation and rotation estimate to correctly predict the trajectory and navigate around obstacles. A rotation estimate is also crucial for estimating the velocity of a tracked object, for instance when stalking moving prey.

Despite the importance of representing rotations during self-motion, most psychophysical and electrophysiological studies refer to rotations as something to be 'subtracted out' in order to estimate heading (Royden et al., 1992; Royden, 1994; Bradley et al., 1996; Crowell et al., 1998; Page and Duffy, 1999; Zhang et al., 2004). These studies have shown that the translation representation in the visual system can be invariant to rotations, but do not address the question of representing rotations in the presence of translations. Furthermore, many of these studies emphasize the importance of non-visual cues (such as efference copies of eye movements) to derive an estimate of the rotational component. But efference copy signals are often noisy or in some situations non-existent (e.g. travelling along a curved path). Therefore, it is important that the rotational component of self-motion can also be estimated based purely

on optic flow when translations and rotations are present simultaneously.

Theoretically, optic flow can be decomposed into its T and R components, owing to the differences in the properties of translational and rotational flow fields (Longuet-Higgins and Prazdny, 1980; Rieger and Lawton, 1985; see Chapter 1 for details). However, the joint representation of translations and rotations in the visual system, based solely on the decomposition of optic flow has not been previously addressed. Given the ecological importance of having both translation and rotation estimates, we hypothesize that the visual system can decompose optic flow into its translational and rotational components and jointly represent both at the single neuron level.

In this study, we test this hypothesis by recording single cell responses in the macaque ventral intraparietal area (VIP) during the presentation of visual stimuli simulating combined translations and rotations. Our results provide the first evidence for a joint representation of translations and rotations at the level of single neurons based purely on visual cues. We further show that the gain fields observed in the joint tuning are similar to the responses observed during pure rotations, indicating a true representation of rotation velocity in area VIP. Lastly, we show that the rotation responses are similar during self-generated eye rotations and simulated rotations, indicating that rotation tuning in VIP is largely independent of the source of the rotation information (i.e. visual and non-visual cues).

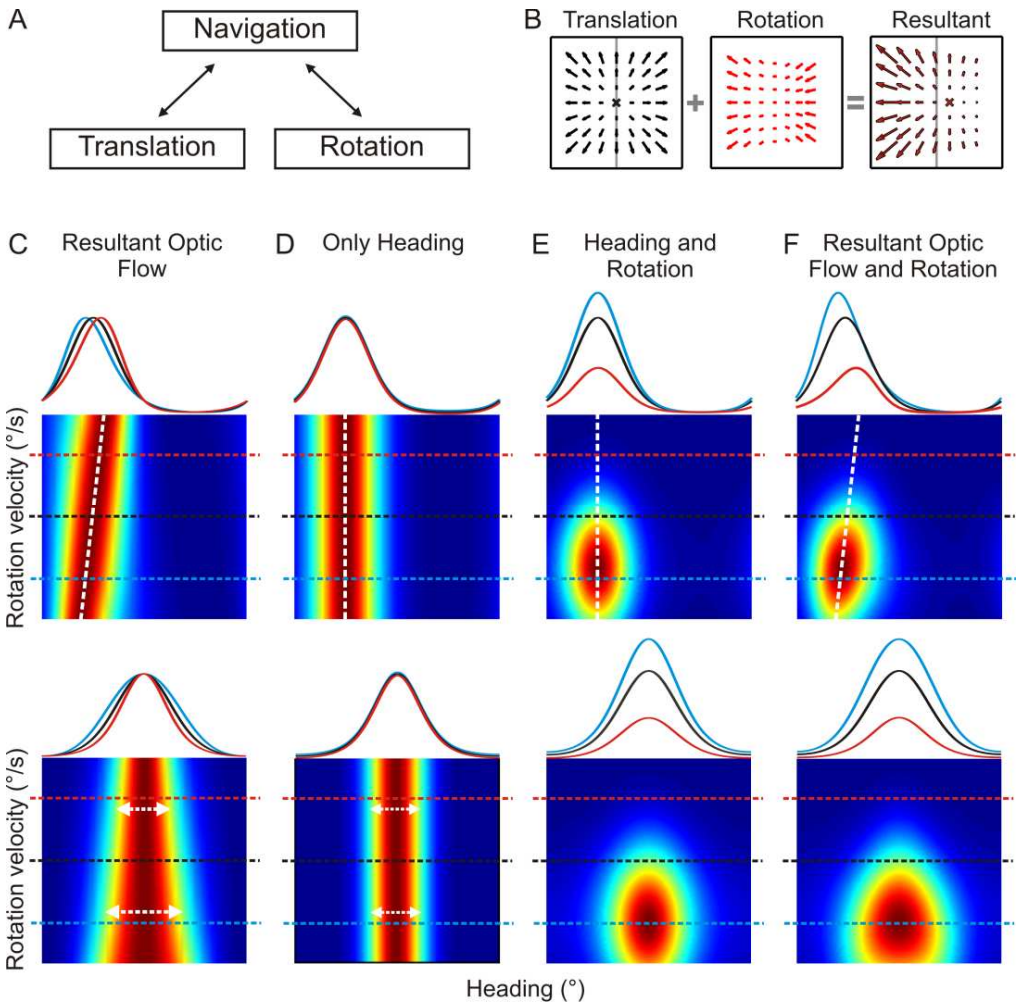

Figure 3-1. Simulation of joint T and R representations. A) Self-motion consists of both translations (T) and rotations (R) and navigation requires information about both these components. B) The optic flow patterns projected onto the retina during self-motion are a combination of the translational and rotational component patterns. C-F) Simulations of the predicted range of joint translation and rotation tuning curves for cells preferring forward heading (row 1) and lateral heading (row 2). C) Simulation of joint tuning representing the resultant optic flow without decomposing the T and R components. D) Only the translation component is extracted and represented. E) Both the T and R components are represented. F) The translation representation is influenced by added rotations (i.e. not rotation-invariant), but the rotation component is still represented.

3.2 Results

To investigate whether neurons in area VIP jointly represent both the translational and rotational components of self-motion, we used visual stimuli simulating different combinations of heading and rotation velocities (see Methods). The translational component of the stimuli consisted of 8 headings in the horizontal plane (Figure 3-2A) and the rotational component consisted of either seven or three rotation velocities ranging from 15°/s leftward to

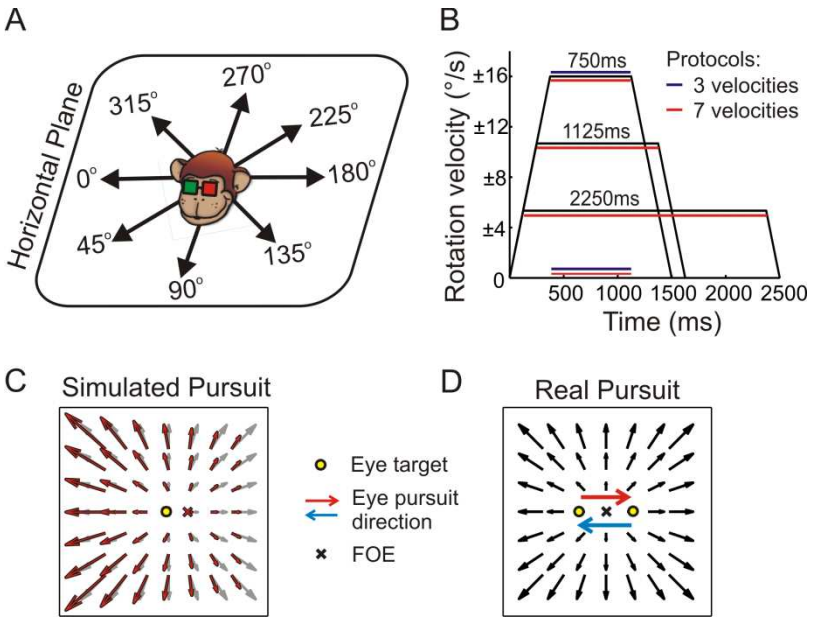

Figure 3-2. Methods. A) Translation was simulated using optic flow stimuli in the horizontal plane in 8 azimuth directions. B) Trapezoidal rotation velocity profiles for different peak rotation speeds. The total displacement of the eye (or simulated displacement) was identical at all rotation velocities. The blue and red lines indicate rotation speeds used in the 3 and 7 rotation velocity protocols, respectively. C) Simulated pursuit (SP) condition where there monkey fixated while optic flow stimuli representing combined translations and rotations was presented. D) Real pursuit (RP) condition, where the monkey smoothly pursued a visual target while translational optic flow was presented in the background. Both conditions had nearly identical retinal stimulation, but the RP condition had both visual and non-visual (efference copy) cues of rotation.

15°/s rightward (Figure 3-2B). The neural responses from these purely visual stimuli (Simulated pursuit, Figure 3-2C) were evaluated to determine the role of visual optic flow processing in decomposing and jointly representing the translational and rotational components of self-motion. These representations were also compared to the responses resulting from pursuit eye movements, where non-visual cues about rotation are also present (Real pursuit, Figure 3-2D), to evaluate the relative roles of visual and non-visual cues. Importantly, the SP and RP stimuli (both T and R) were compared to their corresponding pure rotation stimuli – with only visual information (simulated rotation-only) and both visual and non-visual signals (real rotation-only).

3.2.1 Joint Translation and Rotation Tuning

The simulations presented in Figures 3-1C-F depict the range of joint translation and rotation tuning that might be observed in area VIP. If VIP neurons represent the resultant optic flow, without decomposing it into T and R components, we expect the translation tuning to vary with added rotations. Depending on the preferred direction of the cell, this would result in shifts or bandwidth changes that vary systematically with rotation speed (Figure 3-1C; see Chapter II for details). On the other hand, if the translation and rotation components are perfectly decomposed, we expect to see no changes in the shape of the translation tuning curves. In such a scenario, the cells can represent only the translation component, or both T and R. If only the translation component is represented, we expect to see no gain fields or changes in the shape of the tuning curve in the presence of rotations (Figure 3-1D). However, if the information about rotation velocity is not discarded, we expect to see gain modulation in the translation tuning curves. The responses would then resemble the simulations shown in Figure 3-1E, where both translations and rotations are jointly represented. It is also possible

that the neurons only partially decompose the optic flow, resulting in neural responses for which the translation tuning varies to some extent with the rotation velocity, but the rotation is still represented in the form of gain fields (Figure 3-1F), i.e. combination of shift and gain changes in the translation tuning curves. We evaluated joint tuning curves generated from the neural data to see if VIP neurons do indeed represent rotation velocities in the form of gain fields, as simulated in Figures 3-1E, F.

Joint tuning curves for each cell were generated by measuring the neuronal firing rates during combined translation and rotation stimuli. Figure 3-3 shows joint tuning curves for two example neurons and their corresponding conditional translation and rotation tuning curves on the marginals. The response of the first example neuron during simulated pursuit (Figure 3-3A, left column) is modulated by translation direction, but not rotations. This is apparent from the lack of gain modulation in the translation tuning curves as well as the lack of significant tuning in the rotation tuning curves (1-way ANOVA, $p>0.05$). The joint responses observed here are similar to the simulations presented in Figures 3-1C, D and correspond to the expected features of a cell that only represents translations (T-only cell). The cell in Figure 3-3B, on the other hand, shows significant rotation tuning (1-way ANOVA, $p<0.01$) during simulated pursuit with a preference for -10°/s. The response of this neuron is similar to the expected tuning of a neuron that jointly represents translations and rotations (T&R cell), as simulated in Figures 3-1E, F. Importantly, for both examples, the features of the joint tuning curves are conserved even when the rotations are added in the form of real eye pursuit (right column). These examples show that some VIP neurons may jointly represent T and R, whereas others may represent only the translational component.

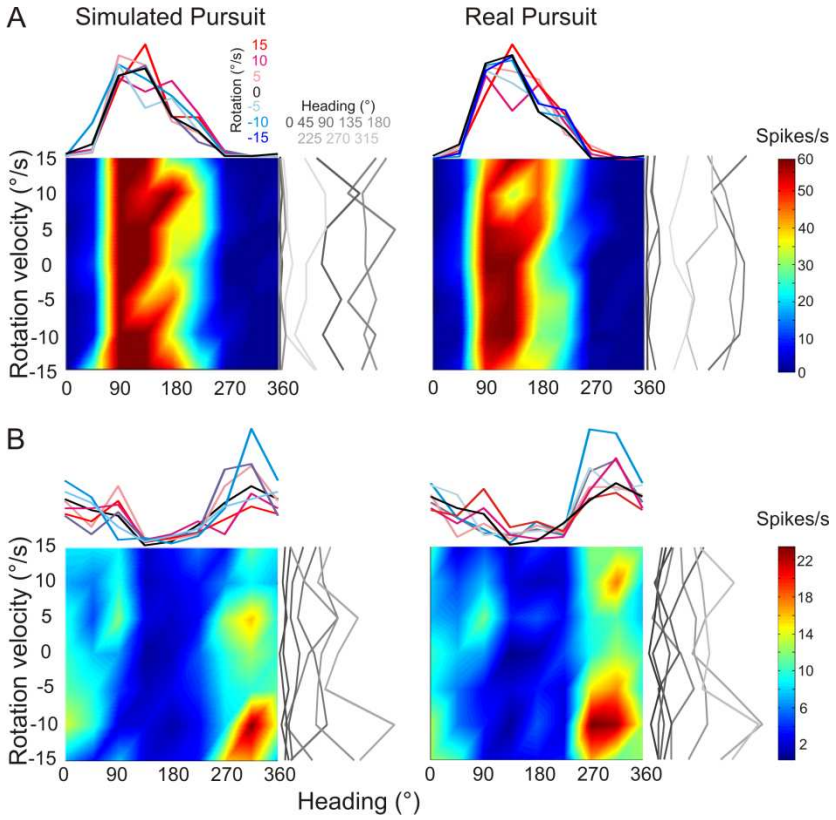

Figure 3-3. Example joint tuning curves. A, B) Joint tuning curves of these two example neurons were generated based on firing rate during the combined translation and rotation stimuli. The curves on the marginals are the conditional tuning curves for translations and rotations. Left column shows SP responses and right column shows RP responses. A) Joint tuning response in this cell were similar to the simulations in Fig. 1 C,D. This cell only represented heading and did not show significant gain fields in the presence of rotations, making it a 'T-only' cell (during both RP and SP). B) This neuron showed significant gain fields and preferred a rotation velocity of -10°/s during both RP and SP, making it a 'T&R' cell. This response is similar to the simulations shown in Figure 3-1E, F.

In order to quantify the presence of rotation tuning during combined translation and rotation, we evaluated the rotation tuning curve at the preferred heading for each cell. Based on this analysis, 28/72 cells during the 3 rotation protocol and 24/50 cells during the 7 rotation protocol showed significant rotation tuning (1-way ANOVA, $p<0.05$) during simulated

pursuit and were classified as T&R cells (52/122, ~43%). Of these T&R cells, only 2 did not have significant gain fields during real pursuit and 5 neurons showed significant gain fields during real, but not simulated pursuit. In the next section we will evaluate if these gain fields observed during added rotations correspond to the rotational velocity present in the optic flow.

3.2.2 Tuning Shifts and Rotation Velocities

The addition of rotations to translational optic flow results in a shift in the focus of expansion (FOE; Figure 3-1B) and may therefore result in a distortion of the translation tuning curve of a neuron (Figure 3-1C, F). In order to quantify these distortions, we measured the shifts between the translation-only tuning curve and the tuning curves measured in the presence of different rotation velocities (see Methods). A shift of 0° indicates a cell that perfectly represents the translation component of the stimulus even in the presence of rotations. A positive shift indicates a shift in the tuning curve in the direction of the FOE shift (under-compensating for rotations) and a negative shift indicates a shift in the direction opposite to the FOE shift (over-compensating for rotations).

FOE shifts are dictated by the rotation velocity, resulting in larger expected shifts in the tuning curve for larger rotation velocities (when translation velocity is constant). Based on the translation and rotation velocities used in our stimulus, the largest shift in tuning is expected at the rotation velocities of ±15°/s. The nearest plane in the 3D cloud stimulus was 25cm from the monkey, which resulted in an FOE shift of 20°. Any plane farther away would result in a larger FOE shift, making an expected tuning curve shift of 20° the most conservative estimate for a non-compensating neuron (see Chapter II for details). We first measured the tuning

curve shifts at the largest rotation velocities (±15°/s) for all cells recorded during the 3 rotation and 7 rotation velocity protocols (n = 122 cells from 3 animals). The median shift at this rotation speed during simulated pursuit (13.7°) was significantly greater than the median shift during real pursuit (9.8°) (Wilcoxson rank-sum test, p=0.04), but both were significantly less than the minimum expected shift of 20° (Wilcoxon signed-rank test; RP: p<0.001, SP: p = 0.01; see also, Chapter II). However, we found no significant difference between the shifts present in T&R cells and T-only cells for either real pursuit or simulated pursuit (Figure 3-4A, B; blue vs. red bars). This implies that the presence or absence of rotation tuning is not dependent upon the degree to which a neuron's translation response is tolerant to rotations.

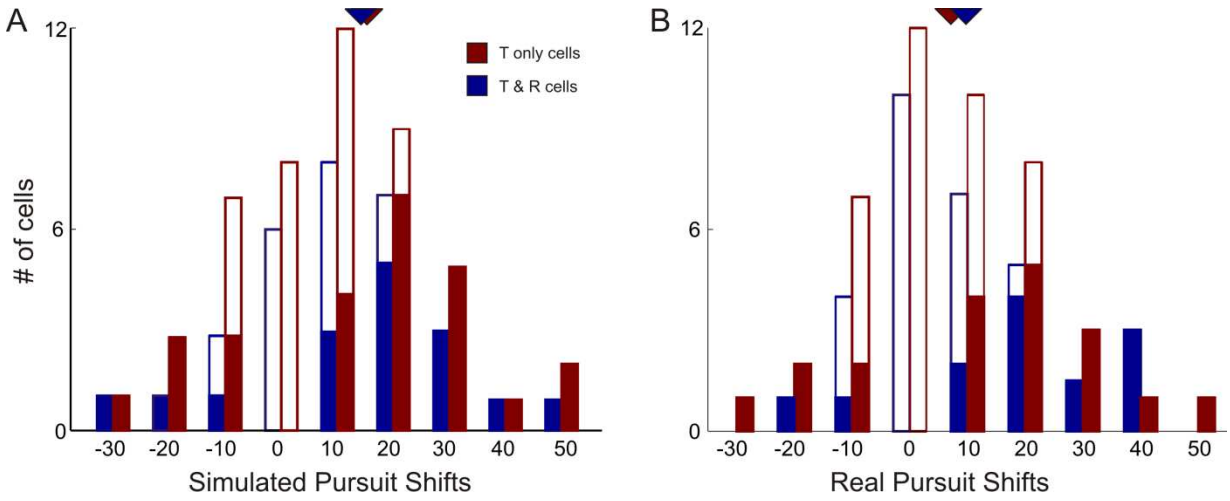

Figure 3-4. Maximum tuning curve shifts for T-only and T&R cells. A, B) Mean tuning curve shift for each cell was calculated by comparing the translation only tuning curve with the RP (A) and SP (B) tuning curves at the largest rotation speeds (±15°/s). Positive shifts indices indicate a shift in same direction as the FOE due to added rotations. Negative shifts indicate a shift in the opposite direction of the FOE shifts and may correspond to an over-estimation of the rotational component. The histograms are separated based on the type of cell (T-only and T&R cells) and pointers at the top indicate median shifts. Filled histograms indicate shifts that were significantly >0°, based on the CIs calculated from bootstrapping. Neither SP, nor RP, showed significant difference in shifts between the two groups of cells.

As shown in the simulations in Figure 3-1, rotations can result in systematic shifts in translation tuning based on rotation velocities. In order to determine if such systematic shifts exist, we evaluated the shifts in translation tuning as a function of rotation velocity for cells recorded using the 7 rotation velocity protocol (n = 50 cells, 2 monkeys). Figure 3-5 shows example joint tuning curves for two T&R cells and two T-only cells along with their corresponding shifts as a function of the rotation velocity. Here, negative values correspond to an expected shift consistent with leftward rotations and positive shifts to that expected with rightward rotations. If a cell does not compensate for the added rotations, we expect the minimum shift to be ±20° at the ±15°/s rotation velocity respectively. Similarly the minimum shift expected is ±13° at ±10°/s and ±6° at the ±5°/s. On the other hand, a perfectly compensating cell would have 0° shifts at all rotation velocities and therefore a slope of 0. The example neural responses in the left column of Figure 3-5 show significant shifts at larger rotation velocities. Importantly, they also show significant slopes describing the relationship between rotation velocities and tuning curve shifts (linear regression, 95% CI of slope not overlapping 0). The examples in the right column, have smaller shifts and slopes that are not significantly different from 0 (linear regression, 95% CI of slope overlaps 0).

Of the 50 cells tested with the 7 rotation protocol, only 14 had significant slopes during SP and 13 during RP (Figure 3-6A, B). The mean slope, however, was not significantly different from 0 during either condition (Wilcoxon signed-rank test; $p<0.001$). It should be noted that two of the cells with significant slopes had negative slopes during both real and simulated pursuit. This corresponds to tuning shifts that are in the direction opposite of the expected FOE shift. A significant negative slope therefore implies that these neurons were over-compensating for the added rotations, but the shifts were still dependent on R.

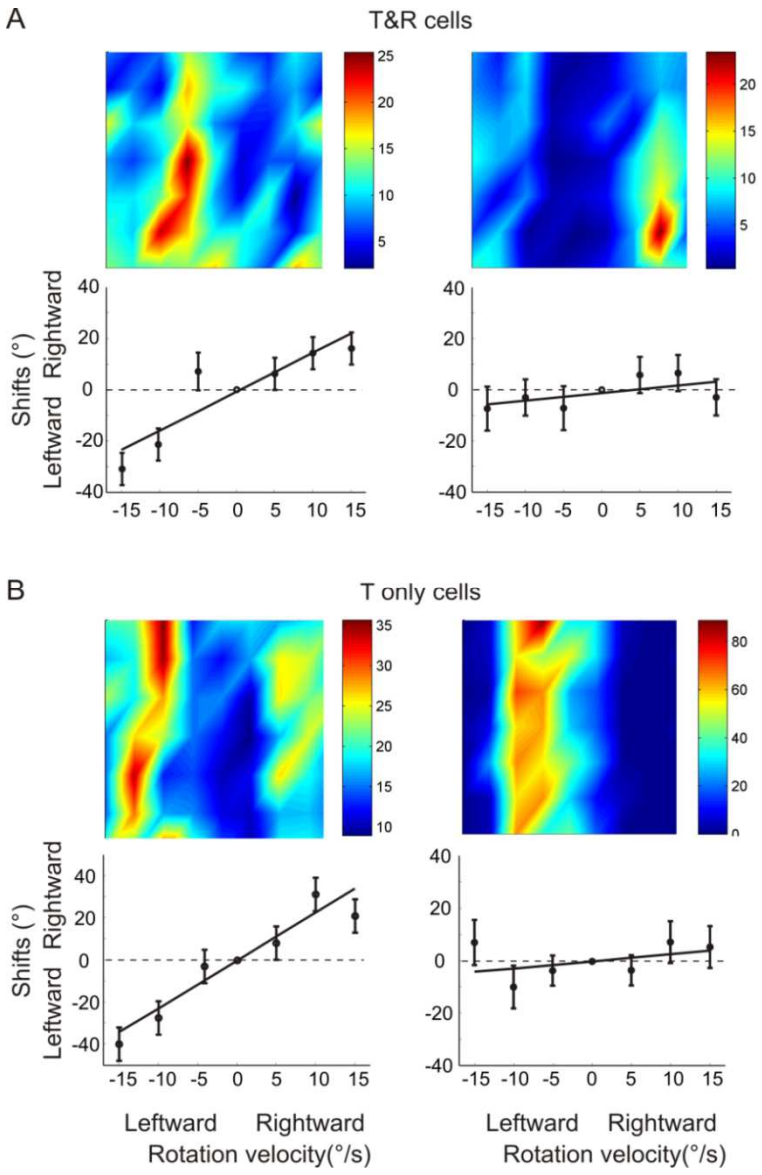

Figure 3-5. Examples of tuning curve shifts due to added rotations. A) Two T&R cells are shown with their joint tuning curves and shifts as function of rotation velocity. The left column shows an example neuron where the shifts are generally larger and vary significantly based on the rotation velocity. The solid line shows the linear regression fit to the bootstrapped data which has a significant slope (95% CI on slope does not overlap 0). The right column shows an example neuron that does not have significant shifts and where the slope of the linear fit is not significantly different from 0 (95% CI overlaps 0). B) Similarly, two example T-only cells – one with a significant dependence of shifts on rotation velocity and one without a significant slope.

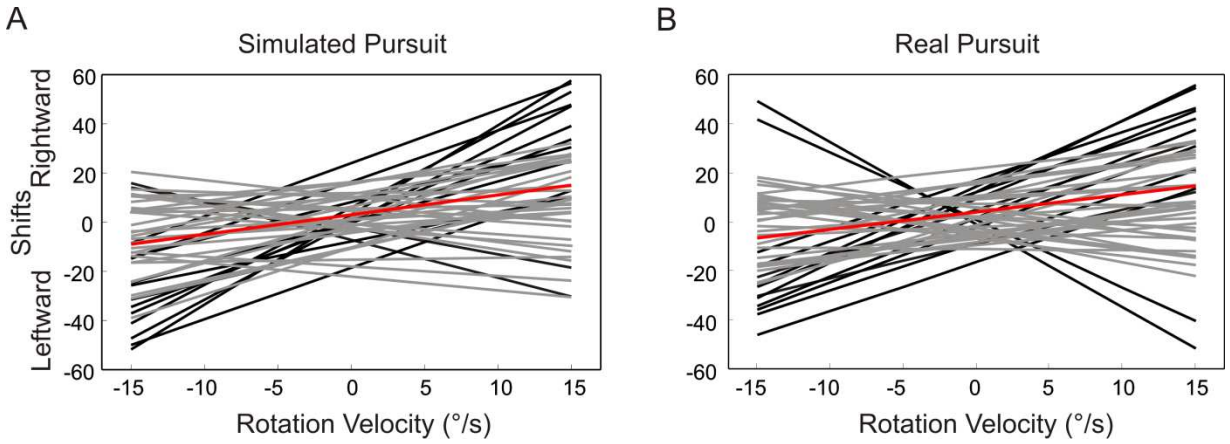

Figure 3-6. Shifts as a function of added rotations. A, B) Linear fits to shifts as a function of rotation velocity for cells collected using the 7 rotation velocity protocol. Dark colored lines indicate slopes that were significantly larger than 0 (based on 95% CI). Red lines indicate mean slope. A) 14/50 cells during SP showed significant slopes, with two cells having significant negative slopes. B) 13/50 cells during RP showed significant slopes, with the same two cells as SP having significant negative slopes.

During simulated pursuit, of the cells that showed no systematic shifts (36/50), 16 were T&R cells (i.e. like Figure 3-1E) and 20 were T-only cells (Figure 3-1D). The 14 cells showing significant shifts were split evenly between T&R (Figure 3-1F) and T-only cells (Figure 3-1C). Hence, the range of joint tuning responses observed in area VIP was well described by the predicted responses in Figure 3-1 C-E. These results, in conjunction with Figure 3-4, show that there was no correspondence between the rotation invariance of a cell (i.e. shifts) and the presence of rotation tuning (i.e. gain).

3.2.3 Rotation Representation in VIP

Theoretically, the visual system should be capable of decomposing self-motion into its translational and rotational components based purely on optic flow (Longuet-Higgins and Prazdny, 1980; Rieger and Lawton, 1985; Koenderink and van Doorn, 1987; Tomasi and Shi,

1996). Results from the previous chapter provide support for the hypothesis that the visual system is capable of extracting the translational component using only visual cues. However, there has been no prior evidence that rotations can similarly be represented based purely on optic flow processing in the presence of translations. We hypothesize that the 43% of neurons that showed significant gain modulation during simulated pursuit (i.e. T&R cells) encode the rotational component of self-motion present in the optic flow stimuli. To quantify these rotation responses, the gain modulation of each cell's response (i.e. rotation tuning) was evaluated at the cell's preferred heading (Figure 3-7A; see Methods).

We evaluate the hypothesis that the gain fields observed during the 'simulated pursuit' condition (combined T and R) are indeed a representation of the rotation velocity in area VIP, by comparing them to the 'simulated rotation-only' (only R) response (Figure 3-7B). We find that the gain modulation of the translation tuning curves correlated significantly with the modulation based on the simulated rotation-only response (Pearson $r = 0.61$, $p = 0.01$). While some neurons showed significant response modulation during the pure rotation stimulus, but not simulated pursuit (18/122), it must be noted that none of those cells were classified as T&R cells. Moreover, only 3/52 T&R cells did not show significant responses during the pure rotation stimulus.

Next, to test if the rotation representation based on visual and non-visual cues are similar, we compared each cell's response to real rotation-only (pursuit over a static field of dots) and simulated rotation-only conditions (see Methods). Figure 3-7C shows that the responses to these two rotation-only stimuli are also correlated (Pearson $r = 0.59$, $p = 0.01$). Some neurons tended to have significant responses to real rotation, but not simulated rotations (12/122) or vice versa (15/122). However, a majority of the neurons that showed significant gain fields

during the combined translation and rotation stimuli (i.e. T&R cells, filled symbols) showed significant response modulation during both real and simulated pure rotations (44/52 cells).

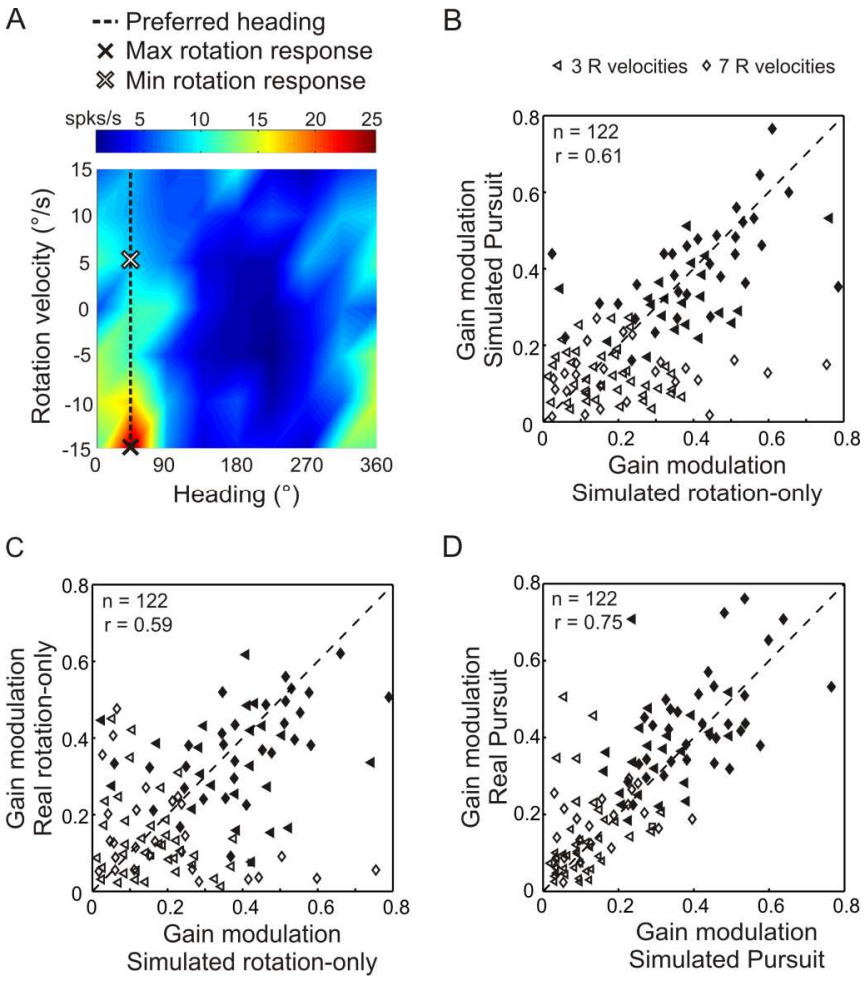

Figure 3-7. Response modulation during rotations. A) Example neuron showing the calculation of the modulation index based on the maximum and minimum response at the preferred heading. B) Scatterplot of response modulation during simulated rotation-only condition and SP (combined T and R), calculated as shown in A. Filled symbols correspond to cells that showed significant gain fields during both RP and SP. C) Scatterplot of response modulation during real and simulated rotation-only conditions. Filled symbols represent cells with significant gain fields during both RP and SP (same as in B). D) Scatterplot showing the modulation of the rotation responses during SP and RP. ~43% of neurons (filled symbols) showed significant gain fields during both RP and SP (1-way ANOVA, $p \leq 0.05$).

The results thus far establish that the response modulation during pure rotations is similar to the modulation during combined T and R and also that the amplitude of responses during pure rotation based on visual and non-visual cues is similar. Next, we evaluate if the gain fields observed in VIP neurons are similar during real and simulated pursuit. Figure 3-7D shows that they are indeed correlated (Pearson $r = 0.75$, $p<0.01$). The filled symbols indicate cells that have significant gain fields during both SP and RP (51/122). Importantly, only 4/122 cells showed significant gain fields in one condition, but not the other (only SP: 1, only RP: 3). These results show that the response modulation in VIP is remarkably similar in the presence or absence of non-visual cues and supports the hypothesis that the estimate of the rotational component is independent of the source of the information, i.e. visual or non-visual.

In order to establish the similarity in rotation responses during combined T and R (SP, RP) and the pure rotation stimuli, we examine the rotation velocity preference of each cell during the two conditions. For the purposes of this analysis, we considered the T&R cells from 7 rotation velocity data ($n = 24$) and fit Gabor functions to the rotation tuning curves obtained from both types of stimuli (median $r^2 = 0.93$). The example rotation tuning curves for two cells, shown in Figure 3-8A, B have similar preferred rotation velocities during both simulated pursuit and pure simulated rotation. The preferred rotation velocities based on these fits during the pure rotation and translation with rotation stimuli are highly correlated for both real (Pearson $r = 0.76$, $p <0.01$) and simulated (Pearson $r = 0.72$, $p<0.01$) rotations (Figure 3-8C). The rotation preferences were uniformly distributed between -15°/s and 15°/s.

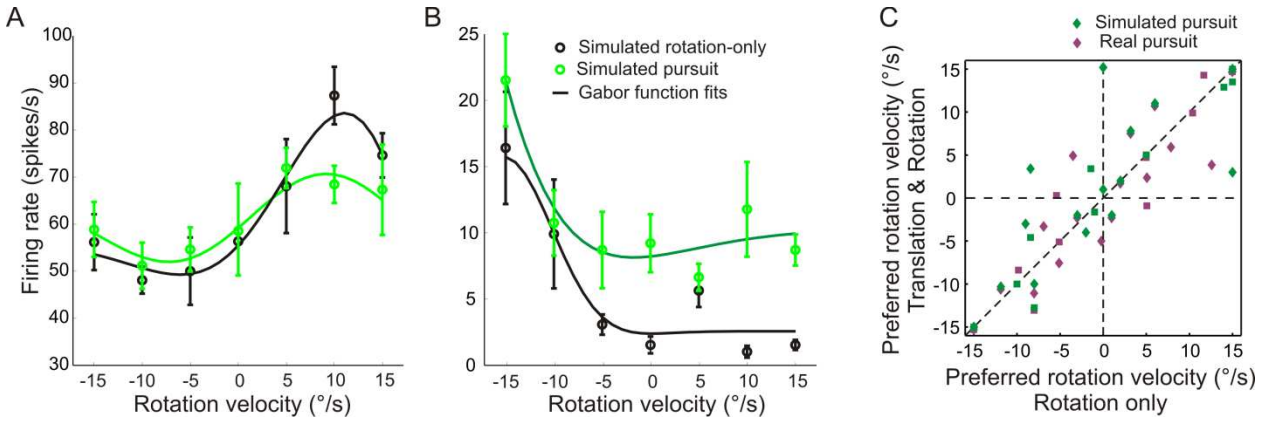

Figure 3-8. Rotation preferences. A,B) Two example neuronal rotation tuning curves measured during simulated pursuit (combined translation and rotation) and simulated rotation-only. The responses in (B) are for the same cell shown in Figure 3-7A. Solid lines represent Gabor function fits to each of the four rotation tuning curves. C) Cells recorded during the 7 rotation velocity protocol with significant rotation tuning were fit with Gabor functions. The scatterplot of preferred rotation velocity (based on the peak of the Gabor function fit) during rotation-only conditions and combined translation, rotation conditions.

These results highlight the similarity in the rotation tuning curves during rotation-only and combined T and R. Therefore, the gain fields observed as a result of added rotations indeed represent the rotational component of self-motion in the presence of translations. Importantly, this rotation representation can be achieved by the visual system through the decomposition of optic flow into its translational and rotational components based on purely visual cues.

3.3 Discussion

Deciphering both the translational and rotational components of self-motion is an important task that the visual system must perform for navigating through the world. Previous studies, both electrophysiological (Bradley et al., 1996; Page and Duffy, 1999; Zhang et al.,

2004) and psychophysical (Royden et al., 1992; Royden, 1994; Crowell et al., 1998) largely focused on how translations (T) are estimated and represented in the presence of rotations (R), neglecting how the visual system represents rotations. Our study is novel in its approach to understanding the representation of rotations, and shows for the first time, the joint representation of T and R. We found that the gain of the translation tuning curve during rotations was modulated in ~43% of neurons (T&R cells) recorded from macaque area VIP, demonstrating a joint representation of T and R. These gain modulations as well as rotation preferences were correlated with neuronal responses to pure rotation stimuli, suggesting that T&R cells encode an estimate of rotation velocity. Importantly, we also show that even in the absence of non-visual cues, T and R components of optic flow are represented in a sub-population (43%) of VIP neurons. This implies that the visual system may indeed be capable of decomposing self-motion into T and R components based purely on retinal signals.

3.3.1 Importance of a Rotational Estimate based on Retinal Cues

Efference copies of eye or head rotations have long been implicated in maintaining a stable visual percept as well as sensorimotor coordination (Bell, 1823; Purkinje, 1823; von Holst and Mittelstaedt, 1950). Later studies implicated these extra-retinal signals in estimating heading in the presence of self-generated rotations (Bradley et al., 1996; Page and Duffy, 1999; Zhang et al., 2004). This focus on self-generated rotations resulted in the prevailing idea that the visual system must discard rotations in order to estimate self-motion, but this idea has two fundamental weaknesses.

First, as discussed in Chapter I and the Introduction, there are several scenarios in which a rotational estimate is just as important as translations for navigation. Therefore, it seems highly unlikely that the representation of self-motion in the visual system is limited to

encoding translations; there must also be an estimate of the rotational component of self-motion. Several studies have evaluated how eye rotations are encoded in the brain (mostly in the form of efference copies), but not in the context of combined T and R self-motion (Fukushima et al., 2002; Duhamel et al., 1997; Schlack et al., 2003). Other studies conducted in area MSTd recorded tuning curves based on spiral optic flow patterns, which occur on ground planes during combined T and R (Graziano et al., 1994; Duffy and Wurtz, 1997; Mineault et al., 2012). However, these studies place the spiral and expansion stimuli on a continuum and do not attempt to study the underlying T or R components separately. Despite the importance of rotations, these previous studies reveal very little about how rotations are encoded in the visual system, especially in the presence of translations.

Second, efference copy signals have several limitations with regards to estimating rotation velocities. Efference copies are, by definition, feedforward signals that inform the brain of the motor command signal. Thus, error between the motor command signal and the actual movement of the eye results in errors in the efference copy. Moreover, studies have shown that visual updating based on efference copies has a long time course (Bridgeman, 1995; Sommer and Wurtz, 2008), which would be a drawback for self-motion processing (Grigo and Lappe, 1999), since our movements are often dynamic. Also consider that we generally make simultaneous eye, head or torso rotations, each of which has a different motor command signal. Retinal optic flow, on the other hand, contains information about the rotation of the eye in the world (i.e. a sum of all rotations). Therefore, in the presence of multiple rotation sources, the noise and errors associated with the efference copies are likely additive and would therefore largely reduce the efficacy of a rotation estimate based on efference copies. But perhaps the largest limitation of relying on extra-retinal signals for a rotation estimate is

that not all rotations are accompanied by efference copies of eye or head rotations. This is especially true for the case of travelling along a curved path where the angular velocity is a result of the curvature of the path and not independent of it, as in self-generated rotations. These limitations of efference copies during navigation highlight the importance of decomposing self-motion into T and R components based purely on optic flow. Indeed, our results show, for the first time that both translation and rotation estimates are derived from purely retinal cues and are represented jointly in the visual system.

3.3.2 Separability of T and R

Figure 3-1D represents a neural response where the translation component is perfectly extracted from optic flow even in the presence of rotations. However, rotations can manifest as either shifts in the translation tuning (Figure 3-1C) or gain fields (Figure 3-1E), or a combination of both (Figure 3-1F). If the representation of translations and rotations are independent of each other, they can be considered 'separable'. This would imply that the visual system perfectly decomposes optic flow into its T and R components and represents them jointly (like the simulation in Figure 3-1E). Such a joint and separable representation has certain advantages for flexibly decoding either T or R or both, depending on the navigational task. A separable representation would imply that a simple decoding strategy, such as marginalizing the responses over either T or R would provide an estimate of the necessary self-motion feature. If the representations are not separable, it may still be possible to estimate T and R, but would require more complex decoding strategies.

In contrast, the simulations in Figures 3-1C, F correspond to situations where the translation tuning curve is deformed (shifts) due to the added rotations – i.e. the two

representations are not completely separable. This inseparability manifests as large shifts in the translation tuning that vary systematically with rotation velocities. It should be noted that even for cells with relatively large shifts, the inseparable component of the joint tuning is small, as can be seen in the simulations in Figure 3-1 (simulated shift of 20° at rotation velocity of 15°/s). The observed shifts in tuning were on average even smaller (Figure 3-4). In addition to small shifts, the large bandwidth of VIP neurons and the sampling of the entire horizontal plane in the joint tuning curves results in inseparabilities that are hard to detect. Therefore, the most robust way of estimating these small, but potentially significant inseparabilities was to evaluate the systematic influence of rotation speed on the shifts in the translation tuning curve. Using this robust estimation of inseparabilities in the joint tuning curves, we found that a majority of the neurons (72% during SP, 74% during RP) showed no significant systematic shifts as a function of added rotations. These results all point to a largely separable encoding of T and R in a sub-population of VIP neurons.

It is also important to note that there were no significant differences in the shifts observed in the population of T-only cells and T&R cells, i.e. they were similarly separable. Hence, the errors associated with the translation representation (based on tuning curves) in the presence of rotations from T-only cells are not significantly different from T&R cells. Moreover, as shown in Figures 3-4, 6, the tuning shifts in both populations are small (~13.7°) and varied systematically only in a minority of cells (28%). This supports the hypothesis that the visual system can employ simple decoding strategies that rely on the separability of T and R (such as, marginalization), for both sub-populations.

3.4 Experimental Procedures

3.4.1 Subjects and Surgery

Three adult rhesus monkeys (*Macaca mulatta*), weighing 8-10kg, were chronically implanted with a circular molded, lightweight plastic ring for head restraint and a scleral coil for monitoring eye movements (see Gu et al., 2006; Fetsch et al., 2007; Takahashi et al., 2007 for more detail). Following recovery from surgery, the monkeys were trained to sit head restrained in a primate chair. They were subsequently trained using standard operant conditioning to fixate and pursue a small visual target for liquid rewards, as described below. All surgical and experimental procedures were approved by the Institutional Animal Care and Use Committees at Washington University and Baylor College of Medicine, and were in accordance with NIH guidelines.

The primate chair was affixed inside a field coil frame (CNC Engineering, Seattle, WA, USA) with a flat display screen in front. The sides and top of the coil frame were covered with a black enclosure that restricted the animals' view to the display screen. A three-chip DLP projector (Christie Digital Mirage 2000, Kitchener, Ontario, Canada) was used to rear-project images onto the 60 x 60 cm display screen located ~30cm in front of the monkey (thus subtending 90° x 90° of visual angle). Visual stimuli were generated by an OpenGL accelerator board (nVidia Quadro FX 3000G). The display had a pixel resolution of 1280×1024, 32-bit color depth, and was updated at the same rate as the movement trajectory (60 Hz). Behavioral control and data acquisition were accomplished by custom scripts written for use with the TEMPO system (Reflective Computing, St. Louis, MO, USA).

3.4.2 Stimuli and Task

Visual stimuli were presented during each trial and consisted of various combinations of eight heading directions in the horizontal plane (Figure 3-2A) and two rotational directions (leftward and rightward). The rotation velocity ranged from -15°/s (leftward) to 15°/s (rightward). The protocol consisted of either 3 rotation velocities (-15:15:15 °/s) or 7 rotation velocities (-15:5:15 °/s). The velocity profile of the rotations were trapezoidal and chosen such that the constant velocity portion of the trapezoid resulted in equivalent eye displacement over the duration of a trial. Therefore, the trial duration ranged from 1500ms for the fastest rotation velocity to 2500ms for the slowest (Figure 3-2B). The translation velocity profile was also trapezoidal with a maximum velocity of 24cm/s. When translations and rotations were presented together, the duration of the trial was dictated by the rotational component. For the pure translation stimuli, the trial duration was 1500ms.

The optic flow stimuli were generated using a 3D rendering engine (OpenGL) to accurately simulate combinations of observer translation and rotation. In the 3D cloud protocol, the virtual scene consisted of a cloud of dots that was 150cm wide, 100cm tall, 160 cm deep and had a density of 0.002 dots/cm^3. In order to maintain a constant volume of dots during the 27cm translation over the duration of a trial, the cloud was clipped in depth to range from 25cm to 125cm in front of the monkey at all times. The stimulus was rendered as a red-green anaglyph that the monkey viewed stereoscopically through red/green filters.

During each session, the monkey's eye position was monitored online using the implanted scleral search coil. Only trials in which the monkey's eye remained within a pre-determined eye window (see below) were rewarded with a drop of juice. Trials were aborted if the eye position constraints set by the eye window were violated.

The experimental protocol consisted of pure translations, combined translations and rotations, and pure rotations. The rotational component was introduced either by the monkey performing smooth pursuit eye movements or by simulating the rotational component using OpenGL. Therefore, the experiment had 5 types of visual stimuli: 1) Translation only, 2) Translations with smooth pursuit eye movements (real pursuit), 3) Translations with simulated eye rotations (simulated pursuit), 4) Real rotation-only, resulting from smooth eye pursuit, 5) Simulated rotation-only.

1) For the pure translation condition, the monkey fixated a visual target at the center of the screen and maintained fixation within a 2° eye window while the optic flow stimuli were presented. Optic flow stimuli simulated 8 headings within the horizontal plane, corresponding to all azimuth angles in 45º steps (Figure 3-2A). The pure translation stimuli were rendered by translating the OpenGL camera along one of the 8 headings with the velocity profile corresponding to the 1500ms duration shown in Figure 3-2B.

2) For the real pursuit (RP) condition, the animal actively pursued a moving target while the translational optic flow stimuli were presented on the display screen. Hence, both retinal and extra-retinal cues to rotation were present in these trials. The translation stimuli simulated 8 headings sampling the horizontal plane, in 45º azimuth angle steps (Figure 3-2A). A rightward rotation trial started when the fixation target appeared to the left of center. Once the monkey fixated this target (within 1000ms), it moved to the right following a trapezoidal velocity profile corresponding to the rotation speed as described above (Figure 3-2B). Analogously, leftward pursuit trials began with the target appearing on the right and moving leftward. The monkey was required to pursue the moving visual target and maintain gaze within a 4° eye window during the acceleration and deceleration periods of the trapezoid.

During the constant velocity phase, the monkey was required to maintain gaze within a 2° window around the visual target.

3) For the simulated pursuit (SP) condition, optic flow stimuli were presented that accurately simulated combinations of the same 8 headings with leftward or rightward rotations at different velocities, while the monkey fixated at the center of the screen (2° window). These stimuli were rendered by translating and rotating the OpenGL camera with the same trapezoidal velocity profiles of the moving target in the RP condition. This ensured that the retinal optic flow patterns in the RP and SP conditions were identical (assuming accurate pursuit in the RP condition). Therefore, rotational information was present only as retinal cues from the optic flow stimulus during these trials.

4) The real rotation-only condition consisted of the animal pursuing a visual target across a static field of dots. This resulted in both retinal and extra-retinal cues signaling the rotation component. Therefore, this condition was analogous to real pursuit, but without the translation component.

5) The simulated rotation-only condition consisted of the animal fixating at the center, while the optic flow stimuli simulated pure rotations. This resulted in only retinal cues signaling the rotation component (analogous to simulated pursuit).

Importantly, during the eye pursuit conditions, the optic flow stimuli were windowed with a software rendered aperture that moved simultaneously with the pursuit target. Thus, the area of the retina being stimulated during the eye pursuit trials remained constant over time, eliminating potential confounds from moving the stimulus across the receptive field over time. Furthermore, this ensured that the area of retinal stimulation was identical when the monkey was fixating or pursuing the visual target.

The protocol also included a blank screen during visual fixation and a static field of dots during fixation (1500ms) to measure the spontaneous activity and baseline visual response of the neurons, respectively (controls). Therefore, the three rotation velocity protocol consisted of 46 unique stimulus conditions: 8 translation only + 8 translation * (2 RP + 2 SP) + 2 real rotation-only + 2 simulated rotation-only + 2 controls. Similarly, the seven rotation velocity protocol consisted of 118 unique stimulus conditions: 8 translation only + 8 translation * (6 RP + 6 SP) + 6 real rotation-only + 6 simulated rotation-only + 2 controls.

3.4.3 Electrophysiological recordings

To record from single neurons extracellularly, tungsten microelectrodes (FHC; tip diameter, 3 µm; impedance, 1-3 MΩ at 1 kHz) were inserted into the cortex through a transdural guide tube, using a hydraulic microdrive. Neural voltage signals were amplified, filtered (400–5000 Hz), discriminated (Plexon Systems), and displayed on SpikeSort software (Plexon systems). The times of occurrence of action potentials and all behavioral events were digitized and recorded with 1ms resolution. Eye position was monitored online and recorded using the implanted scleral search coil. Raw neural signals were also digitized at a rate of 25 kHz using the Plexon system for off-line spike sorting.

VIP was first identified using MRI scans as described in detail in Chen et al. (2011). Electrode penetrations were then directed to the general area of gray matter around the medial tip of the intraparietal sulcus with the goal of characterizing the entire anterior-posterior extent of area VIP – typically defined as the intraparietal area with directionally selective visual responses (Colby et al., 1993; Duhamel et al., 1998). To determine direction selectivity, we presented a patch of drifting dots for which the size, position, and velocity could be

manipulated manually with a computer mouse. We used this mapping procedure to characterize the presence or absence of strong visual drive as well as the direction and speed selectivity of multi-unit and single-unit activity. At each location along the anterior–posterior axis, we first identified the medial tip of the intraparietal sulcus and then moved laterally until there was no longer a directionally selective visual response in the multi-unit activity.

During each experimental session, we inserted a single microelectrode into the region of cortex identified as VIP. Single unit action potentials were then isolated online using a dual voltage-time window discriminator. Within the region of gray matter identified as VIP, we recorded from any neuron that showed robust visual responses during our search procedure. Once a single unit was isolated, we ran the experimental protocol with all conditions randomly interleaved. Each stimulus condition was repeated at least 4, and usually 5, times. The three rotation velocity protocol was used for all neurons recorded from the monkey X (n = 38). In monkey E, 34 neurons were recorded using the 3 rotation protocol and 20 were recorded using the 7 rotation velocity protocol. All cells recorded from monkey R (n = 30) were recorded using the 7 rotation velocity protocol.

3.4.4 Analyses

Analysis of spike data and statistical tests were performed using MATLAB (MathWorks). Firing rates for each stimulus were calculated based on the neural response during the middle 750ms of each trail since the shortest constant velocity period was 750ms. This ensured that the tuning curves were based on data from identical durations despite the differences in total trial duration for different rotation velocities. This also ensured that the mean eye position during each trial was the same – straight ahead. The joint tuning curves for RP and SP were

generated using these measured firing rates. The joint tuning shown in Figures 3-3, 5 were generated using the firing rates calculated for each combination of translation direction and rotation velocity including 0°/s (i.e. translation only trials). The colormap was generated by representing higher firing rates with warmer colors and linearly interpolating the firing rates in between the difference translation, rotation combinations. The tuning curves shown on the marginal were conditioned on either the rotation velocity (top) or the translation direction (right).

To determine the effect of rotations on neural responses, the translation only tuning curve (0°/s rotation) was compared to the tuning curves measured at the different rotation velocities during both real and simulated pursuit. Since the 'shifts' expected in the tuning curves in the presence of rotations are complex and exhibit bandwidth changes as well as skew, we used a previously described method to measure the tuning curve deformations (Sunkara et al, 2014). Briefly, we used a 3-step analysis procedure. *Step 1*: We normalized the responses of the rotation added tuning curve to match that of the translation only tuning curve based on the maximum and minimum firing rate of the tuning curve (i.e. eliminated gain fields and vertical offsets). *Step 2*: Because the predicted effects of rotation are opposite for forward and backward headings, the tuning curves in the presence of rotations were split into heading ranges of 0-180° and 180-360° and linearly interpolated to a resolution of 1º. *Step 3*: For the half-curves that showed significant tuning (ANOVA, $p \leq 0.05$), we performed a shift analysis. The pure translation tuning curve was circularly shifted (in steps of 1º) to minimize the sum-squared error with each half of the RP/SP tuning curves. For neurons that were significantly tuned in all conditions and in both direction ranges, this analysis yielded two shift values for real pursuit and two shifts for simulated pursuit for each non-zero rotation velocity. In order

to quantify the transformation of heading tuning due to rotations, the two shift values were averaged to arrive at one shift value for real pursuit and one shift for simulated pursuit for each rotation velocity, for each cell. The leftward and rightward shifts at 15°/s were averaged together for the shifts reported in Figure 3-4; everywhere else in this paper, the leftward and rightward shifts were not averaged.

The 95% confidence intervals (CIs) for each of the shifts measured were calculated using a bootstrap analysis. Bootstrapped tuning curves for translation only, real pursuit, and simulated pursuit at different rotation velocities were generated by resampling responses with replacement. The offset, gain and shift calculations described above were performed on each one of 300 bootstrapped tuning curves to produce a distribution of shifts for each neuron and each rotation velocity from which the 95% CI was calculated by the percentile method. For the 7 rotation velocity protocol, the bootstrapped shifts as a function of the rotation velocity were fit with a line (Figure 3-5) and the 95% CI on the slope was determined using MATLAB's 'regress' function. Slopes with CI not overlapping 0 were determined to be significant (dark lines in Figure 3-6).

Modulation indices were calculated to quantify the rotation tuning of a cell during combined translation and rotation as well as rotation-only stimuli (Figure 3-7). For RP and SP joint tuning curves, first the preferred heading was determined based on the heading angle that elicited the largest response during pure translations. The modulation of the rotation tuning curve at this preferred heading was then evaluated using the equation [(max-min)/(max+min)]. The same equation was used to determine the modulation index during rotation-only stimuli, where the 0 rotation velocity response corresponded to the baseline visual response of the cell (static field of dots). The rotation tuning curves that were

significantly modulated (1-way ANOVA, $p \leq 0.05$) were also fit with Gabor functions for cells with 7 rotation velocities. The peak of the Gabor fit was reported as the preferred rotation velocity of the cell (Figure 3-8C).

Chapter 4 – Discussion

The visual system plays a key role in helping us navigate through the world. Optic flow or the movement of objects on the retina as we move through the world, carries complex, but useful information about self-motion. Since navigational tasks can be extremely varied, such as deciphering heading or tracking moving prey or estimating one's motion trajectory, it is important to decompose and represent both components of optic flow – i.e. translations and rotations. There are two strategies that the brain can implement to represent the T and R components of self-motion separately. First, non-visual cues, specifically efference copies of self-generated rotations (such as eye movements) can be used to estimate rotations. Second, retinal cues based on optic flow properties can be used to decompose the resultant optic flow into T and R.

A review of the current literature (Chapter I), shows that the there are two important, yet unanswered questions regarding the representation of translations and rotations during self-motion. First, efference copies of self-generated rotations have long been implicated in achieving a rotation-invariant perception of translation. More recent psychophysical results have suggested that the visual system may be capable of extracting the translation component of optic flow in the presence of rotations, based purely on retinal cues. However, conflicting results cast doubt on the extent to which retinal cues are useful in separating T and R. Importantly, there was no conclusive neurophysiological evidence that the visual system is capable of using only retinal cues (optic flow) to decompose self-motion into its T and R components. Therefore, it is unclear whether the visual system performs the computations necessary for estimating the T and R components of self-motion based only on visual cues or

if it relies on extra-retinal signals of rotation.

Secondly, most previous studies were limited to examining translation representation in the presence of rotations. This led to a prevalence of the idea that rotations are a feature of self-motion that needs to be discarded or compensated for. As a result, very little was known about how the visual system may represent rotations in the presence of translations, despite the importance of rotation estimates for many navigational tasks. While several studies examined the presence of efference copy signals in representing R, there was no clear characterization of rotation representation based on purely visual cues, especially during self-motion. An effective strategy for encoding self-motion would be to represent both the T and R components in a way that allows for flexibly estimating either T or R, or both. However, the presence of joint T and R tuning or the interaction between these two components had never been examined.

The two studies detailed in this document aim to address these two fundamental questions about how self-motion is represented in the visual system. Through single-unit electrophysiology in macaque area VIP, we show for the first time that: (1) translation representations are largely tolerant to rotations even in the absence of efference copies, suggesting that the visual system is indeed able to decompose self-motion into its translational and rotational components. (2) Joint and separable representations of T and R are present at the single-unit level in a sub-population of VIP neurons. Together, these results imply that the visual system can dissociate the T and R components of self-motion based purely on optic flow and jointly represent T and R in a separable manner.

These results have several implications in expanding our understanding of vision-based navigation, and opening up new avenues of research in the field. Our results suggest

that the visual system is capable of performing sophisticated computations on optic flow stimuli to decompose and represent self-motion. We also show that both 2D (dynamic perspective) and 3D (motion parallax) cues contribute to the resolving of optic flow into its T and R components. However, the exact neural mechanism by which this computation is performed deserves further investigation. Moreover, computational vision scientists have several other theories about which specific aspects of optic flow are most important in this computation that can be directly tested in the parietal cortex using appropriate stimuli. In addition to the encoding of T and R, it is also important to study how the information may be decoded downstream. We hypothesize that the joint and separable representation of T and R can allow for flexibly decoding the feature of importance depending on the behavioral task. Experiments evaluating the neural response during active behavior may reveal a lot about how the visual network accomplishes the complex navigational tasks we perform on a daily basis. In addition to perceiving our instantaneous motion, we often need to predict our trajectories to navigate successfully. For instance, the instantaneous optic flow during path-independent rotations (as studied here) and curvilinear motion are highly similar, but the trajectory estimates of the two may be very different. Thus, it is important for the brain to resolve this ambiguity. As discussed in Chapter I, despite the similarity in instantaneous optic flow, the temporal evolution of the flow patterns are distinct for the two types of motion. Thus, further investigation into the temporal dynamics of how the visual system processes optic flow would be of interest to scientists interested in human as well as robotic navigation.

Chapter 5 – References

Andersen RA (1997) Neural mechanisms of visual motion perception in primates. Neuron 18:865-872.

Avillac M, Ben Hamed S, Duhamel JR (2007) Multisensory integration in the ventral intraparietal area of the macaque monkey. J Neurosci 27:1922-1932.

Avillac M, Deneve S, Olivier E, Pouget A, Duhamel JR (2005) Reference frames for representing visual and tactile locations in parietal cortex. Nat Neurosci 8:941-949.

Banks MS, Ehrlich SM, Backus BT, Crowell JA (1996) Estimating heading during real and simulated eye movements. Vision Res 36:431-443.

Bell C (1823) On the Motions of the Eye, in Illustration of the Uses of the Muscles and Nerves of the Orbit. Philosophical Transactions of the Royal Society of London 113:166-186.

Bradley DC, Maxwell M, Andersen RA, Banks MS, Shenoy KV (1996) Mechanisms of heading perception in primate visual cortex. Science 273:1544-1547.

Bremmer F, Kubischik M, Pekel M, Hoffmann KP, Lappe M (2010) Visual selectivity for heading in monkey area MST. Exp Brain Res 200:51-60.

Bridgeman B (1995) A review of the role of efference copy in sensory and oculomotor control systems. Ann Biomed Eng 23:409-422.

Bridgeman B (2007) Efference copy and its limitations. Computers in biology and medicine 37:924-929.

Britten KH (2008) Mechanisms of self-motion perception. Annu Rev Neurosci 31:389-410.

Chang SW, Snyder LH (2010) Idiosyncratic and systematic aspects of spatial representations in the macaque parietal cortex. Proc Natl Acad Sci U S A 107:7951-7956.

Chen A, DeAngelis GC, Angelaki DE (2011) Representation of vestibular and visual cues to self-motion in ventral intraparietal cortex. J Neurosci 31:12036-12052.

Chen X, Deangelis GC, Angelaki DE (2013) Eye-centered representation of optic flow tuning in the ventral intraparietal area. J Neurosci 33:18574-18582.

Chen X, DeAngelis GC, Angelaki DE (2014) Eye-centered visual receptive fields in the ventral intraparietal area.

Cheng JC, Li L (2012) Effects of reference objects and extra-retinal information about pursuit eye movements on curvilinear path perception from retinal flow. J Vis 12.

Colby CL, Duhamel JR, Goldberg ME (1993) Ventral intraparietal area of the macaque: anatomic location and visual response properties. J Neurophysiol 69:902-914.

Crowell JA, Andersen RA (2001) Pursuit compensation during self-motion. Perception 30:1465-1488.

Crowell JA, Banks MS, Shenoy KV, Andersen RA (1998) Visual self-motion perception during head turns. Nat Neurosci 1:732-737.

Cullen KE (2004) Sensory signals during active versus passive movement. Curr Opin Neurobiol 14:698-706.

Duffy CJ, Wurtz RH (1995) Response of monkey MST neurons to optic flow stimuli with shifted centers of motion. J Neurosci 15:5192-5208.

Duffy CJ, Wurtz RH (1997) Planar directional contributions to optic flow responses in MST neurons. J Neurophysiol 77:782-796.

Duhamel JR, Colby CL, Goldberg ME (1998) Ventral intraparietal area of the macaque: congruent visual and somatic response properties. J Neurophysiol 79:126-136.

Duhamel JR, Bremmer F, Ben Hamed S, Graf W (1997) Spatial invariance of visual receptive fields in parietal cortex neurons. Nature 389:845-848.

Ehrlich SM, Beck DM, Crowell JA, Freeman TC, Banks MS (1998) Depth information and perceived self-motion during simulated gaze rotations. Vision Res 38:3129-3145.

Fetsch CR, Wang S, Gu Y, Deangelis GC, Angelaki DE (2007) Spatial reference frames of visual, vestibular, and multimodal heading signals in the dorsal subdivision of the medial superior temporal area. J Neurosci 27:700-712.

Fukushima K, Yamanobe T, Shinmei Y, Fukushima J (2002) Predictive responses of periarcuate pursuit neurons to visual target motion. Exp Brain Res 145:104-120.

Gellman RS, Fletcher WA (1992) Eye position signals in human saccadic processing. Exp Brain Res 89:425-434.

Gibson JJ (1950) The perception of visual surfaces. Am J Psychol 63:367-384.

Graziano MS, Andersen RA, Snowden RJ (1994) Tuning of MST neurons to spiral motions. J Neurosci 14:54-67.

Grigo A, Lappe M (1999) Dynamical use of different sources of information in heading judgments from retinal flow. J Opt Soc Am A Opt Image Sci Vis 16:2079-2091.

Gu Y, Watkins PV, Angelaki DE, DeAngelis GC (2006) Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. J Neurosci 26:73-85.

Kaminiarz A, Schlack A, Hoffmann KP, Lappe M, Bremmer F (2014) Visual selectivity for heading in the macaque ventral intraparietal area. J Neurophysiol.

Kim HR, Angelaki DE, DeAngelis GC (Under revision) A novel role for visual perspective cues in the neural computation of depth. Nature Neuroscience.

Klier EM, Angelaki DE, Hess BJ (2005) Roles of gravitational cues and efference copy signals in the rotational updating of memory saccades. J Neurophysiol 94:468-478.

Koenderink JJ, van Doorn AJ (1976) Local structure of movement parallax of the plane. Journal of the Optical Society of America 66:717-723.

Koenderink JJ, van Doorn AJ (1981) Exterospecific component of the motion parallax field. J Opt Soc Am 71:953-957.

Koenderink JJ, van Doorn AJ (1987) Facts on optic flow. Biol Cybern 56:247-254.

Lewis RF, Gaymard BM, Tamargo RJ (1998) Efference copy provides the eye position information required for visually guided reaching. J Neurophysiol 80:1605-1608.

Li L, Warren WH, Jr. (2000) Perception of heading during rotation: sufficiency of dense motion parallax and reference objects. Vision Res 40:3873-3894.

Li L, Warren WH, Jr. (2002) Retinal flow is sufficient for steering during observer rotation. Psychol Sci 13:485-491.

Li L, Warren WH, Jr. (2004) Path perception during rotation: influence of instructions, depth range, and dot density. Vision Res 44:1879-1889.

Li W, Matin L (1992) Visual direction is corrected by a hybrid extraretinal eye position signal. Ann N Y Acad Sci 656:865-867.

Longuet-Higgins HC, Prazdny K (1980) The interpretation of a moving retinal image. Proc R Soc Lond B Biol Sci 208:385-397.

Mineault PJ, Khawaja FA, Butts DA, Pack CC (2012) Hierarchical processing of complex motion along the primate dorsal visual pathway. Proc Natl Acad Sci U S A 109:E972-980.

Mullette-Gillman OA, Cohen YE, Groh JM (2009) Motor-related signals in the intraparietal cortex encode locations in a hybrid, rather than eye-centered reference frame. Cereb Cortex 19:1761-1775.

Newsome WT, Wurtz RH, Komatsu H (1988) Relation of cortical areas MT and MST to pursuit eye movements. II. Differentiation of retinal from extraretinal inputs. J Neurophysiol 60:604-620.

Page WK, Duffy CJ (1999) MST neuronal responses to heading direction during pursuit eye movements. J Neurophysiol 81:596-610.

Probst T, Krafczyk S, Brandt T, Wist ER (1984) Interaction between perceived self-motion and object-motion impairs vehicle guidance. Science 225:536-538.

Purkinje J (1823) Beobachtungen und versuche zur physiologie der sinne.

Rieger JH, Lawton DT (1985) Processing differential image motion. J Opt Soc Am A 2:354-360.

Rosenberg A, Angelaki DE (2014) Gravity influences the visual representation of object tilt in parietal cortex. J Neurosci 34:14170-14180.

Royden CS (1994) Analysis of misperceived observer motion during simulated eye rotations. Vision Res 34:3215-3222.

Royden CS, Banks MS, Crowell JA (1992) The perception of heading during eye movements. Nature 360:583-585.

Royden CS, Cahill JM, Conti DM (2006) Factors affecting curved versus straight path heading perception. Percept Psychophys 68:184-193.

Saunders JA, Niehorster DC (2010) A Bayesian model for estimating observer translation and rotation from optic flow and extra-retinal input. J Vis 10:7.

Schlack A, Hoffmann KP, Bremmer F (2003) Selectivity of macaque ventral intraparietal area (area VIP) for smooth pursuit eye movements. J Physiol 551:551-561.

Schmerler J (1976) The visual perception of accelerated motion. Perception 5:167-185.

Shenoy KV, Bradley DC, Andersen RA (1999) Influence of gaze rotation on the visual response of primate MSTd neurons. J Neurophysiol 81:2764-2786.

Shenoy KV, Crowell JA, Andersen RA (2002) Pursuit speed compensation in cortical area MSTd. J Neurophysiol 88:2630-2647.

Sommer MA, Wurtz RH (2002) A pathway in primate brain for internal monitoring of movements. Science 296:1480-1482.

Sommer MA, Wurtz RH (2008) Visual perception and corollary discharge. Perception 37:408-418.

Squatrito S, Maioli MG (1997) Encoding of smooth pursuit direction and eye position by neurons of area MSTd of macaque monkey. J Neurosci 17:3847-3860.

Swindale NV (1998) Orientation tuning curves: empirical description and estimation of parameters. Biol Cybern 78:45–56.

Takahashi K, Gu Y, May PJ, Newlands SD, DeAngelis GC, Angelaki DE (2007) Multimodal coding of three-dimensional rotation and translation in area MSTd: comparison of visual and vestibular selectivity. J Neurosci 27:9742-9756.

Tanaka K, Saito H (1989) Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. J Neurophysiol 62:626-641.

Tanaka K, Hikosaka K, Saito H, Yukie M, Fukada Y, Iwai E (1986) Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. J Neurosci 6:134-144.

Tomasi C, Shi J (1996) Image deformations are better than optical flow. Mathematical and Computer Modelling 24:165-175.

Upadhyay UD, Page WK, Duffy CJ (2000) MST responses to pursuit across optic flow with motion parallax. J Neurophysiol 84:818-826.

van den Berg AV, Brenner E (1994) Why two eyes are better than one for judgements of heading. Nature 371:700-702.

von Helmholtz Hermann (1924) Helmholtz's treatise on physiological optics. [Rochester, N.Y.: Optical Society of America.

von Holst E, Mittelstaedt H (1950) Das reafferenzprinzip. Naturwissenschaften 37:464-476.

Wallach H (1987) Perceiving a stable environment when one moves. Annu Rev Psychol 38:1-27.

Warren WH, Hannon DJ (1988) Direction of self-motion is perceived from optical flow. Nature 336:162-163.

Warren WH, Jr., Hannon DJ (1990) Eye movements and optical flow. J Opt Soc Am A 7:160-169.

Warren WH, Jr., Saunders JA (1995) Perceiving heading in the presence of moving objects. Perception 24:315-331.

Warren WH, Jr., Morris MW, Kalish M (1988) Perception of translational heading from optical flow. J Exp Psychol Hum Percept Perform 14:646-660.

Yang Y, Liu S, Chowdhury SA, DeAngelis GC, Angelaki DE (2011) Binocular disparity tuning and visual-vestibular congruency of multisensory neurons in macaque parietal cortex. J Neurosci 31:17905-17916.

Zhang T, Britten KH (2011) Parietal area VIP causally influences heading perception during pursuit eye movements. J Neurosci 31:2569-2575.

Zhang T, Heuer HW, Britten KH (2004) Parietal area VIP neuronal responses to heading stimuli are encoded in head-centered coordinates. Neuron 42:993-1001.