Integration of Head-Centric Optic Flow and Head Rotation signals account for the perception of surface tilt by the active observer

Giovanni Mancuso

Doctoral School in Cognitive and Brain Sciences

26th Cycle

To my family

Contents

List of Figures

Chapter 1

Introduction

1.1 The phenomenon of depth from motion

When we move about in the environment, the relative movement between us and the objects in the world generates a rich pattern of retinal motion: the Optic Flow (OF). The OF contains information about: (i) self-motion, (ii) moving objects, and (iii) the three-dimensional ($3D$) layout of the environment and therefore could potentially be used to control navigation (Warren [2003]). This manuscript is specifically concerned with the perception of the $3D$ shapes from motion information during active exploration of the environment. In particular, the thesis is focused on the local properties of the OF generated by the relative motion of an observer and a planar surface. From a geometrical standpoint, the planar surface is one of the simplest $3D$ structures nevertheless it is very important in the field of spatial vision as $3D$ objects can be described as arrangements of local planar surfaces. Thus, understanding how planar surfaces are perceived from the OF may help reveal how complex surfaces representation is constructed.

The phenomenon of depth perception from the OF was recognized more than a century ago and Hermann von Helmholtz is often acknowledged as one of the pioneers in this field. Helmholtz gives an intuition of the effectiveness of motion cues as a source of depth perception:

> Suppose, for instance that a person is standing still in a thick woods, where it is impossible for him to distinguish, except vaguely

> and roughly, in the mass of foliage and branches all around him what belongs to one tree and what to another, or how far apart the separate trees are, etc. But the moment he begins to move forward, everything disentangles itself, and immediately he gets an apperception of the material contents of the woods and their relations to each other in space, just as if he were looking at a good stereoscopic view of it (cited in Simpson 1993, pp. 35-36).

Another instance of the human ability of extracting $3D$ shape information from two-dimensional ($2D$) transformation of the image is the phenomenon of kinetic depth effect originally studied by Wallach and OConnel (1953). The shadow of a rotating wire-frame is projected onto a screen where the observer experiences the display. The impression of a three-dimensional shape is compelling when the object is rotating but in the absence of motion no depth can be seen. In biological and artificial models of vision, this ability of extracting the $3D$ structure and object shape from motion cues is referred to as *Structure from Motion* (SfM).

1.2 The human ability to extract depth from the motion

The study of the OF, like other areas of perception, includes many fascinating but puzzling problems. For instance, simultaneous observer and object motion generates a complex pattern of retinal stimulation as well as extra-retinal signals (vestibular, proprioceptive, efferent copies of motor commands). In order to accurately estimate object structure and motion, the OF must be parsed into components due to *self* and *object* motion. At every moment in time, in fact, the retinal motion contains two major sources: a flow resulting from self-motion, and a flow resulting from the movement of objects relative to the observer. The brain, however, contains no internal signals related to object motion. In engineering this problem is known as "the source separation" problem and it is well characterized by following example:

> Consider a soccer player running downfield to intercept a pass and head the ball toward the goal. This athlete must be able to accurately

judge the trajectory of the ball relative to the trajectory of his/her self-motion, in order to precisely time his/her head thrust to meet the ball (DeAngelis and Angelaki [2012]).

How this problem can be solved has been a matter of debate and ground for theoretical works. According to the classical approach, $3D$ perception from the OF is an inverse-inference and the goal of the perceptual system is to deduce from sensory evidence the $3D$ shape that most likely led to the perceptual experience. In order to achieve this goal, this approach postulates an inversion of the generative model of the OF. An example will clarify this point. Deriving the retinal stimulation given a $3D$ object geometry and motion is a fairly simple task:

$$S = f(W)$$

The function f relates the distal properties of the world W with the proximal input S. The function f is defined by the laws of physics (optics) and it is well understood. Perception however faces the opposite problem: derive the state of the world W (in our case the $3D$ geometry) from a given sensory input (the OF). We can write:

$$W = f^{-1}(S)$$

This is an ill-posed inverse problem as the same sensory input (the OF) could be generated by an infinite number of states of the world W (combinations of $3D$ geometry and motion). To overcome this problem, vision has been described with probabilistic terms in order to include prior knowledge as well as uncertainty about the state of the world (Yuille and Kersten [2006]). Suppose, that the retinae receive the OF of that is generated by the relative motion M of an observer in the layout of the environment L. The goal of a visual system, artificial or biological, is to estimate L from of and M. This goal can be re-framed, thanks to Bayes rule, as follows:

$$P(L|of, M) \propto P(of, M|L)P(L)$$

The likelihood function $P(of, M|L)$ rules out all the interpretations L that are not consistent with the of and the M. Then the prior knowledge of the state of the world $P(L)$ is intended to narrow down the remaining possible interpretations. In a further step, vision is believed to maximize the posterior probability of $L^* = argmax P(L|of, M)$.

In the context of the SfM example, the knowledge of the of is not sufficient for the recovery of the $3D$ structure. In fact, in order to specify the layout L the knowledge of the relative motion M is needed. Note that the parameter M depends both from object and observer's motion. Therefore, for a veridical interpretation of the OF, is not sufficient an estimate of the ego-motion component; the knowledge of the object's motion is also required. Nevertheless, according to the classical approach a veridical estimate of L can be achieved by using non-visual information and introducing some assumptions regarding the motions of the objects. In fact, according to the rigidity and stationarity assumptions, the visual system is believed to favor the perceptual solution that maximizes: (i) the most rigid transformations and (ii) select, among many alternatives, the most stationary one (most static object) in an allocentric (earth-fixed) reference frame. Provided that these assumptions are satisfied it follows that all the motion cues are due to the observer's motion, hence if the visual system had access to extra-retinal information the SfM would become a simple multisensory integration problem (Wexler et al. [2001b], Wexler et al. [2001a], van Boxtel et al. [2003], Wexler [2003], Wexler and van Boxtel [2005]). In other words, to derive the motion of the object it would be sufficient to subtract from the overall OF the component due to the motion of the observer.

In the last few years, empirical evidence has been accumulated against the classical approach. A theoretically different perspective has been advocated by other Authors (see for example: Domini and Caudek [2003a], Fantoni et al. [2010], Caudek et al. [2011], Fantoni et al. [2012]). This line of research suggested that the extra-retinal signals from ego-motion are seldom used in the $3D$ perceptual interpretation of the OF. These findings are not in agreement with the idea that the visual system uses extra-retinal signals to discard the part of the OF generated by the self motion in order to encode a veridical and world-centered representation of the $3D$ objects. Instead, the results are consistent with the view that the

visual system operates a heuristic analysis based only on some optical elementary components of the OF (the deformation - *def*) with little or no contribution of extra-retinal signals. According to this hypothesis, the strategy applied by the visual system derives the most likely perceptual solution given the magnitude of the *def*; however it doesn't necessarily guarantee a veridical solution to the SfM problem (Domini and Caudek [1999], Domini and Caudek [2003a]). In agreement with this line of research are also the empirical findings undermining both the assumption of stationarity and rigidity witnessed in several experiments by Fantoni et al. (2013, submitted).

1.3 Overview of the present work

Given these conflicting premises, it is reasonable to ask why different researchers found a differential effects of the extra-retinal signals on the $3D$ perceptual interpretation of the OF. It has been proposed (Fantoni et al. [2010], Fantoni et al. [2012]) that the classical approach focused on the affine properties, while the alternative approach was mainly concerned with the perception of Euclidean $3D$ properties. The present work enters this debate and proposes a new theory of the human ability to extract the tilt direction of planar surfaces from actively generated OF. We suggests that, when the relevant sensory input (i.e. the head-centric OF, see below) is carefully controlled, also affine properties such as the tilt direction of a planar surface, cannot be recovered in a veridical fashion. According to the model presented in the following chapters, the main input to the visual system is the *head-centric* OF as opposed to the *retinotopic* OF postulated in previous researches. Within the framework of the *head-centric* OF analysis we were able to derive the relationship between the orientation in depth of a planar surface (in this case the tilt direction) and the translational components of the OF. Translational components were previously thought to be *not informative* of the $3D$ shape of the object and therefore they have always been neglected from the computation of depth from the OF.

The distinction between previous approaches and our proposal is well described by the following example: Wexler et al (2001b, 2001a) for instance, by comparing Active and Passive trials, repeatedly reported an advantage of active

over passive vision as the proof of optimal integration between the retinal and extra-retinal signals. The passive condition consisted in a re-play of the *retino-topic* OF with no translational components. This was done because during active trials the eye keeps a steady fixation on the center of the stimulus, where the OF vanishes. In contrast, if we assume a head-centric frame of reference, then translations are included in the analysis of the OF. This fact rises the confounding between the role of the extra-retinal signals and the global translation of the head-centric OF. In other words, the veridical performance found in the active vision was a consequence of the presence of the extra-retinal signals, as advocated by Wexler et al. (2001b, 2001a), or it is the byproduct of the presence of the translational components of the head-centric OF? The answer to this question may have a profound influence in the field of the perception of $3D$ shapes from the OF.

It must be said that even the Authors that argued against the classical approach consistently found a minor contribution of the extra-retinal signals (Caudek et al. [2011]). This leaves open the possibility that the visual system has access and may use extra-retinal information for the perceptual analysis of the OF, at least to some extent. A closer look at the literature reveals that, as far as the $3D$ perception from the OF is concerned, only the linear translations of the head were studied (Wexler et al. [2001b], van Boxtel et al. [2003], Fantoni et al. [2010], Caudek et al. [2011]) whereas, to our knowledge, none studied the contribution of the angular rotations. Therefore the reported confusing results might be due to differential effects of translations and rotations of the head. These two types of head motion are sensed and processed by different receptors in the peripheral vestibular system that might have different sensitivity. However, none systematically studied the influence of these two signals in the perceptual interpretation of the $3D$ from the OF.

The aim of this manuscript is to systematically study the head-centric OF and its relations with the perceived $3D$ shapes. This goal is achieved by breaking down the problem into its basic constituents and then systematically analyzing them both theoretically and empirically. The thesis contains one theoretical chapter (Appendix) and four experimental chapters, each one specifically designed to test different aspects of the model:

- in the *Appendix* we formally derive our theory from the mathematical analysis of the OF. In a second step, we advocate the notion of head-centric OF by demonstrating the relation between the tilt direction with both (i) the translational components of the OF and (ii) the angular rotations of the head. Finally we uncover the predictions of our theory by framing the model into probabilistic terms.

- *Chapter 2* describes a typical active-vision experiment where the participants are asked to provide perceptual judgments about motion-defined $3D$ stimuli. The goal of this chapter is to compare the head-centric versus the retinotopic framework for the perceptual analysis of the OF. In particular, the experiment is carefully designed to disentangle the contributions of two variables that were confounded in previous versions of the experiment: (i) the linear ego-motion signals and (ii) the translations of the OF. The results clearly support the role of the head-centric as opposed to the retinotopic OF. The advantage of active over passive vision, previously attributed to the extra-retinal signals, is found to be associated with the translational components of the OF and not with availability of the linear ego-motion signals.

- In *Chapter 3* we further explore the head-centric OF model focusing on the difference between linear and angular head movements. Specifically, from our theoretical analysis, different outcomes are expected when the same translational components of the OF is induced by linear or angular motion of the head. Remarkably, the empirical results support the predictions of the model. This experiment not only supports our previous findings from *Chapter 2*, but it also point out the differential contribution of linear and angular extra-retinal signals. The perceptual interpretation of the OF relies on the angular, but not on the linear component of observer's ego-motion.

- In *Chapter 4* we try to quantify the difference between head translation and rotation in terms of bias and precision. In Experiment 1 we compared the amount of visual shift that can be tolerated when the observers are: (i) laterally translating or (ii) laterally rotating the head. In Experiment 2, static

participants were presented with a replay of the visual motion that they themselves generated during the Experiment 1. By keeping constant the visual stimulation in the two experiments, we were able to directly compare the performance of the observers when they were translating or rotating the head. Angular movements are associated with less biased and more precise estimates of object motion, whereas linear translation are biased and are associated with a more noisy representation of object motion.

- In *Chapter 5* we run a parametric study on the role of the translational components of the OF and how these translations are interacting with the gradients of the velocity field. The study featured one single passive condition where a wide variety of velocity gradients, translational velocities and directions of motion (and all possible combinations) are tested. The rationale of this chapter is to test the model when angular extra-retinal signals are not available. The results are highly consistent across all the conditions tested and show little modulation of the factor tested. When the modulation is indeed present, it outlines a complex balance between OF translations and the gradients of the velocity field for the recovery of surface tilt.

- The experimental Chapters are summarized in *Chapter 6*.

Chapter 2

Perceived Surface Tilt from Self-Generated Optic Flow in Different Reference Frames

Abstract

The three-dimensional structure of a planar surface is not fully specified by the Optic Flow (OF) generated by the relative motion between an observer and a planar surface. For instance, a moving observer who approaches a planar surface that rotates about the vertical axis (e.g., a rigid flag hinging from a pole) generates the same OF produced by a planar surface that rotates about a horizontal axis and it is viewed by a static observer. In spite of this ambiguity, previous studies reported that perceived surface orientation by the active observer is usually veridical (Wexler et al. [2001a]). These Authors postulate that the visual system interprets the OF by combining retinal measurements of image velocities with extra-retinal information about the observer's ego-motion. By introducing some assumptions in the interpretation process, veridical estimates of Euclidean $3D$ structure can, in principle, be derived from the OF. However, recent empirical findings hinder the biological plausibility of such approach by showing that perceived $3D$ structure is not veridical and that ego-motion signals are seldom used in the perceptual interpretation of the OF (Fantoni et al. [2010], Caudek et al. [2011]). Here we test an alternative interpretation based on a probabilistic model that takes as the relevant input: (a) the head-centric OF and (b) the angular motion of the head (while ignores the linear translations). An implication of our model is that perceived orientation undergoes a 90° flip whenever a translational motion, in a direction parallel to the surface axis of rotation, is added to the OF. In the previous experiment, the head-centric OF always contained translational motion when the observer moved towards the stimulus display, due to the natural rotations and translations of the head. In

the present experiments, we tested our alternative explanation by asking observers to judge surface orientation in two main conditions: (1) when the axis of rotation was tethered to a coordinate system centered on the observers head, so as to eliminate the translational components of the head-centric OF or vice versa (2) when a random-dot planar surface rotated about a stationary axis. The results are consistent with the predictions of our model. Perceived surface orientation in (1) is ambiguous since there are no translational components of the OF and in (2) is veridical. A similar pattern of results is found when the same OFs, generated by the observer's movements, were replayed to a static observer.

Fantoni, C., Mancuso, G., Caudek, C., and Domini, F. (2012) Linear ego-motion signals are mostly ignored in the interpretation of the self-generated optic flow. doi:10.1167/12.9.1046 *Journal of Vision* August 13, 2012 vol. 12 no. 9 article 1046

2.1 Introduction

2.1.1 Background

Moving observers can perceive the three-dimensional ($3D$) shape of objects from the Optic Flow (OF) (Gibson [1950], Koenderink and van Doorn [1975], Koenderink and van Doorn [1978], Ullman [1979], Longuet-Higgins and Prazdny [1980], Longuet-Higgins [1984]). This is a remarkable ability because the OF is often ambiguous: the same motion pattern is consistent with multiple perceptual interpretations. In the case of a moving observer, for instance, the brain faces a "source separation" problem (DeAngelis and Angelaki [2012]): in order to recovery the *veridical* object structure and motion, the visual system must factor out the components of the OF that are produced by the motion of the observer and retain the components due to the object motion. This goal is believed to be achieved by an "inverse geometry" analysis: (i) the OF is processed in conjunction with an estimate of the observer's motion (computed thanks to the extra-retinal signals) and then (ii) some assumptions (i.e. rigidity and stationarity) about the motion of the objects are introduced in the perceptual interpretation process. Provided, that these two hypothesis are satisfied, the veridical $3D$ shape can be recovered by the moving observers.

The research on active shape perception however, led to somewhat conflicting results. On one hand, several experiments have witnessed the advantage (in terms of precision and bias reduction) of active over passive vision: stimulus ambiguities are resolved when extra-retinal signals are available (Ullman [1979], Cornilleau-Peres et al. [2002], Wexler [2003], Wexler [2005], Colas et al. [2007]). On the other hand, another line of research demonstrated that the extra-retinal signals are rarely employed in the perceptual interpretation of the OF and the retinal input, when carefully controlled, explains the vast majority of the data (Domini and Caudek [2003a], Fantoni et al. [2010], Caudek et al. [2011], Fantoni et al. [2012]). To reconcile this two views, it has been proposed (Fantoni et al. [2010], Fantoni et al. [2012]) that the first studies focused mainly on the recovery of the *affine* properties of the objects such as the tilt of planar surfaces. Whereas, the second list of studies was primarily concerned with the perception of slant for

which a computation of the *Euclidean* $3D$ properties is required.

2.1.2 The present experiment

The present study enters this debate by re-examining the role of the extra-retinal signals in the perception of affine $3D$ properties. Contrary to what has been proposed previously, we advocate the idea that if the relevant input variables (i.e. head-centric OF, see below) are carefully controlled, also affine properties cannot be recovered in a veridical fashion. In order to achieve this goal we first review a previous experiment done by Wexler et al. (2001a). Then we propose a different interpretation based on a new model for the perception of surface tilt from the OF. Finally, the model's predictions are tested on the same perceptual ambiguities described by Wexler et al. (2001a).

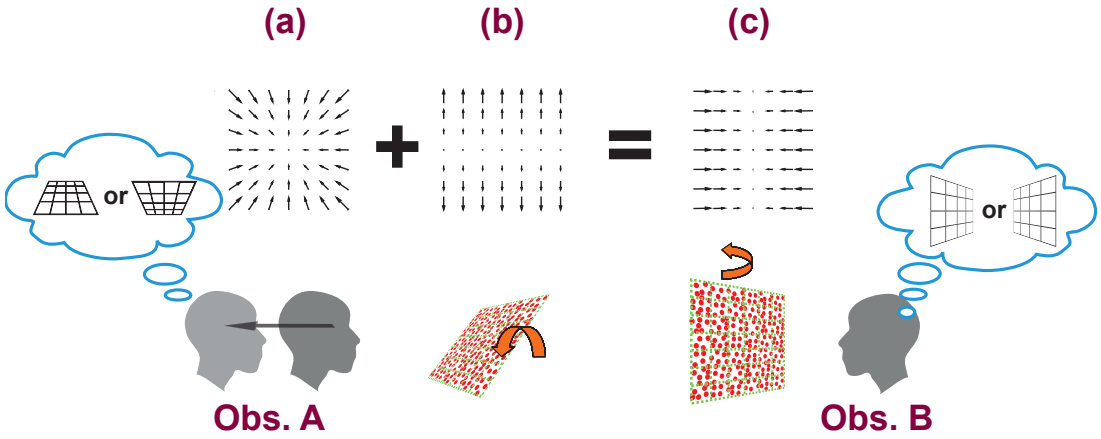

Figure 2.1: **Stimulus tilt ambiguity.** The bottom part of the panel shows the active observer (Obs. A) that is moving toward the stimulus while the surface is rotating. The upper part of the sketch shows the retinal flow: the expanding OF (**a**) is generated by the observers motion while the vertical expansion (**b**) is due to objects rotation. The combination of these two flows generates the OF depicted on the upper right (**c**). The resulting flow (**c**) is perceived: (i) in a veridical fashion in active vision by the Obs. A while (ii) in passive vision (Obs. B) the perception of the opposite tilt is favored. Please note that the same ambiguity can be achieved if the simulated surface is slanted and rotating about the vertical axis. In that case the overall resulting retinal flow would be constituted of a vertical gradient of the velocity field (as if **b** and **c** were interchanged). Adapted from Wexler et al. 2001a.

The experiment conducted by Wexler et al. (2001a) was divided into active and passive trials. In the active trials, the observer is approaching a slanted rotating surface; the amount of surface rotation and the translation in depth of the observer are adjusted in such a way that the induced OF is consistent with a different surface that is slanted and is rotating about a different axis. Figure 2.1 shows the basic outline of the active trials: the participant moves backward relative to the object and therefore generates a contracting OF (**a**); this translation in depth is coupled with the rotation of the object which in turn generates a pure vertical expansion (**b**); the combination of these two flows generates an OF that is consistent with a surface with a different tilt and a different axis of rotation (**c**). Despite this ambiguity of the retinal input, the participants perceived the correct (simulated) tilt in the active trials, as if the ego-motion component of the OF (figure 2.1 **a**) was disregarded and the perceptual interpretation was based only on the OF generated by the object motion (i.e. figure 2.1 **b**). In contrast, when the same OF shown in figure 2.1 **c**, was presented to passive observers, instead of perceiving the correct (simulated) surface, the observers perceived a surface with the opposite tilt. This finding was taken as the proof that extra-retinal signals are employed to disambiguate ambiguous OF.

Here we propose a new approach, developed in further details in the Appendix A, for the perception of surface tilt from the OF. Our model relies on two principles: (i) the analysis of the OF is conducted in a head-centric reference frame and (ii) the visual system doesn't make use of the linear ego-motion signals while it incorporates the angular ego-motion signals in the perceptual interpretation of the OF. Our head-centric framework has a series of important implications. Consider the case described in figure 2.2. A planar surface, which is rotating about an axis of rotation, is also moving laterally relative to a stationary observer. The relative motion between the object and the observer generates a translational components of the head-centric OF that was previously neglected. It was neglected because a retinotopic frame of reference was assumed; in fact, the eye tracks the motion of the stimulus and keeps a steady fixation on the center of the stimulus where the OF vanishes. Therefore, in this case, there are no translational components. The translational components are very important in our approach. Indeed, in the Appendix A we showed that the OF gradients specify the relationship between

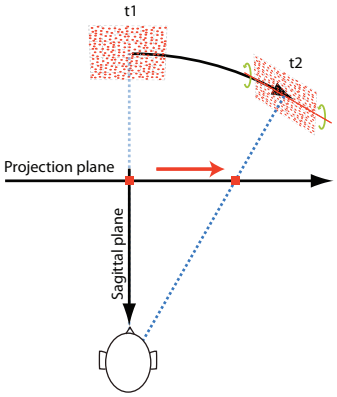

Figure 2.2: **The head-centric frame of reference.** The rotating planar surface is moving relative to the observer. The computation of the OF depends on the frame of reference that we adopt. According to the head-centric account a translational components of the OF is induced by the relative motion. According to the retinotopic approach, no translational components are present in the OF as the eye precisely tracks the motion of the object and keeps a firm fixation where the OF vanishes.

the tilt (τ) and the direction of the observer translation (θ) in the frontal plane. The angle θ in turn can be estimated knowing the *translational components* of the OF (a_1 and a_4), the gradients of the velocity field (a_2 and a_6) and the angular motion of the head about the horizontal and vertical axes (ω_{Rx} and ω_{Ry}):

$$\theta = \arctan\left[\frac{\mid a_4 + \omega_{Rx} \mid + a_6}{\mid a_1 - \omega_{Ry} \mid + a_2}\right] = -\tau \tag{2.1}$$

In the following we will examine an example in order to understand how the knowledge of the direction of self motion (θ) can constitute a solution for the extraction of the surface tilt. Given the OF depicted in figure 2.1 **c**, there are, at least, two possible tilt interpretations: (i) a horizontal tilt ($\tau = 0\,^\circ$, figure 2.1 **c**) or (ii) a vertical tilt associated with a sagittal translation ($\tau = 90\,^\circ$, figure 2.1 **b**). So, if we think about this problem in probabilistic terms, the parameter τ has two, equally likely, peaks ($L_{Gradients}(\tau)$). The angle θ also defines a likelihood distribution, $L_\theta(\theta)$. However, since $\theta = -\tau$ then $L_\theta(\theta) = L_\theta(-\tau)$. Therefore, we can obtain the likelihood of the tilt direction $L(\tau)$, by multiplying $L_{Gradients}(\tau)$ and $L_\theta(-\tau)$ (see figure 2.3). In summary, we are proposing that the perceived tilt

is derived through the computation of the angle θ by combining the head-centric OF with the angular movements of the head.

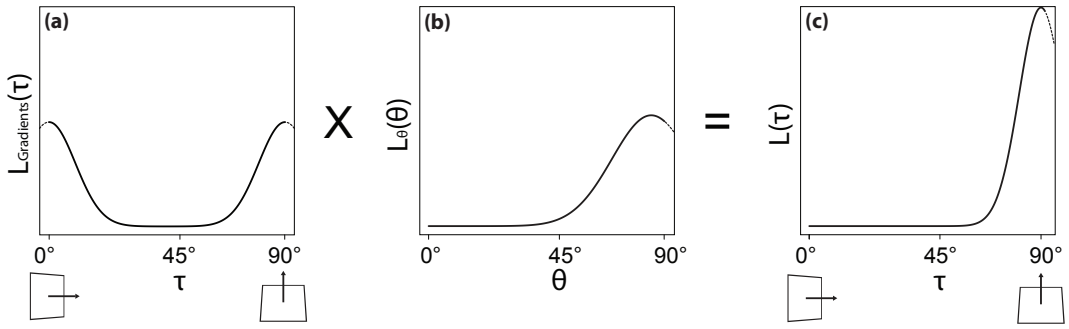

Figure 2.3: **Hypothetical computation of surface tilt from the head-centric OF.** The perceptual interpretation of the surface tilt from the OF can be expressed in terms of Likelihood distributions. **Panel a** represents the likelihood distribution of the tilt direction given the gradients of the OF, $L_{Gradients}(\tau)$. The two peaks represent the ambiguity of the OF. **Panel b** represents the likelihood distribution of the angle θ, $L_\theta(\theta)$. Therefore the likelihood that τ has induced the observed OF when head linear ego-motion is ignored, is represented in **Panel c** as the product of two previous likelihood distributions. The ambiguity is now vanished thanks to the translational components of the OF.

Our approach applies to active experiments as well. During active movements of the head the eye keeps a steady fixation on the center of the stimulus, where the OF vanishes. The retinal input therefore lacks of translational components. In contrast, the head-centric OF *does* contain translations since the head is moving relative to the object. This means that in previous experiment, during the passive condition, when stationary observers were presented with the replay of the *retinal* OF, important information available in the head-centric OF (i.e. translational components), were missing.

2.1.2.1 Predictions of the model

The goal of this chapter is to compare the head-centric versus the retinotopic framework for the perception of surface tilt from the active and passive OF. Figure 2.4 illustrates our purpose with an example of how our head-centric framework applies to Wexler's experiment. More specifically, the experiment is carefully designed to disentangle the contributions of two variables that were confounded in

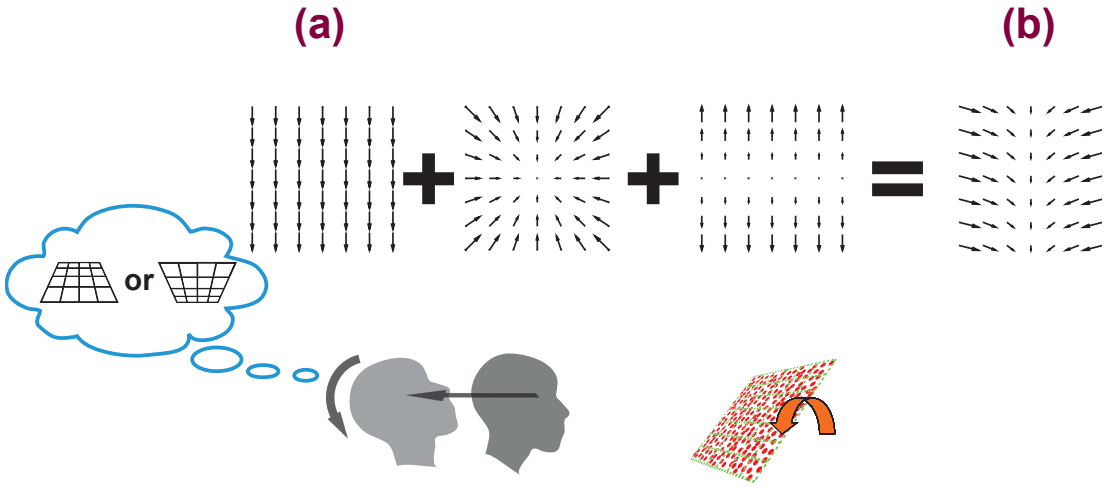

Figure 2.4: **Wexler's experiment revised within the head-centric framework.** The rotations of the head induce translatory components of the head-centric OF (**a**) that was previously neglected. According to our head-centric model, in order to make a fair comparison between active and passive vision, this translatory component should be included in the passive block (**b**).

previous versions of the experiment: (i) the linear ego-motion signals and (ii) the translations components of the OF. Therefore, in the present experiment, we asked our participants to report the perceived surface tilt in two main conditions: (i) when the axis of rotation was tethered to a coordinate system centered on the observers head, so as to eliminate the translational components of the head-centric OF or vice versa (ii) when a random-dot planar surface was allowed to translate relative to the observer. These conditions were run both in active and in passive vision. The manipulation of these variables allowed us to make counterintuitive predictions that cannot be reconciled with the previous approach:

- tilt perception should *not* be accurate for active observers if the OF is deprived of the translational components. Indeed, OF translational components are diagnostic of linear self-motion. According to previous proposals instead, this manipulation shouldn't affect perceived tilt as long as extra-retinal signal are available to disambiguate the interpretation of the OF.

- when the translational components of the head-centric OF are artificially enhanced, the tilt perception should be biased by the combination of the

translations and the gradients of the velocity field. This should be true both for active and passive vision. In contrast, this manipulation should not be effective according to the retinotopic account as no translational components are present in the retinal flow.

The novel result of the present investigation is that the advantage of active over passive vision, previously attributed to the extra-retinal signals, is found to be associated mainly, although not exclusively, with the translations of the head-centric OF and not with availability of the linear ego-motion signals. The results demonstrate also that human observers employed a perceptual strategy that doesn't necessarily guarantee the recovery of the veridical affine properties. Not even when, in principle, all the information for a correct analysis of the OF are available. Instead, the perceptual interpretation can be biased when the relevant input (i.e. the angle θ) is manipulated.

2.2 Methods

2.2.1 Participants

Nine participants (age 20 to 36, 5 males) took part in the experiment. All had normal or corrected-to-normal vision and signed a consent form as directed by the local Ethical Committee.

2.2.2 Apparatus

A detailed description of the viewing apparatus can be found in Fantoni et al. [2010],Caudek et al. [2011] and Fantoni et al. [2012], here we provide only a brief report. Stimuli were presented on a 19-inch CRT monitor (ViewSonic 9613, $19W$) which was viewed through a high-quality front-silvered mirror (Figure 2.5).
This arrangement produced a viewing distance (i.e., distance between the pupil and the center of the display reflected by the mirror) of 568.5 mm relative to the right eye of an observer at rest. The translational displacements and orientation of the participants head were recorded real-time by an Optotrak Certus system (Northern Digital Inc.) with two position sensors (with a spatial resolution below

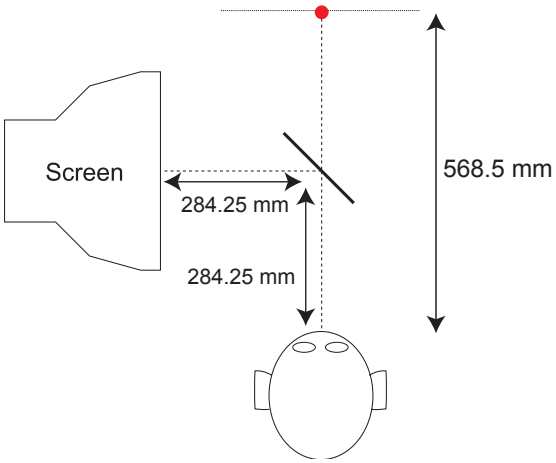

Figure 2.5: **Viewing Apparatus.** The visual stimuli are seen through a front-silvered mirror which is 45 ° slanted away from the monitor screen. This setting generates a viewing distance of 568.5 mm.

0.01 mm). The two position sensors recovered the $3D$ positions of three markers (infrared emitting diodes) which were attached to the participant's head during the active movements. Thanks to a Dell Precision $T3400$ $525W$ we were able to control the stimulus display and sample the tracker using a standard PCI card. This configuration allowed us to update in real time the position of the visual stimuli on the screen as the participants moved. A custom Visual C++ program supported by OpenGL Libraries and combined with Optotrak API routines was used for stimulus presentation/response recording.

2.2.3 Visual stimulation and the task

The participants were asked to judge the tilt direction of the simulated planar patch by pressing one of four possible alternative keys on the keyboard (left, right, up and down). As described below in greater details (see section 2.2.4), the experiment was performed in monocular viewing and the surface orientation was defined by the actively generated OF due to the relative motion between the observer and the concurrent rotation of the object.

We now describe the visual stimuli by using a reference frame described in the Appendix A with xy-plane coplanar with the monitor screen. In this setting the x-axis is pointing to the subjects right, the y-axis upward, and the z-axis away

from the subject and the origin is set in the center of the screen. The stimulus displays was made of 75 random anti-aliased red dots simulating the projection of a squared random dots planar surface (approximately $5\,^\circ \times 5\,^\circ$). On different trials the planar surface was slanted of about $\pm 45\,^\circ$ around the y or the x-axis. An Optotrak Certus System (Northern Digital Inc.) was employed to update the arrangement of the dots according to the observers head position and his/her orientation with respect to the simulated surface.

The dots were set at the maximal electron-gun value of $82cd/m^2$; the black background was $3cd/m^2$. Also, to remove texture (non motion) cues, the dots were randomly distributed on the projected image (not on the simulated surface) by imposing $z_0 = tan(g_x)x_0 + tan(g_y)y_0$, with x_0 and y_0 randomly selected in the range between ± 25 mm from the screen center, and g_x and g_y representing the amount of surface rotation around the y- and x-axes, respectively (see Appendix A). In terms of slant (σ) and tilt (τ) of a planar surface, g_y corresponds to $\sin(\tau)\tan(\sigma)$ and, g_x to $\cos(\tau)\tan(\sigma)$, so that $\sigma = \arctan\sqrt{\tan(g_x^2) + \tan(g_y^2)}$ and $\tau = \arctan(\tan(g_y)/\tan(g_x))$. In different trials, the tilt of the simulated surface was changed: either we set $g_x = \pm 45\,^\circ$ and $g_y = 0\,^\circ$ or viceversa $g_x = 0\,^\circ$ and $g_y = \pm 45\,^\circ$.

In the active condition, the stimulus rotation and the translation in depth of the observer (back-and-forward sagittal movements) were coupled thanks to the equation $\omega = \arccos(\cos(\sigma)\frac{f_z - e_z}{f_z}))$. At any given time-frame, the equation returns the appropriate angular speed of rotation of the surface (ω) given the instantaneous distance of the eye (e_z) from the fixation point (f_z) and the slant of the surface σ. This coupling between sagittal motion of the observer and surface rotation creates an ambiguous OF which is consistent with multiple surfaces that are slanted and rotating about different axis. The perceptual interpretation of this ambiguity is the subject matter of the present chapter. The stimulus displays were generated following a two-steps procedure: in a first stage the center of the stimulus was defined according to the experimental condition and then for each stimulus frame the dots of the simulated planar surface were projected onto the screen by using a generalized perspective pinhole model. In the present experiment, there were four different conditions: in (i) the stimulus was always centered and fixed in the center of the screen and the Center Of Projection (COP)

positioned at the observers cyclopean eye at rest; in (ii) the center of the stimulus was dynamically updated in order to be always centered and aligned with the line of sight of the observer while the camera was located at the exact position of the right eye; in (iii) and (iv) a global translational components was artificially added to the OF. The translations were proportional to the motion of the observer along the depth axis (40% of the instantaneous motion) and were oriented either rightward (Horizontal-translation) or upward (Vertical-translation). The COP was centered and aligned with the center of the object.

2.2.4 Procedure

The procedure included three phases: (I) instructions; (II) a practice in which participants familiarized with the task (III) the experimental task in which participants were asked to monocularly judge the tilt direction of the planar surface by pressing one of four buttons on the keyboard (corresponding to the following tilt directions: 0 °, 90 °, 180 ° and 270 °). This task was performed twice in two separate blocks: the first block corresponded to the Active Block condition and the second for the Passive Block condition.

In the Active Block four different types of OF were tested, either the surface: (i) was rotating about an earth-stationary axis of rotation or (ii) it was tethered to the head of the observer while it was also rotating about its own axis of rotation or (iii) it was tethered to the head of the observer while a global horizontal translatory component was added the OF, or finally (iv) it was tethered to the head of the observer while a global vertical translatory component was added to the OF. In a head-centric reference frame, these four manipulations led to four different conditions (see section 2.2.5 and figure 2.6):

- in (i) translatory components were present in the head-centric OF;
- while in (ii) where absent (as the stimulus was precisely aligned on the line of sight);
- similarly in (iii) and (iv) translatory components of the OF were artificially added by dynamically displacing the surface as the observer was moving.

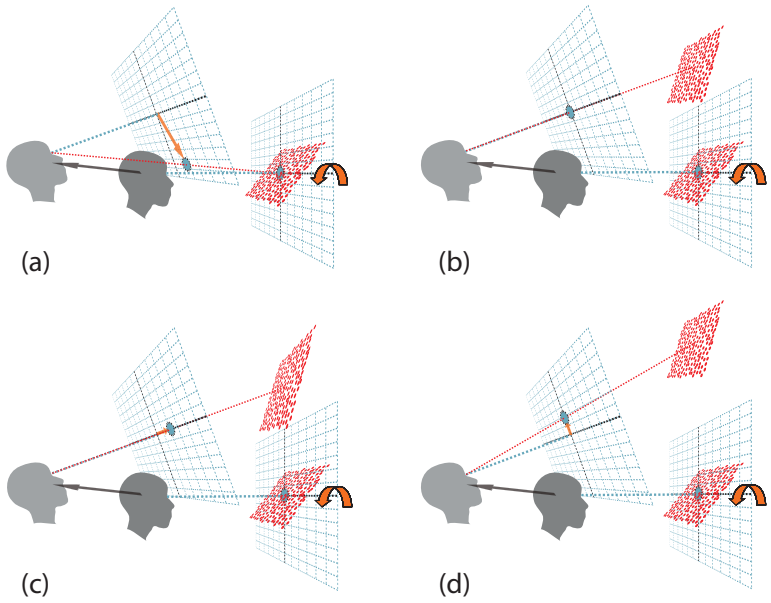

Figure 2.6: **Four Conditions tested in the experiment.** Each caption depicts two temporal intervals of an observer that is moving backward relative to a rotating surface. **Panel a** The surface is rotating about a stationary axis of rotation. Even though the observer keeps a steady fixation on the center of the image, in head-centric coordinates any movement that deviates from back-and-forward generates a translatory component of the OF (in the second frame the object is no longer centered). In contrast in **Panel b**, the image is always centered on the line of sight of the observer, and therefore no translational movements are present in this condition (in both frames, the image is centered in the reference frame). In **Panels c** and **d** an artificial translatory component is added to the OF (Horizontal and Vertical respectively).

The active block started only after 5 minutes of dark-adaptation, when the observers head was positioned at the ideal starting location (568.5 mm) from the monitor screen. At that point, the orientation and the position of the head were recorded and considered to be the starting values. If, at the beginning of a new trial, the head position and orientation were not near the original values (more than 50 mm and 5 $^\circ$ respectively) the participant was not allowed to start the trial until the experimenter located the head in the correct position. At the beginning of each trial, a red fixation mark was shown in the center of the screen and the observer was required to start the movement of his head backward (away from the screen). The participant was instructed to reverse the direction of head motion

after hearing a tone signaling a shift of $\pm 50\ mm$ relative to the starting position. After one cycle and a half, when the observers head passed through the starting position moving away from the screen, the fixation mark was replaced by the stimulus display. The simulated random-dot planar surface remained visible for two cycles and a half. The trial restarted if, during the translations, the subjects rotated the head more than 4° about the x, y, or z-axis.

In the Passive block, observer's head was located on a chinrest at the ideal distance of 568.5 mm. The left eye was blindfolded and the participants experienced the same head-centric OF sensed in the Active block (for a total amount of 80 trials). The head-centric OF was generated by replaying the $2D$ transformations generated by the corresponding active trials. In Passive block, four different scenarios were recreated (see section 2.2.5):

- when the subject was actively moving and the stimulus was centered on the cyclopean eye at rest, the passive head-centric OF contained all different sorts of translatory components relative to the sagittal and horizontal plane of the head. The relative motion was due to the normal swinging of the head: it is implausible that during sagittal motion the head remained completely still and always centered on the stimulus;

- when, during active vision, the stimulus was always aligned and centered on the line of sight of the observer, there were no translatory components and therefore the passive replay of the head-centric OF was that of a surface which was rotating about a stationary axis of rotation;

- when, a horizontal component was added to the head-centric OF, that component remained also in the Passive Block;

- similarly, when a vertical translation was added to the head-centric OF, that component also remained in the Passive Block.

Both experimental block were preceded by a training session to familiarize with the task: during the Active block observers were trained to maintain a constant velocity during head movement ($\approx 15 deg/sec$), whereas during the passive block, participant performed 20 preliminary trials.

2.2.5 Design

The combination of the factors employed resulted in a $2\times2\times4$ within-subjects design: simulated Gradient of the velocity field (Vertical and Horizontal - remember that these gradients were generated respectively by Horizontally and Vertically oriented surfaces through the process described in section 2.1.2), Head Motion Block (Active vs. Passive) and OF Translation (All, No-translation, Horizontal-translation and Vertical-translation) were the three independent variables. In both Blocks each subject viewed 10 presentations, in random order, of the eight different stimuli, for a total of 80 trials each block.

2.3 Results

We start the result section with a qualitative inspection of the behavioural results focusing on the Passive Block first (figure 2.7). The plot represents the proportions of horizontal responses (y-axis) as a function of the different levels of the OF Translation factor (x-axis) both for vertical (**Panel a**) and horizontal gradients (**Panel b**) of the velocity field. We will consider the *No-Translation* level as the baseline reference for all the other levels: in this case indeed, the stimuli were pure horizontal or vertical gradients with no other information in them. In the *No-Translation* level participants reported 0.10 ± 0.045 and 0.89 ± 0.047 proportion of "horizontal" responses for the vertical and horizontal gradients respectively. This result replicate previous findings from Wexler et al (2001a) where ambiguous OF led to a biased performance (remember that the veridical perception would be the exact opposite as the vertical and horizontal gradients originated respectively from horizontally and vertically oriented surfaces - see section 2.1.2).
In agreement with the predictions of our model, when all the translational components were included in the display (*All-Translation* condition) the performance showed a small, although consistent modulation: the proportion of "horizontal" responses for the vertical and horizontal gradients respectively became 0.27 ± 0.151 and 0.72 ± 0.155. Similarly, the proportion of "horizontal" responses was strongly modulated by the *Horizontal-Translation* level: both for the vertical and horizontal gradients, responses reached a ceiling effect (0.89 ± 0.047 and 0.93 ± 0.039).

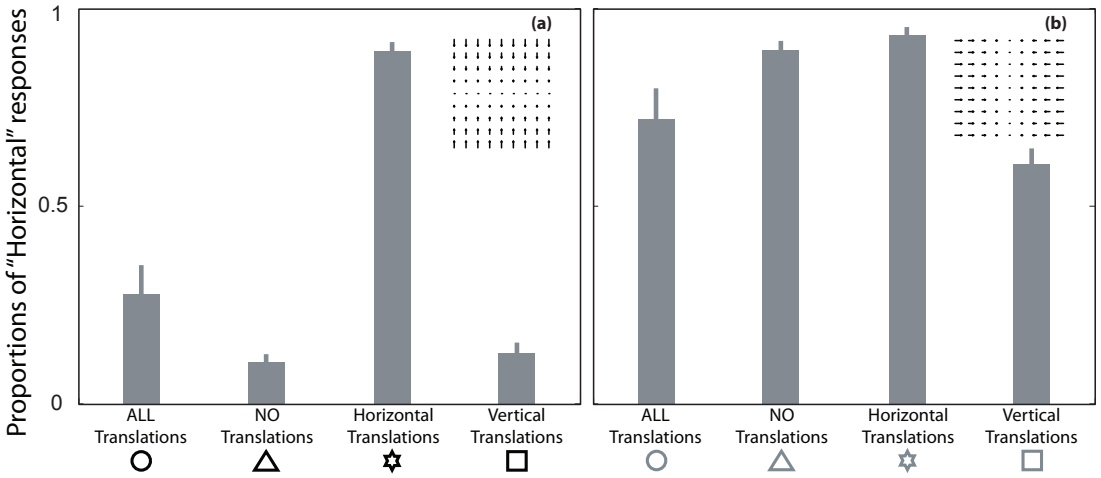

Figure 2.7: **The results of the Passive Block.** The proportions of "horizontal" responses is plotted as a function of the OF Translation factor both for the vertical (**Panel a**) and horizontal (**Panel b**) gradient of the velocity field. The empty symbols identify the levels of the OF Translation factor and the gradient of the OF (black-vertical, grey-horizontal) for further analysis in figure 2.10.

Instead, the effect of an artificial vertical translation (*Vertical-Translation* level), was less pronounced and overall asymmetric. The proportion of "horizontal" responses remained approximately the same (0.13 ± 0.051) for the vertical gradient. In contrast the vertical translation was less effective in the horizontal gradient and the proportions dropped by approximately 20% as opposed to 30% or more (0.60 ± 0.075).

In summary, the inspection of the Passive Block unravels a first important point. The perceived surface orientation is modulated by the presence/absence of translatory components of the OF. Moreover, this modulation is predicted by the direction of the translation: OF that are translating Horizontally are more likely to be perceived as Horizontally oriented surfaces (relative to the non-translating stimuli, namely *No-Translation*); similarly Vertically translating OF were perceived more often as Vertically oriented surfaces (compared to the non-translating stimuli). This pattern of results is predicted by our model in which the analysis of the OF is conducted in a head-centric frame of reference; while it remains unexplained according to a retinotopic model. In fact, in a retinotopic representation, the eye keeps a steady fixation on the center of the stimulus,

where the OF vanishes and there are no translatory components.

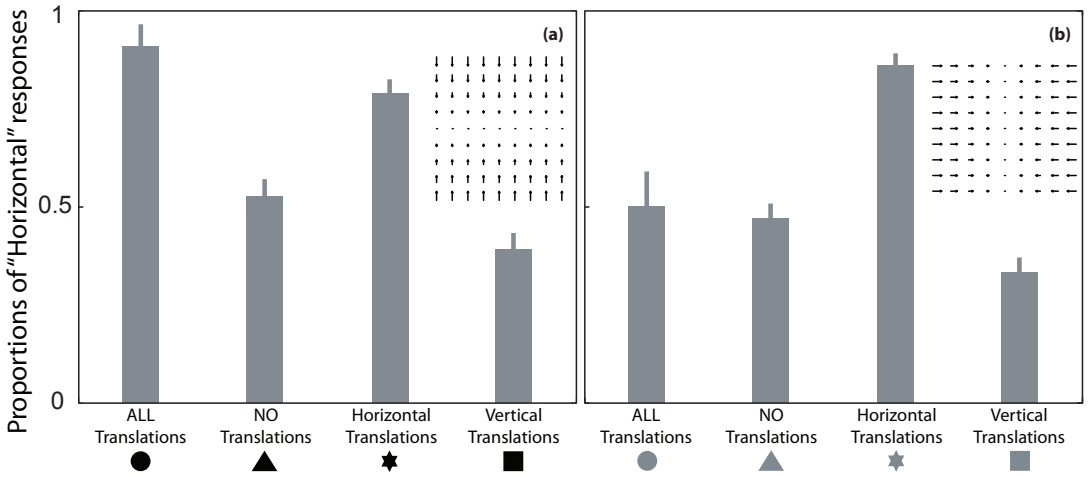

Figure 2.8: **The results of the Active Block.** The proportions of "horizontal" responses is plotted as a function of the OF Translation factor both for the vertical (**Panel a**) and horizontal (**Panel b**) gradient of the velocity field. The filled symbols identify the levels of the OF Translation factor and the gradient of the OF (black-vertical, grey-horizontal) for further analysis in figure 2.10.

Having described the Passive Block we can now review the Active Block always keeping the level *No-Translation* as a reference. Being this an active Block, the main focus now becomes the contribution of the extra-retinal signals and their interaction with the Translations of the OF.

Interestingly, when the extra-retinal signals are available but there are no translations of the OF (i.e. Active *No-Translation*) the performance was at chance level in both the vertical and the horizontal gradient conditions (respectively, 0.52 ± 0.078 and 0.47 ± 0.078). In line with our predictions, we observed an improvement when the translations were available (*All-Translation* level). This improvement however, was present only in the vertical gradient condition (proportion of "horizontal" responses: 0.90 ± 0.098) but not in the horizontal gradient condition (0.50 ± 0.09). We speculate that this asymmetry might be due to differential effects of the translations on the horizontal and vertical gradients.

The comparison between the *No-Translation* and *All-Translation* condition uncovers another important result of the present experiment: the trend toward the veridical perception cannot be explained by the knowledge of the forward

translation (estimated through the extra-retinal signals). Indeed, this information is available in both conditions. Instead, what is diagnostic is the presence/absence of the translational components of the OF.

Finally, when an artificial translation was added to the OF, the direction of translation was able to bias the perceptual judgments of the stimuli that we employed. The Horizontal translation led to the following proportions of "horizontal" responses: 0.90 ± 0.098 and 0.86 ± 0.053 for the vertical and horizontal gradients respectively. Similarly, the vertical translation prompted the perception of vertically oriented surfaces both for vertical and horizontal gradients: 0.39 ± 0.078 and 0.33 ± 0.073.

In order to capture this difference we run the following non-linear regression model that is based on our approach developed in the Appendix A:

$$y \quad = \phi_{\mu\sigma}(\theta_{w_{1\ldots n}}(a_1, a_4, a_2, a_6, \omega_{Ry}, \omega_{Rx})) \tag{2.2}$$

As explained in section 2.1.2.1 the perceived orientation of a planar surface (y - Horizontal or Vertical) is function of the angle θ. The angle θ, in turn, is expressed as a function of several input variables (OF translations: a_1, a_4, OF gradients a_2, a_6 and angular rotations of the head ω_{Ry}, ω_{Rx}). This parametrization allowed us to flexibly evaluate the relative importance (weights, $w_1, ..., w_n$) for the different sensory inputs, namely retinal and extra-retinal signals. It is important to remember the reader that this perceptual strategy doesn't include information about linear sagittal motion of the observer. The binary responses are then modeled with a probit link function, indeed the function ϕ represents the classical cumulative Gaussian distribution with a center (μ) and a scale (σ) parameter.

We estimated the model parameters in two separate steps. We begun with the analysis of the Passive Block and we obtained the weights associated with the retinal variables and the parameters of the Gaussian distribution. Then, in a second stage, we employed the estimated retinal parameters and the model to fit the Active data and extract the weights for the extra-retinal signals. Employing the procedure of model simplification by means of Akaike's Information Criterion

(AIC[1]) and Transformed Likelihood Ratio test (TLR[2]), we achieved the best trade-off between number of free parameters and explained variance. In the context of the present analysis of the Passive condition data set, the best model ($\chi^2_{(4)} = 28.93, p < 0.001$) included the standard deviation of Gaussian cumulative distribution[3] (σ) and four parameters for the retinal inputs (i.e. w_1, w_4 for the translations and w_{gx}, w_{gy} respectively for the horizontal and vertical gradient of the velocity field). Then, we employed the estimated parameters values to fit the Active condition data. In this case the goal was to estimate the remaining parameters associated with the extra-retinal signals (i.e.: $w_{\omega x}, w_{\omega y}$) and assess the predictive power of our model. The model perfomed significantly better when compared with a null model ($\chi^2_{(1)} = 11.83, p < 0.001$) and its predictive power can be inspected both in figure 2.9 and in figure 2.10.

Figure 2.9, represents the psychometric function relating the perceived orientation with the angle θ. This fit well predicts the performance of all the sixteen conditions tested in this experiment (2 Gradient directions $\times$ 2 Head Motion Blocks $\times$ 4 OF Translation).

The prediction of our model can be inspected also by looking at figure 2.10, **Panel a**. Filled and empty symbols refer respectively to the Active Condition and the Passive condition. The colors refer to the Vertical or Horizontal gradients of the OF stimuli (black and grey respectively). And, finally, different shapes are associated with different levels of the OF Translation factor: *All-translations* is represented by the circle, *No-translations* by the triangle, *Horizontal-translations* by the star and *Vertical-translations* by the square. The symbols are also displayed in the legend of figure 2.8 and figure 2.7.

The predictive performance of the model is remarkable as it captures most of the sixteen experimental conditions. There are however two points that behaved

[1]The corrected version of the Akaike's Information Criterion was employed Akaike [1974], Sugiura [1978], Hurvich and Tsai [1989].

[2]When nested models are compared within the framework of composite Hypothesis Testing, the ratio between the likelihood functions approximates a χ^2 distribution with *degrees of freedom* equals to the number of free parameters of the more general model minus the number of free parameters of the restricted model (see for instance, Kingdom and Prins [2010]).

[3]The mean of the distribution was found to be close to 45 ° and therefore was fixed at that value. Interestingly, the center of the psychometric function, the point of maximum uncertainty where the performance is at chance, corresponds to the angle 45 ° which is halfway between 0 ° and 90 °.

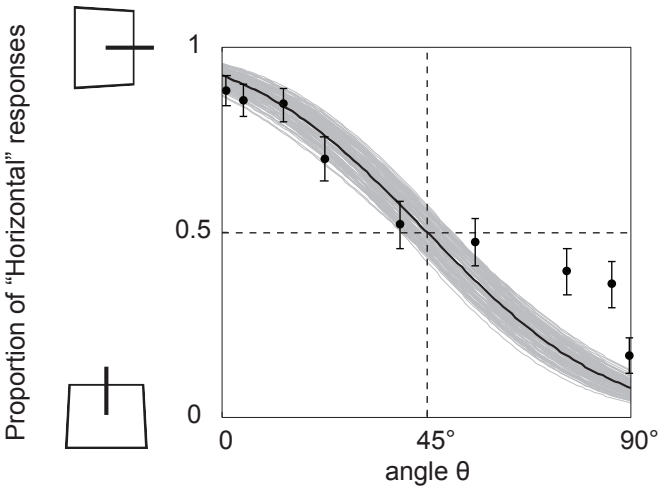

Figure 2.9: **The psychometric function fit.** The proportions of "horizontal" responses decrease as a function of the angle θ (light lines indicate uncertainty in the regression). The empirical proportions where calculated dividing the angle θ into nine equally dense bins. The fit shows another important property: the maximal uncertainty coincides with θ equal to 45 °, namely when is halfway between 0 ° and 90 °.

differently from what would be predicted by the model. These two points may reveal something very important about the underlying processes that are operating under these circumstances. One such possibility is that retinal and extra-retinal signals are combined through a *non-additive* interaction. This means for instance that a pitch rotation (i.e. a rotation of the head about the horizontal axis that induce a vertical translation of the OF) may be weighted differently when the optical signal is congruent (for instance, a vertical gradient of the velocity field). At this moment the estimates for $w_{\omega x}$ (0.43 ± 0.14) and $w_{\omega y}$ (0.46 ± 0.08) are very similar between each other.

In order to test this intuition, we run a separate model in which the active data set was divided into horizontal and vertical gradients. Then the same model fitting procedure described above was followed. The result of this analysis showed the presence of a non-additive interaction between the gradients and the weights assigned to the extra-retinal signals. When only the vertical gradient was fitted, then the estimated parameters were: $w_{\omega x} = 2.76 \pm 0.46$ and $w_{\omega y} = 0.32 \pm 0.17$. In contrast, when only the horizontal gradient was fitted, the estimated $w_{\omega x}$ was

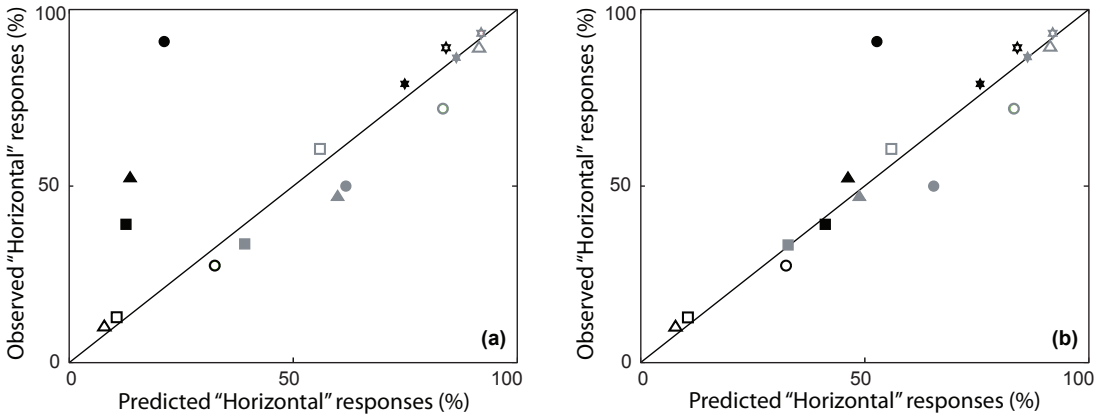

Figure 2.10: **Model prediction and Observed data. Panel a** Plot of predicted against measured proportions of "horizontal" responses. **Panel a** outlines the presence of two outliers due to the interaction between retinal and extra-retinal signals in the two OF gradients conditions. The improvement of the predictive power of the model is evident in **Panel b** when the fit is run taking into account the interaction. Different symbols are associated with different condition (see the text, figure 2.8 and figure 2.7)

about zero ($3.56e-004 \pm 0.13$ while $w_{\omega y}$ took the value of 0.46 ± 0.09. This differential weighting of the extra-retinal inputs is diagnostic of an interaction between retinal inputs and extra-retinal signals. In **Panel b** of figure 2.10 are shown the predictions of the regression model when horizontal and vertical gradients are treated separately.

2.4 Discussion

In previous experiments, the veridical $3D$ perceptual interpretation of ambiguous OFs was found to be improved when the observers were actively generating the OF as opposed to when the observers were passively experiencing the retinal flow (Wexler et al. [2001b], Wexler et al. [2001a], van Boxtel et al. [2003], Wexler [2003]). This result was taken as the proof that the *veridical* shape recovery is possible thanks to an inverse geometry analysis (Ullman [1979], Cornilleau-Peres et al. [2002], Wexler [2003], Wexler [2005], Colas et al. [2007]): (a) the OF is analyzed in conjunction with the extra-retinal signals and (b) some assumptions

about the motion of the objects are introduced in the perceptual interpretation process. Here we proposed and tested an alternative explanation based on two assumptions: (i) the relevant input to the visual system is the *head-centric* OF and not the *retinotopic* OF; and (ii) linear ego-motion signals are mostly ignored in the perceptual analysis of the $3D$ structure from the OF (Domini and Caudek [2003a], Fantoni et al. [2010], Caudek et al. [2011], Fantoni et al. [2012]). Both the Active and Passive Block presented here were specifically devised to test this hypothesis by carefully dissociating the information content of the *head-centric* vs. *retinotopic* OF and the presence/absence of the extra-retinal signals. The contribution of the present study is twofold. The results indicate that the presence/absence of extra-retinal signals is negligible while translational components of the *head-centric* OF systematically bias the perceived plane orientation. The perceptual strategy employed for the extraction of the tilt direction is based on the computation of the self-motion (θ) from the head-centric OF and the sensed angular rotations, but with no contribution of the linear ego-motion signals. This strategy doesn't necessarily guarantee the veridical extraction of $3D$ affine properties.

2.4.1 Is the Advantage of Active over Passive vision attributable to extra-retinal signals?

In the Active Block the Translational components of the *head-centric* OF were made (i) either available (like in previous studies) or (ii) not available to the active observers while the presence of extra-retinal signals was kept constant. The responses of our participants followed previous results when translational components of the OF were present. However, the responses became markedly different when the translational components were *not* available: the participants performed at chance level. This result constitutes a novel finding. It clearly indicates that even when are available, the extra-retinal signals are not employed for the $3D$ interpretation of the affine relationships such as the tilt of the surface. The result stands in contrast to previous findings where the perception of tilt direction from ambiguous OF benefited from the availability of the extra-retinal signals (Dijkstra et al. [1995], Cornilleau-Peres et al. [2002], van Boxtel et al.

[2003]). Our finding, instead, is in line with a of studies that found that extra-retinal signals are seldom employed for the perceptual interpretation of Euclidean $3D$ properties of objects (such as the slant, see for instance Domini and Caudek [2003a], Fantoni et al. [2010], Caudek et al. [2011], Fantoni et al. [2012]).

We propose that this conflicting findings may be explained by the translational components of the OF. This means to embrace a head-centric frame of reference for the analysis of the OF. Indeed, in previous studies on tilt perception the presence of extra-retinal signals covaried with the presence of the translational components as a retinotopic frame of reference was assumed. Therefore the contribution of the two variables was confused. In the Appendix A we demonstrated how a head-centric analysis of the OF is related to the perceived tilt direction. Our approach doesn't include any knowledge about the linear ego-motion nor does it make assumptions about the motion of the objects. Nevertheless, the model captures the vast majority of the data under the different conditions tested (see figure 2.10). The relevance of this analysis and the head-centric frame of reference becomes also apparent, in the experimental results, when translational components were artificially included and enhanced. The perceived $3D$ surface orientation was strongly modulated by the presence of the translational components.

2.4.2 The sensitivity of Passive vision

The passive Block was a re-play of the head-centric OF experienced by the participants during the active trials and was performed in order to further disentangle the relative contribution of retinal and extra-retinal information (indeed in this case no extra-retinal information were available).

Consistently with previous version of the present experiment (Wexler et al. [2001a]), the ambiguities in the perception of the tilt were present when the observers viewed pure gradients of the velocity field (i.e. *No-Translation*). This result was taken to advocate the view according which, in order to disambiguate the OF, the extra-retinal signals are needed. However, this possibility is now ruled out when the *No-Translation* condition is compared with the experimental condition *All-Translation*. The *All-Translation* condition includes all the OF information

that was available to the obsever during the Active Block. Interestingly, a small but consistent trend (17%) toward the opposite perceptual solution was found when translation components were made available. Again, this evidence speaks in favor of the role of the translational components in the context of a head-centric analysis of the OF. The reader may argue that the ammount of the trend is not sufficient to reach a veridical performance (like in the Active Block). Therefore, this leaves open the possibility that, at least to some extent, the visual system does make use of extra-retinal information. In our opinion, this residual effect may be accounted for by the retinal stabilization hypothesis (Cornilleau-Peres and Droulez [1994]). During active movements, ego-motion signals may be used to stabilize and conduct a better measurement of the retinal inputs; this process however, doesn't necessarily improve the processing of depth from motion (Domini and Caudek [2003a],Domini and Caudek [2003b]).

Finally, similarly to the active Block, the effect of the Translational components was even more pronounced when the amount of translation was artificially added to the OF. The perceive tilt was clearly biased by the presence of the translations.

2.4.3 Head-Centric OF in a broader context

Overall, the empirical results of the Active and Passive Block provide support to the notion of a head-centric analysis of the OF. Indeed, a representation of the OF centered on the head or on the body might be more relevant when the observer is navigating in a complex environment. For instance, a head-centric flow signals the motion of the head in space (Beintema and van den Berg [1998]) and computing the head-centric translations is an important step for the perception of heading direction (Royden et al. [1992], van den Berg [1992]). Our study suggests that similar processes may operate also during the shape perception of small objects defined by the OF.

Goossens et al. (2006) identified the regions, in the human visual hierarchy, responsible for the representation of the head-centric OF. This study suggests that neurostimulation techniques (such as the Transcranial Magnetic Stimulation) could be employed to interfere with the normal OF processing and provide

causative and converging evidence about the functional role of the head-centric OF in visual tasks.

Chapter 3

Non-informative components of retinal and extra-retinal signals affect perceived surface orientation from optic flow

Abstract

We test two predictions of a new computational model for the interpretation of the Optic Flow (OF) (Domini et al. [2012]): (1) perceived orientation of a planar surface rotating about an axis coplanar to the surface undergoes a 90° flip whenever a translational component orthogonal to the axis of rotation is added to the OF; (2) the perceptual interpretation of the OF relies on the angular, but not on the linear component of observers ego-motion. In the previous chapter, a static observer viewed the OF produced by a random-dot planar surface rotating about a vertical axis (e.g., a rigid flag hinging on a vertical pole). This OF induces a veridical perception of the surface orientation. However, consistently with prediction (1), when a vertical translational component was added to this OF, the $3D$ interpretation underwent a 90° flip (i.e., a rigid flag hinging on a horizontal pole). In the present experiment, the OFs were actively produced by observers head movements. The observer looked at rotating planar surfaces while performing either a lateral translation of the head or a horizontal head rotation (yaw). In one condition the motion of the planar surface was tethered to the motion of the observer, so that the translational component of the OF was nil. In another condition a surface rotating about a static axis of rotation produced the same lateral translation of the OF tested in the previous chapter. Consistently with prediction (2), perceived surface orientation depended on the lateral motion of the OF in the head-translation, but not in the in the head-rotation condition. Experimental results support the model proposed by Domini, et al., (2012).

Mancuso, G., Fantoni, C., Caudek, C., and Domini, F. (2012) Non-informative components of retinal and extra-retinal signals affect per-

ceived surface orientation from optic flow. doi: 10.1167/12.9.241 *Journal of Vision* August 13, 2012 vol. 12 no. 9 article 241

3.1 Introduction

The Optic Flow (OF) is an important source of information about the three-dimensional ($3D$) shape of objects and the layout of the environment (Gibson [1950], Koenderink and van Doorn [1975], Koenderink and van Doorn [1978], Ullman [1979], Longuet-Higgins and Prazdny [1980], Longuet-Higgins [1984]). Our previous chapter showed that, both in passive and active vision, perceived surface tilt[1] is modulated by the translations of the head-centric OF. Also, in active vision we found evidence that linear ego-motion signals are mainly, although not completely, ignored in the perceptual interpretation process. Under normal circumstances, however, when the observers are freely moving in the environment, the linear ego-motion signals are not the only extra-retinal signals available: indeed angular movements are also largely represented. Our modeling work (Appendix A) would suggest that angular ego-motion signals are employed for the perceptual analysis of the OF. Although there seems to be preliminary evidence in the domain of motion detection (Jaekl et al. [2004]), this assumption was never tested empirically with structure-from-motion stimuli. Therefore, it was important to understand if the OF translation induced by the angular or linear motion of the head differently affects the perceptual interpretation of surface tilt.

In order to achieve this goal, we measured the perceived tilt in different viewing conditions. In one case, the subjects were laterally moving and the translations of the head-centric OF were either absent or present (depending on whether the object was or was not kept aligned on the line of sight). In another condition, the same presence/absence of the head-centric OF translations was manipulated while the participants were rotating their head along the horizontal plane (i.e. yaw rotation). In line with our model, we found evidence of a differential effect of linear vs. angular ego-motion signals:

- for a linearly-moving observer, the perceived surface tilt is (i) biased by the translation of the head-centric OF and (ii) accurate when the surface is kept aligned on the line of sight (so as to eliminate the translations);

[1] The tilt of a slanted surface is the projection of its normal on the fronto-parallel plane.

- interestingly, the opposite pattern is observed when the observers are rotating the head. The perceived surface tilt is (i) biased when there are *no* translations of the head-centric OF and (ii) accurate when the translations are present.

These empirical findings provide support to our head-centric model. The linear ego-motion signals are not taken into account during the computation of the OF. Therefore, when available, the observer makes use of the translational components to derive the surface tilt although this process doesn't necessarily guarantee a veridical performance. On the other hand, observers compensate for the OF translations induced by the rotation of the head. This compensation suggests that the system precisely estimates the angular motion of the head that is directly employed for the perception of the surface orientation. This notion is further confirmed by the fact that the performance is instead biased by objects that are moving in a earth-centered reference frame.

3.2 Methods

3.2.1 Participants

Eleven participants (age 20 to 36, 8 males) took part in the experiment. All had normal or corrected-to-normal vision and signed a consent form as directed by the local Ethical Committee.

3.2.2 Apparatus

The apparatus employed in the present experiment has been described in the previous chapter; also a detailed account of the experimental setup could be found in Fantoni et al. [2010],Caudek et al. [2011] and Fantoni et al. [2012].

3.2.3 Visual stimulation and the task

For a detailed description of the visual stimuli, the interested reader is again referred to the previous chapter. For the sake of our discussion here it is sufficient

to remind that the visual stimuli were small (approximately $5\,^\circ \times 5\,^\circ$) random-dot planar patches. On different trials, the surface was slanted of about $45\,^\circ$ and it was rotating about the y or the x-axis. An Optotrak Certus System (Northern Digital Inc.) was employed to update the arrangement of the dots according to the observers head position and his/her orientation with respect to the simulated surface.

According to the OF condition (see below section 3.2.5) in half of the presentations the stimulus displays were either (a) fixed on the center of the screen (the center of the stimulus corresponded to the center the observers cyclopean eye at rest) or (b) jointly moving with the observer motion (the center of the stimulus was constantly updated and centered on the actual observers cyclopean eye). Then, for each stimulus frame the dots of the simulated planar surface were projected onto the screen by using a generalized perspective pinhole model. In the case described in (a) the Center Of Projection (COP) was the observers cyclopean eye at rest whereas in the second condition (b) the actual observers right eye position was employed as COP. This manipulation allowed us to minimize the shear components of the OF, despite the observer's motion.

3.2.4 Procedure

The procedure included three phases: (I) instructions; (II) a practice in which participants were trained to maintain a constant velocity ($\approx 10 deg/sec$[1]) during head translation or rotation and (III) the experimental task in which participants were asked to monocularly judge the tilt direction of the planar surface by pressing one of four buttons on the keyboard (corresponding to the following tilt directions: $0\,^\circ$, $90\,^\circ$, $180\,^\circ$ and $270\,^\circ$). This task was performed twice in two separate blocks: one block for the Head Translation condition and one for the Head Rotation condition. The order of the blocks was randomly intermixed between participants.

As described with greater details below, the OF experienced by the observers were horizontal or vertical gradients of expanding flow. The OF was generated by the rotation of the surfaces and the concurrent motion of the observer. The

[1] empirical translational and rotational velocities were respectively: 9.9 deg/sec 95% $CI[8.8, 10.8]$ and 9.8 deg/sec 95% $CI[8.2, 11.2]$

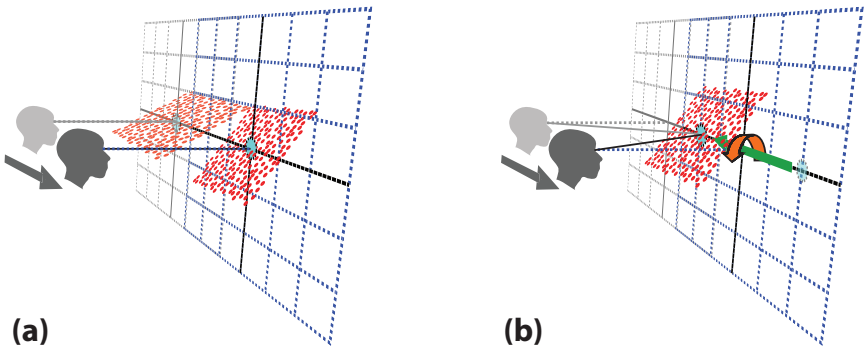

Figure 3.1: **OF Translation conditions. Panel a**. The surface is rotating about its horizontal axis and is also aligned to the line of sight of the translating observer. In this condition, the visual direction is always aligned with the center of the stimulus where the OF vanishes, therefore there is no relative translation between the object and the observer (OF Translation Absent). **Panel b**. The surface is rotating about a stationary axis of rotation, therefore a translational component of the OF is induced by the relative motion of the observer and the object (OF Translation Present).

experiment was carefully designed in order to compare two OF conditions. In one case the stimulus was rotating about an earth-stationary axis of rotation; in the second case the surface was tethered to the head of the observer while it was also rotating about its own axis of rotation. This manipulation allowed us to study the perceptual interpretation of the head-centric OF in the presence or absence of translatory components of the OF. Indeed, the relative motion of a moving observer and a stimulus that is rotating about an earth-fixed axis of rotation generates a translational component in the head-centric OF (condition: OF Translation Present), whereas when the stimulus is always aligned on the line of sight of the observer, no translational components are present in the OF (condition: OF Translation Absent). Figure 3.1 provides and intuition of this distinction.

At the beginning of each trial, after 5 minutes of dark-adaptation, a red fixation mark, corresponding to the projection of the cyclopean eye on the screen, was shown on the monitor and the observer was required to move his/her head rightward. Depending on the block (i.e. Head Translation or Head Rotation) the participants were either translating laterally (side-to-side) or rotating the head on

the horizontal plane (yaw-rotation). The observer was instructed to reverse the direction of head translation after hearing a beep signaling a head shift of 50 mm (to the right) relative to the center of the screen and after hearing a beep at -50 mm (to the left) signaling a shift in the opposite direction. In the Head Rotation block the subject was instructed to rotate the head of about $5\,^{\circ}$. This value was chosen because, at a distance of 568.5 mm, the projection of the cyclopean eye on the monitor covers exactly the same amount of space (50 mm) spanned in the Head Translation block. After two cycles of head movement, when the observer's head was in the leftmost position and was moving rightward, the fixation mark was replaced by the stimulus display. The simulated surface remained visible for an entire cycle and then was replaced by a black screen. After stopping the motion of the head, subjects were asked to classify, with a button press, the direction of the tilt. The amount of surface slant was related with the motion of the head of the observer and spanned approximately 10 deg (9.6 deg 95% CI[7.0, 12.4]) ranging from about $50\,^{\circ}$ to $40\,^{\circ}$ during each trial. In the Head Translation Block the rotation of the surface was related with the lateral displacement of the head thanks to the equation: $\omega = \arctan\left[\frac{(e_x)}{|f_z+e_z|}\right] + 45(\frac{\pi}{180})$. ω is the instantaneous slant of the surface, e_x is the x-position of the right eye, f_z is the viewing distance and e_z the instantaneous right eye position along the z-axis. In the Head Rotation block the equation employed was $\omega = \arctan\left[\frac{P_x - e_x}{P_z}\right] + 45(\frac{\pi}{180})$, where P_x and P_z are respectively the x and z-positions of the projection of the cyclopean eye at the viewing distance whereas e_x is the x-position the right eye on the x-axis.

3.2.5 Design, Data Analysis and Predictions

The combination of the factors employed resulted in a $2 \times 2 \times 2$ within-subjects design: Head Motion Block (Head Translation vs. Head Rotation), OF Translation (Present vs. Absent) and Gradient of the velocity field (Horizontal and Vertical) were the three independent variables. In both Head Motion Blocks each subject viewed 20 presentations, in random order, of the four different stimuli.

As explained in the Appendix A, our model relates the observer's translation on the frontal plane (angle θ) and the tilt direction (τ) through the analysis of the head-centric OF and the angular (not linear) ego-motion signals:

$$\theta \quad = \arctan\left[\frac{|a_4+\omega_{R_x}|+a_6}{|a_1-\omega_{R_y}|+a_2}\right] = -\tau \;\; or \;\; (\pi-\tau) \tag{3.1}$$

The experiment presented in this chapter was carefully designed to test our model by selectively manipulating each term of the equation 3.1. In the context of the present experiment, we limited our analysis to the angular rotations on the horizontal plane ω_{R_y} and lateral linear translations. This means that the terms a_4, and ω_{R_x} were always null and therefore we can safely ignore them in the following analysis (thus, the variables manipulated were: a_1, ω_{R_y} and the two gradients a_2 and a_6).

In two separate blocks, in order to manipulate the presence or the absence of the angular rotation (ω_{R_y}), the participants looked at a rotating planar surface while performing: (i) linear lateral head movements ($\omega_{R_y} = 0$) or (ii) angular head movements ($\omega_{R_y} \neq 0$). Similarly, in order to manipulate the presence or the absence of OF Translations, in each block the surface could have been: (i) rotating about a stationary axis of rotation or ($a_1 \neq 0$, OF Translation Present) (ii) tethered to the motion of the observer ($a_1 = 0$, OF Translation Absent). The two gradients of the velocity field (a_2 and a_6) were equally balanced in all the combinations tested. These manipulations allowed us to make specific predictions based upon our model. In the following we will focus only on the Vertical Gradient because this condition is more informative about the interaction of the head-centric OF and the ego-motion. The combination of all these variables gives rise to two different scenarios.

First, for the case of a translating observer and a stationary object, the relative motion generates a horizontal translational component of the head-centric OF. In contrast, when the object is tethered to the line of sight of the observer, the sagittal plane of the head is always centered on the center of the image and there are no translational components in the OF. According to previous approaches in this field, the two conditions shouldn't be different. The observer keeps a steady fixation on the target (therefore there are no translational components in the OF); also the visual system incorporates the knowledge of body motion into the analysis of the retinal input. Our model, instead, makes the counter intuitive

		Head-Motion Block	
		Linear Translation	Angular Rotation
OF Translations	Absent	$\theta = \arctan\left[\frac{a_6}{0}\right] \approx 90\,^\circ$	$\theta = \arctan\left[\frac{a_6}{\omega_{R_y}}\right] \approx 0\,^\circ$
	Present	$\theta = \arctan\left[\frac{a_6}{a_1}\right] \approx 0\,^\circ$	$\theta = \arctan\left[\frac{a_6}{0}\right] \approx 90\,^\circ$

Table 3.1: **The Experimental Design:** the angle θ approaches zeros only if the denumerator is much bigger compared to the numerator. In our experiment, a more realistic value of the ratio was about 1, which means that the angle θ was about 45 °

predictions that: (i) the perceived tilt is biased by the ego-motion when the stimulus is rotating about a *stationary* axis of rotation while (ii) the perceived tilt is not affected when the image is fixed in a head-centric coordinate system (but moving in the earth-centered reference frame). Formally, this can be seen by looking at the direction of the angle θ which is computed as $\arctan\left[\frac{a_6}{a_1}\right]$ in (i)[1] while is computed as $\arctan\left[\frac{a_6}{0}\right]$ in (ii).

The Head-Rotation Block leads to the complete opposite configuration: when the subject is rotating the head and the stimulus is tethered, although there is no relative translation ($a_1 = 0$), the tilt perception should be biased toward a horizontal solution. This can be easily seen by looking at how the angle θ is computed $\arctan\left[\frac{a_6}{\omega_{R_y}}\right]$. Vice versa, when the stimulus is stationary in a earth-centered reference frame, the translational component (a_1) is equal and opposite to the head rotation ω_{R_y}, so these two quantities are canceling each other out and the result of this process leaves the perception of the surface tilt not altered ($\arctan\left[\frac{a_6}{0}\right]$). Table 3.1 summarize how the relevant manipulations of the present experiment systematically affected the computation of the angle θ and therefore the perceived tilt.

3.3 Results

Figure 3.2 **Panel a** shows the results of the Head-Rotation and Head-Translation Block for the case of a Vertical Gradient of the OF. It is immediately apparent

[1] if a_1 is sufficiently strong relative to a_6, that is $a_1 \gg a_6$, a horizontal perceived solution becomes more and more likely.

the interaction between the OF Translation and the type of Head-Motion. The presence of the OF Translation biased the perceived tilt when the participants where moving linearly from side to side (from 0.32 ± 0.03 to 0.61 ± 0.03). Whereas, the opposite scenario was observed during angular rotations of the head; the biased performance happened when no translational signals where available (from 0.64 ± 0.03 to 0.24 ± 0.03). This inspection reveals a first very important point: the extra-retinal signals arising from linear and angular motion are differentially employed for the perceptual analysis of the surface orientation. In agreement with our hypothesis, the linear translation of the head, is *not* compensated during the perceptual judgments: a Vertical Gradient is perceived mostly as a horizontally oriented surface when a horizontal translation is added (*OF Translation Present*). No such bias was found when the horizontal translation was induced by the angular rotation of the head. This means that, in this case, the angular motion of the head is taken into account during the computation of the OF.

In **Panel b** it's shown the performance for the Horizontal Gradients of the OF. In this case we can see very little modulation of the OF Translation factor. The proportions of "horizontal" responses were higher for the Head-Rotation (0.97 ± 0.01 and 0.96 ± 0.01, NO-Translations and Horizontal-Translation respectively) than for the Head-Translation Block (0.87 ± 0.02 and 0.85 ± 0.02, NO-Translations and Horizontal-Translation respectively). This result is consistent with our idea that angular rotations of the head are measured and encoded more precisely. Finally, the Horizontal Translation of the OF didn't affect the perceived orientation. This may have happened for two reasons. First, in this case the Horizontal Translation was in the same direction of the Gradient of the velocity field, therefore we would have expected at the most a strengthening of the same horizontal response. Also, the fact that participants performed at the ceiling level make the detection of possible differences more difficult.

Having described the results we can now focus on the regression model that we built for the statistical analysis. Similarly to the previous chapter, the relationship between the perceived tilt and the angle θ is expressed by the following probit[1]

[1]Following the results of the previous chapter, as well as theoretical considerations, the center of the probit was set at $45\,^\circ$ which is halfway between $0\,^\circ$ and $90\,^\circ$.

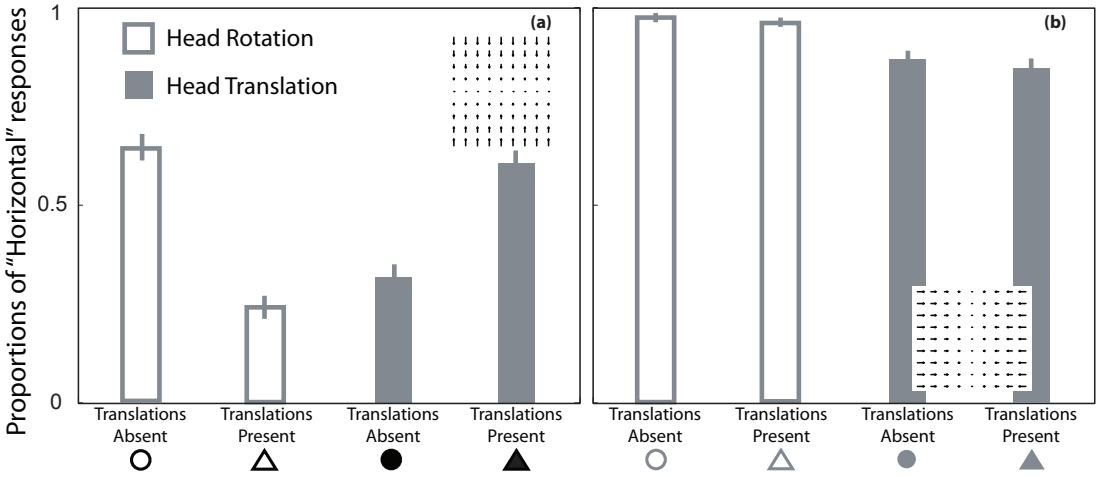

Figure 3.2: **The Results of Head Rotation and Translation.** The Perceived tilt plotted as a function of the different OF *Translation* levels. The empty and filled bars refer respectively to the Head-Rotation and Head-Translation Blocks. In **Panel a** the stimulus is a Vertical Gradient of the velocity field. The Head-Rotation and Translation Block showed a non-additive interaction: the horizontal OF translation was able to bias the judgments only in the Head Translation but not in the Head-Rotation Block. Vice versa, the absence of translation of the OF biased the Head-Rotation but not the Head-Translation Block. In **Panel b** (horizontal gradient) the performance of the participants reached the ceiling level, showing very little modulation of the experimental conditions. The symbols identify the levels of the variable OF Translation for further analysis (see figure 3.3).

function:

$$y \quad = \phi_\sigma(\theta_{w_{1\ldots n}}(a_1, a_2, a_6, \omega_{Ry})) \tag{3.2}$$

The angle θ is a function of several optical (a_1, a_2, a_6) and non-optical (ω_{Ry}) variables and our goal was to estimate, by means of a non-linear regression, the model parameters $(\sigma,\ w_{1\ldots n})$.

The significance level and the model comparisons were achieved using the Transformed Likelihood Ratio statistics (Kingdom and Prins [2010]) and the Akaike's Information Criterion (Akaike [1974], Sugiura [1978], Hurvich and Tsai [1989]). The optimal model contained four free parameters: the standard deviation of the probit function (σ), two parameters associated to the retinal in-

puts (w_{a_1} and w_{a_6}) and one parameter associated to yaw-rotation of the head ($w_{\omega y}$). This model perfomed significantly better when compared with a null model ($\chi^2_{(3)} = 23.76, p < 0.001$) and its predictive power can be inspected in figure 3.3. In the **Panel a** of figure 3.3 it's shown the scatterplot of the Predicted vs. Observed proportions of "horizontal" responses. The simbols are associated with the different levels of the factors (see also figure 3.2). Filled and empty simbols indicate respectively the Head-Translation and Head-Rotation Block. Whereas, black and grey color are associated respectively with the Vertical or Horizontal Gradients of the velocity field. Remarkably, with only four paramters, the model prediction fall very close to the empirical results in all the eight condition tested. Figure 3.3, **Panel b**, represents the probit function that relates the perceived orientation as a function of the angle θ. Again, the fit well predicts the empirical results of all the conditions tested.

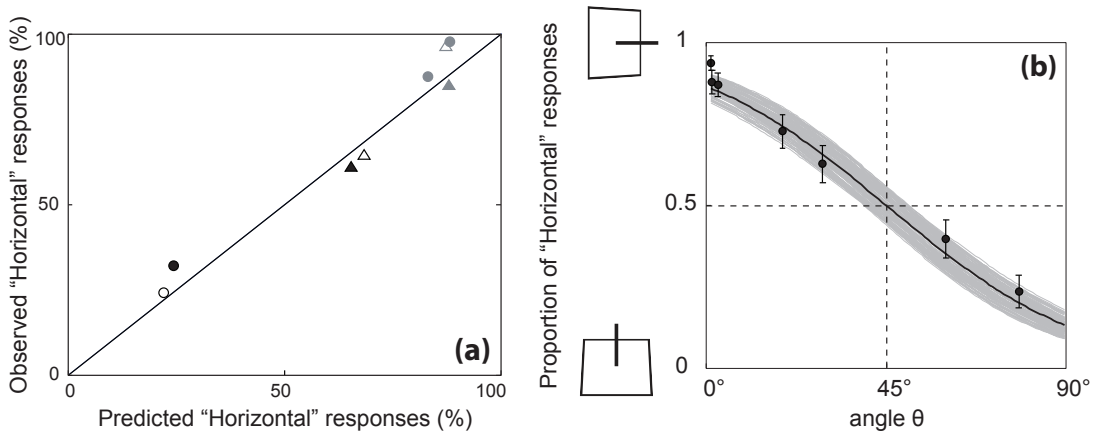

Figure 3.3: **The Model's prediction. Panel a** shows the Predicted VS. Observed proportions of "horizontal" responses. The simbols are coding the different levels of the variables tested: filled and empty simbols indicate respectively the Head-Translation and Head-Rotation Block. Whereas, black and grey color are associated respectively with the Vertical or Horizontal Gradients of the velocity field. **Panel b** The proportions of "horizontal" responses decrease as a function of the angle θ (light lines indicate uncertainty in the regression). The empirical proportions where calculated dividing the angle θ into eight equally dense bins.

3.4 Discussion

In the previous chapter, the perceptual interpretation of ambiguous OF was found to be modulated by the presence of translational components of the head-centric OF. In passive vision the presence of normal translational components improved the performance while, artificially enhanced translations, strongly biased the perceptual interpretation. In contrast, in active vision the effect of normal translational components was overall asymmetric: only in one case (out of two) tested, the presence of translational components was associated with an improvement of the performance. Given the importance of the translations, we must stress that under normal circumstances, the relative translation of the OF may be generated both by linear displacements or rotations of the head (or a combination of the two). Therefore, it was important to understand how the retinal signals were combined with the linear and angular extra-retinal signals for the perception of surface tilt. In the experiment presented here, the variables were carefully manipulated in order to dissociate the information conveyed by the stimuli and the type of head motion (linear vs. angular). The contribution of the study is manifold. The empirical findings provide further support to the notion of a head-centric analysis of the OF. Moreover, the results clearly showed that linear ego-motion signals are *not* combined with the OF for a veridical estimation of the object structure. In contrast, our study demonstrates that the angular ego-motion signals are instead encoded and employed for the perceptual analysis of the OF. These observations are in line with our model that derives the tilt orientation by computing the direction of self motion (θ) thanks to a joint analysis of the head-centric OF and the angular (but not linear) ego-motion signals.

3.4.1 What is the ego-motion signal employed by the visual system for the recovery of surface tilt orientation?

In the experiment presented here, the presence/absence of OF Translations was actively generated by the linear or the angular motion of the observers. This manipulation allowed us to dissociate the relative contribution of two extra-retinal

signals to the perception of surface tilt.

For the case of a Vertical gradient of the velocity field, the responses of our participants markedly differed when the observers were linearly moving vs. when they were rotating their heads. Surface tilt was biased by an OF Translation induced by a linearly-moving observer but not when the *same* OF Translation was induced by the rotation of the head. Vice versa, perceived tilt was accurate when the OF was kept aligned with the line of sight of the linearly-moving observer but was biased for the case of a rotating observer. These results have important implications. The study confirms that the linear ego-motion signals are *not* optimally combined with the retinal information to achieve a veridical perception of the surface orientation. This idea finds support on previous studies on slant perception (Domini and Caudek [2003a], Fantoni et al. [2010], Caudek et al. [2011], Fantoni et al. [2012]). Our data support the view that the visual system makes use of the translations of the head-centric OF to derive the surface tilt, despite the fact that this process doesn't necessarily guarantee the veridical solution (Domini et al. [2012]).

Another important implication of our results is that rotations of the head are precisely estimated during ego-motion and subsequently employed for the analysis of the OF. This notion is based on the observation that OF translations are compensated during the rotation of the head.

3.4.2 The role of linear and angular ego-motion signal

The study of how visual and vestibular information are integrated has a long tradition in human psychophysics (for instance, Benson [1982], Kolev et al. [1996]). Differences between angular and linear motion have already been tested in the presence (Jaekl et al. [2004]) or absence (MacNeilage et al. [2010]) of visual stimulation. These preliminary ideas seem consistent with our results that demonstrated that yaw rotation is employed for the analysis of the OF whereas the linear motion of the head is poorly integrated. In the next chapter, we systematically study the contribution of these two extra-retinal signals to the detection of visual motion.

Chapter 4

The detection of visual motion induced by linear and angular head movements

Abstract

When we move about in a cluttered environment we perceive a stable world although every movement of the head is associated with the visual displacement of images on our retinae. Here, we study how and to what extent the brain carries this compensation by examining the interaction between retinal and extra-retinal signals. In Experiment 1 we compared the amount of visual shift that can be tolerated when the observers are: (I) laterally translating or (II) laterally rotating the head. In Experiment 2, static participants were presented with a replay of the visual motion that they themselves generated during the Experiment 1. By keeping constant the visual stimulation in the two experiments, we were able to directly compare the performance of the observers when they were translating or rotating the head. Angular movements are associated with less biased and more precise estimates of object motion, whereas linear translation are biased and are associated with a more noisy representation of object motion. These findings are consistent with the assumption of our model and with the results obtained in the previous experiment (*Chapter 3*).

4.1 Introduction

When we move about in the environment, we usually perceive a stable world around us although every movement of the head is associated with the visual displacement of images on our retinae. This phenomenon is often referred to as perceptual stability, and it has been a longstanding and unsolved problem in the field of vision sciences (Wallach [1987]). The ability to correctly judge the motion of surrounding objects becomes crucial in order to effectively navigate in a complex environment. However, achieving perceptual stability it's a demanding problem in sensory integration as it requires the neural combination of visual (Lappe et al. [1999], Redlick et al. [2001], Vaina et al. [2004]), vestibular (Benson [1982]) and somatosensory/proprioceptive signals (Kandel et al. [2000]) about head motion. In fact, at every moment in time, the visual input contains two major sources of motion: a source resulting from self-motion, and a source resulting from the movement of objects relative to the observer. Therefore, the brain faces a source-separation problem as it has to parse the visual input into components due to self and objects motion (DeAngelis and Angelaki [2012]). How, and to what extent, the source-separation problem can be solved by the brain has been a matter of debate and ground for theoretical work. In our previous experiment for instance, we found that linear ego-motion signals are not compensated if the subject is asked to judge the tilt orientation of an optic-flow-defined surface. The result of this process is a biased representation of the object structure and motion (see also Fantoni et al. [2010], Caudek et al. [2011], Fantoni et al. [2012]). In contrast, we found evidence that the angular rotation of the head it's accurately estimated during the computation of the OF.

Several psychophysical studies have shown that estimating the object motion during active self motion can be achieved comparing the local motion pattern to the global flow (Rushton and Warren [2005], Warren and Rushton [2009], Royden and Connors [2010], Calabro et al. [2011]). However, all these studies have focused on the perceptual analysis of the interaction between object motion and background motion while it would be important to measure the absolute perceptual sensitivity to object motion without background information. One attempt to answer this question has been advanced by Kolev et al 1996: the Authors

investigated object motion detection in the presence or absence of vestibular signals. In this experiment however, the participants were passively rotated while sitting on a *Bárány* chair and therefore the vestibular system was passively stimulated and not actively prompted. One of the most comprehensive measurements of perceptual stability, during active linear and angular head movements, was conducted by Jaekl et al 2004. One important difference from previous studies was that instead of object motion, the participants were judging the motion of a full-field-view virtual-reality environment which was linked to their own head motion. The subjects were asked to adjust a mismatch between the visual stimulation and the head motion so that the virtual environment appeared stable as they moved. When the subjects were rotating the head (yaw) no bias was found; bias which was instead present when the subjects were translating side-to-side. The Authors explained this result hypothesizing that we are more accustomed to head rotation than side-to-side translations. Alternatively, the visual system could have different sensitivities to translations and rotations on the horizontal plane. We believe that this difference must be searched in the early stages of signal processing, since at the central level of the parieto-insular vestibular cortex, responses to linear and angular motion were found to be equally represented (Chen et al. [2010]). The possibility that linear and angular signals may exert different effects on the perceptual interpretation of visual motion seems likely as these two signals are detected by different receptors in the peripheral vestibular system (otolith organs and semicircular canals for the detection of linear and angular acceleration respectively, Kandel et al. [2000]). Moreover, careful detection experiments have clearly established that humans are able to sense rotation on the horizontal plane in the absence of visual information; also, the sensitivity to rotation is higher relative to (forward) translations (MacNeilage et al. [2010]).

Given these findings we hypothesized a different sensitivity for angular and linear head motion on the horizontal plane. Contrary to previous experiments (Jaekl et al. [2004]), we employed a small stimulus and not a full-view virtual environment. This was done to avoid the stimulation of large portions of the visual field that might have different sensitivities to visual motion (Raymond [1994]). Thanks to a 2AFC procedure, we studied the interaction between retinal and extra-retinal signals under different head motion conditions. In Experiment 1

we compared the amount of visual shift that can be tolerated in order to perceive an object as stationary when the observers are: (i) laterally translating or (ii) laterally rotating the head. In Experiment 2, static participants were presented with a replay of the visual motion that they themselves generated during the Experiment 1. By keeping constant the visual stimulation in the two experiments, we were able to directly compare the performance of the observers when they were translating or rotating the head.

Briefly, we found that angular movements of the head are efficiently integrated with visual motion. The bias induced by head rotation is significantly smaller compared to the bias induced by linear translation. Also, angular rotations are encoded with a higher precision compared to linear translations. Instead, as previously reported (Fantoni et al. [2010], Caudek et al. [2011], Fantoni et al. [2012]), linear ego motion signals are mostly ignored in the perceptual interpretation of the visual motion: the sensory representation of visual motion is strongly biased and poorly reliable. These findings provide empirical support to the results of the previous experiment (*Chapter 3*) where we found a differential effect of the linear and angular head motion during the $3D$ interpretation of the OF. The present experiment, also, validates the assumptions made by our modeling work.

4.2 Methods

4.2.1 Experiment 1

4.2.1.1 Participants

Eight participants (age range 24 - 36, 3 females) gave informed consent, as directed by the local Ethical Committee, and took part in the first experiment. All had normal or corrected-to-normal vision and were naïve about the purposes of the experiment.

4.2.1.2 Apparatus

A detailed description of the viewing apparatus can be found in Fantoni et al (2010), Caudek et al (2011) and Fantoni et al (2012), here we provide only a

brief report. Stimuli were presented on a 19-inch CRT monitor (ViewSonic 9613, $19W$) which was viewed through a high-quality front-silvered mirror (see *Chapter 2*). This arrangement produced a viewing distance (i.e., distance between the pupil and the center of the display reflected by the mirror) of 568.5 mm relative to the right eye of an observer's at rest. The translational displacements and orientation of the participant's head were recorded real-time by an Optotrak Certus system with two position sensors (with a spatial resolution below 0.01 mm). The two position sensors recovered the $3D$ positions of three markers (infrared emitting diodes) which were attached to the participant's head during the active movements. Thanks to a Dell Precision $T3400$ $525W$ we were able to control the stimulus display and sample the tracker using a standard PCI card. This configuration allowed us to update in real time the position of the visual stimuli on the screen as the participants moved. A custom Visual C++ program supported by OpenGL Libraries and combined with Optotrak API routines was used for stimulus presentation/response recording.

4.2.1.3 Visual stimulation and the task

The visual display was a circle of 10 mm of diameter whose position was updated real time according to the viewing position of the participants. In order to study how the brain compensate for natural head motion, a visual gain (mismatch) was introduced between the horizontal motion of the head in space and the horizontal motion of the visual stimulus on the screen. The motion of the stimulus was linearly related to the motion of the observer with the following equation: $s = gt$. s represents the spatial position of the stimulus on the screen; g, the gain, is a multiplicative term and t represents the spatial position of the observer. Therefore, if $g = 1$ then there is no mismatch between the motion of the observer and the motion of the stimulus. In this case the stimulus is perfectly aligned and centered with the line of sight of the observer (Figure 4.1a). If the gain is bigger or smaller than 1, then the translation of the distal object on the monitor will be respectively bigger or smaller than the translation of the observer (Figure 4.1b). The gain can also take on negative values, in this case the motion of the stimulus will be in the opposite direction of the direction of head motion (Figure 4.1c).

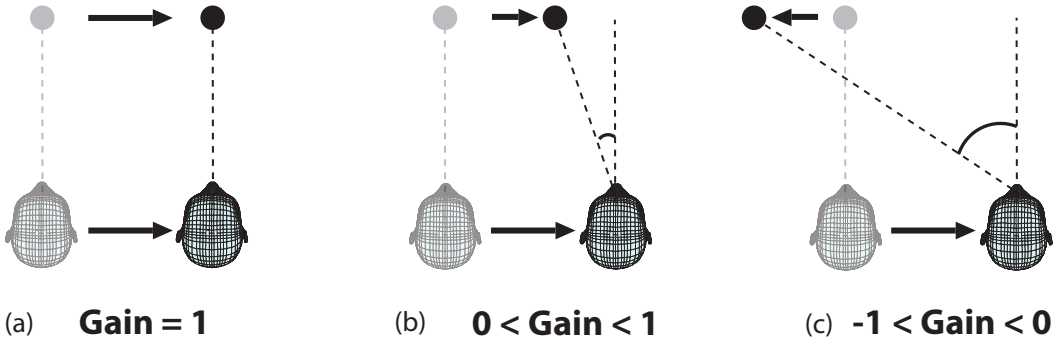

Figure 4.1: **The effect of gain on the visual displacement.** The sketch represents two successive temporal intervals for an observer that is translating laterally. (a) When the gain is 1, there is an exact mapping between the subject and the object motion, the sagittal plane of the head is always aligned with the object. (b) When g is between 0 and 1, the object moves only of a smaller fraction of the total motion of the observer. (c) When g is negative, the motion of the circle is still proportional to the motion of the observer but it is in the opposite direction.

At the end of each trial, the participants reported whether the direction of motion of the visual stimulus was left or right. In general, when the gain is negative, it is easy to detect the direction of motion, because the object on the screen is moving in the opposite direction relative to the observer (Figure 4.1*c*). The task increasingly becomes more difficult when the gain takes on values near zero or positive values. In order to precisely estimate the sensitivity curves for each participant, a four-interleaved-staircase procedure (1 up - 3 down) was employed. Based on previous responses the staircase algorithm adjusts the value of the gain until the visual display appears stable to the observer. If the brain perfectly compensate for the head displacement, then the staircase procedure should converge to a value of the gain that is close to zero; every deviation from this value represents a systematic error and is informative about the underlying neural sensitivity to head translation and rotation.

4.2.1.4 Procedure

The procedure included three phases: (I) instructions; (II) a practice in which participants were trained to maintain a constant velocity ($\approx 15 deg/sec$) during

head translation or rotation and (III) the experimental task in which participants were asked to detect the direction of motion of the circle on the screen. This task was performed twice in two separate blocks: one block for the translation and one for the head rotation. The order of the blocks was randomly intermixed between participants.

At the beginning of each trial, after 5 minutes of dark-adaptation, a red fixation mark, corresponding to the projection of the cyclopean eye on the screen, was shown on the monitor and the observer was required to move his/her head rightward. Depending on the block (i.e. translation or rotation) the participants were either translating laterally or rotating the head.

The observer was instructed to reverse the direction of head motion after hearing a beep signaling a head shift of 50 mm (to the right) relative to the center of the screen and after hearing a beep at -50 (to the left) signaling a shift in the opposite direction. For the head rotation block the subject was instructed to rotate the head of about $5\,^{\circ}$. This value was chosen because, at a distance of 568.5 mm, the projection of the cyclopean eye on the monitor covers exactly the same amount of space (50 mm) spanned in the translation block.

After two cycles of head movement, when the observer's head passed through the center of the screen moving rightward, the fixation mark was replaced by the stimulus display. The moving dot remained visible for half cycle and then was replaced by a black screen. After stopping the motion of the head, subjects were asked to classify, with a button press, the direction of the moving dot as left or right.

4.2.1.5 Design and Data Analysis

In both rotation and translation blocks the participants reported the perceived direction of motion. The probability of a "rightward" response was modeled as a function of the gain value (continuous predictor) by fitting a Cumulative Gaussian distribution (as it has already been done in a similar settings, Jaekl et al. [2004]). Following the Maximum Likelihood procedure described in Wichmann and Hill (2001a, 2001b) four parameters were estimated from the raw data of each participant: (i) the Point of Subjective Equality (PSE); (ii) the Just Noticeable

Difference (JND), defined as the range of stimulus intensity within which the psychometric function increases from 0.5 to 0.8413 (corresponding to 1 standard deviation); (iii) the guess rate (γ) and (iv) the lapse rate (λ). The analysis were conducted using psignifit toolbox version 2.5.6 (Wichmann and Hill [2001a], Wichmann and Hill [2001b]). The two parameters of interest (PSE and JND) were then compared in the two experimental conditions.

4.2.2 Experiment 2

4.2.2.1 Overview

In order to make a direct comparison between the translation and rotation conditions of Experiment 1, we conducted a control experiment to demonstrate that the visual information in the two conditions was not different. Only if this assumption is satisfied any difference found in Experiment 1 could be ascribed to a different sensitivity to translational and rotational movements.

4.2.2.2 Participants

The same pool of eight participants tested in Experiment 1 took part also in Experiment 2. During the first experiment we recorded the kinematics of every movement of the head of our participants and we used this information to recreate in Experiment 2 the visual stimulation experienced during active movements of the Experiment 1.

4.2.2.3 Apparatus and stimuli

Indeed, during Experiment 2 the subjects were not moving and passively observed the visual stimulation that was generated by replaying the 2D transformations generated by the corresponding active trials. This time however, the task was different from the first experiment as it would have been too easy to detect the direction of motion of the circle. Instead we opted for a 2AFC task in which a judgment about the speed of motion was asked. Each trial consisted of two presentations: in one interval the exact same trajectory covered by the subject was replayed (standard stimulus) in the other interval the trajectory was mod-

ulated by the corresponding gain value (comparison stimulus). The task of the participants was to indicate with a button press, in which of the two intervals the stimulus moved faster. The order of presentation of the standard and the comparison was random. Also, in order to avoid that the participants based their decisions on the total length of the trajectory covered on the screen, the starting position of the stimuli followed a random uniform distribution (bounded between $\pm 0.25mm$). The order of the blocks (replay of linear or angular motion) was randomized.

4.2.2.4 Data Analysis

The data gathered in the 2AFC task were fitted using the same procedure applied in Experiment 1. The performance in the passive replay of the translation and rotation blocks was modeled as a Cumulative Gaussian with four free parameters (see section 4.2.1.5) and then the JND was compared in the two conditions.

4.3 Results

4.3.1 Experiment 1

Figure 4.2 displays the amount of translation and rotation along different axis of a typical subject. The panels **a** and **b** show the head kinematics during rotational and translational movements. The experiment was build so that the translation and rotation conditions produced the same angular displacement on the horizontal plane as witnessed by a *t-test* ($Welch - t_{(7)} = -1.4734, p = 0.1841$). Also, the median angular velocity along the horizontal plane was the same in the two conditions ($Welch - t_{(7)} = 0.4758, p = 0.6487$).

Having examined the kinematics of the participants' movements we can now focus on the psychophysical results. In Figure 4.3 are represented the four-interleaved-staircases (1 down - 3 up) for one active observer after having performed the rotation condition. Panel **b** of Figure 4.3 represents the output of the fitting procedure for the same data. Both the translation and the rotation conditions led to a biased estimate of the visual motion in the direction of self-

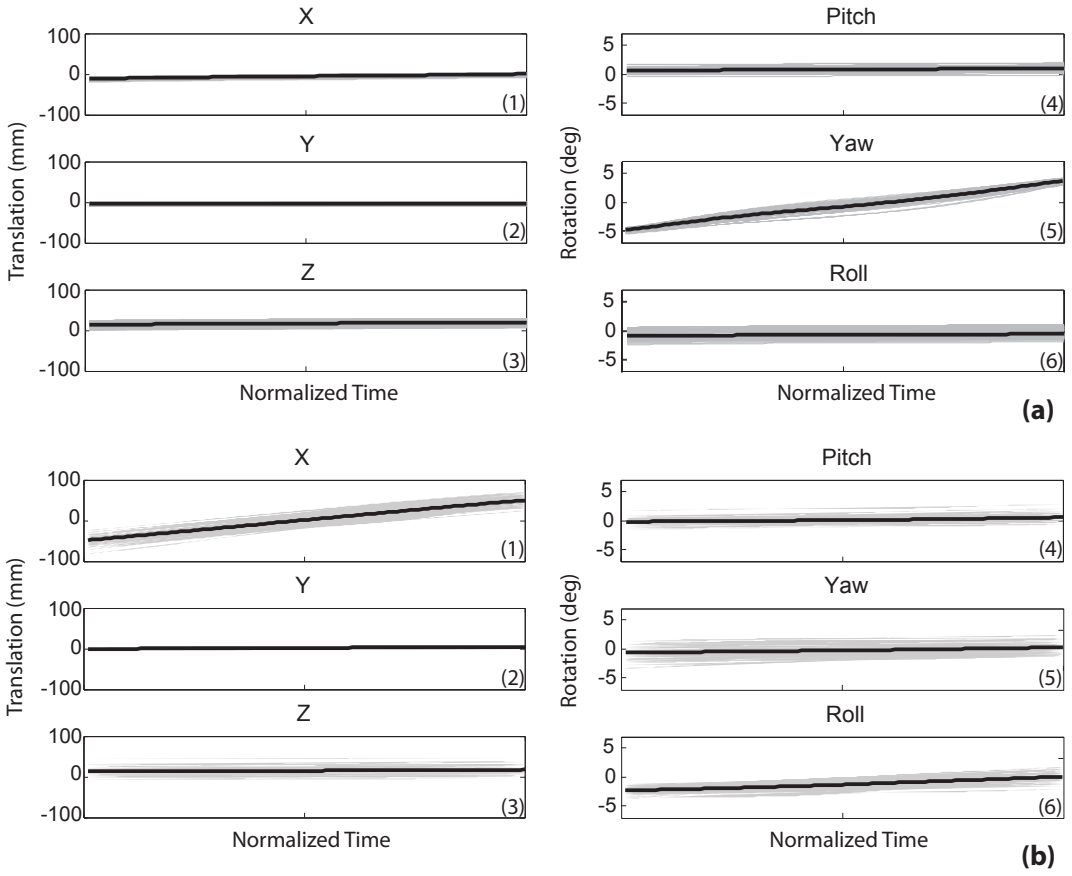

Figure 4.2: **Head kinematics of a typical subject during Rotation (Panel a) and Translation (Panel b).** In the left hand side column we have the translational components along the Horizontal (X), Vertical (Y) and Depth (Z) dimension. On the right column we have the rotation about the Horizontal ($Pitch$), Vertical (Yaw) and Z ($Roll$) axis. **Panel a** refers to the Rotation Block. All the trajectories are oscillating around a central value of 0, with the only exception of the *Yaw* (**a5**) rotation which spanned between $\pm 5\,^{\circ}$. **Panel b** represents the Translation block: again every kinematics is centered at zero except the Translation along the Horizontal axis (**b1**) which covered approximately 100 mm ($\pm 50mm$).

motion[1]: in order to be perceived as stationary, the visual stimulus had to move in the same direction of motion of the observer (rotation: $t_{(7)} = 3.389, p < 0.05$; translation: $t_{(7)} = 5.7324, p < 0.001$). However, as show in Figure 4.4 *a*, the

[1]Since multiple Null-Hypotheses were simultaneously tested, the p-values were corrected accordingly employing the Bonferroni method

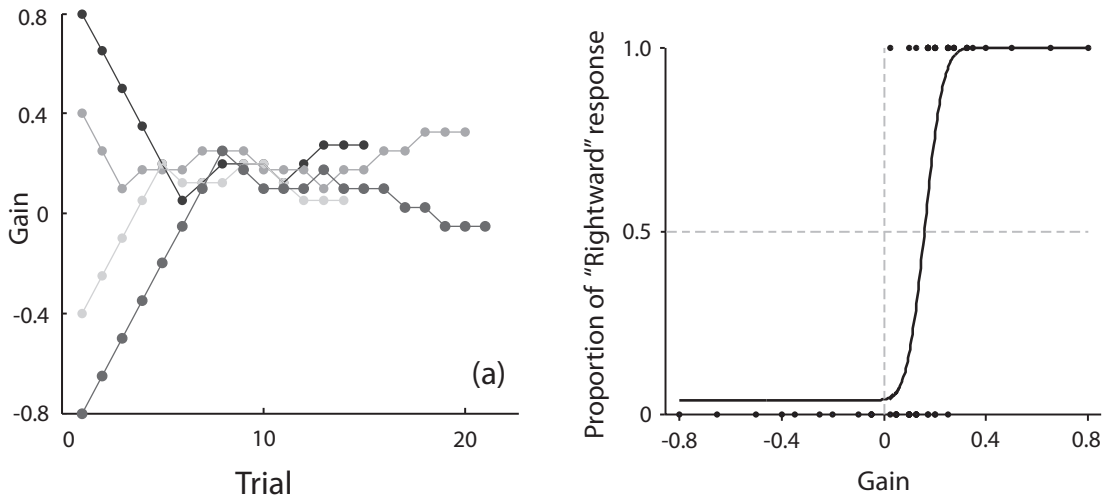

Figure 4.3: **Performance of a typical subject.** Panel **a** represents the four-interleaved staircases used to extract the PSE and the JND after the fitting of a Cumulative Gaussian (Panel **b**).

translation was more biased as revealed by a *Welch t-test* for unequal-variance samples ($t_{(7)} = -6.523, p < 0.0001$). Moreover, the JND was significantly larger (Figure 4.4 *b*) for the translation condition with respect to the rotation condition ($t_{(7)} = -2.6183, p < 0.05$).

4.3.2 Experiment 2

Rotations and translations of the head are encoded by separate receptors in the peripheral vestibular system and the behavioral differences found in Experiment 1 may reflect a different sensitivity of these two systems. Nevertheless, there is also the possibility that the rotation and translation conditions differed in terms of the visual input. For this reason we conducted a second experiment were the observer passively looked at the very same visual stimulation experienced during the active task. By keeping the visual stimulation the same and removing the extra-retinal information available to the active observers in the Experiment 1, we were able to separate the contribution of retinal and the extra-retinal signals (linear and angular motion) to the detection of visual motion.

The bar plots in Figure 4.5 shows the average of the JNDs for the rotation and the translation blocks. A Welch t-test for unequal-variance samples revealed no

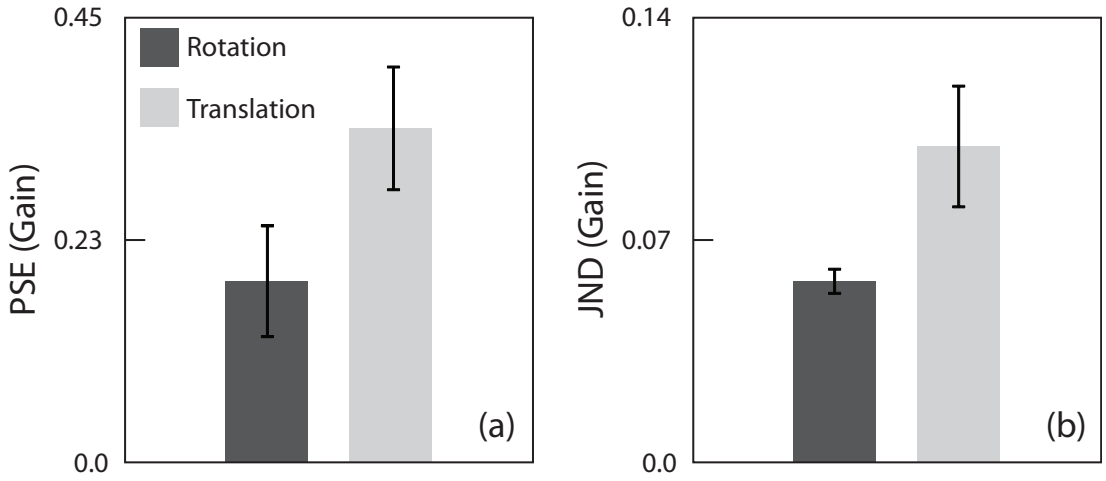

Figure 4.4: **EXPERIMENT 1:PSE and JND results.** Panel **a** represents the PSEs for the Rotation (left bar) and Translation (right bar) condition. Panel **b** shows the difference in JND measure for the Rotation (left bar) and Translation (right bar).

difference between the translation and rotation conditions ($t_{(6)} = -0.3841, p = 0.999$)[1].

4.4 Discussion

Although every movement of the head is associated with the visual displacement of the images on our retinae, we normally perceive a stable world around us. Here, in a series of experiments, we studied how, and to what extent, this compensation for retinal motion is achieved during linear and angular head motion. Our previous experiment (*Chapter 3*) suggested a difference between linear and angular motion of the head in the $3D$ perceptual interpretation of ambiguous optic flow stimuli. Linear ego-motion signals were found to be poorly integrated with the

[1]We applied the Bonferroni correction for simultaneous Null-Hypothesis testing. Also, in this Experiment we excluded 1 participant because she was able to perfectly separate standard and comparisons stimuli: the sigmoid function looked like a *Step Function* and we couldn't extract a meaningful estimate for the JND. Nevertheless, Psignifit provided an estimate for the JND and if that estimate was included in the general analysis, the overall results were not affected. Therefore, we can safely assume that the deletion of the outlier doesn't introduce a systematic error.

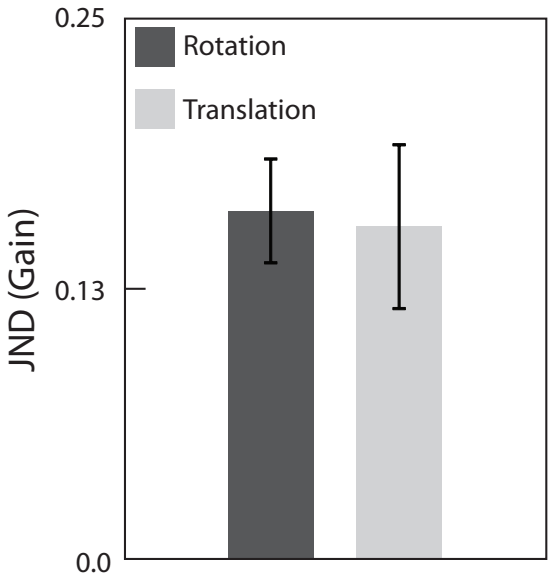

Figure 4.5: **EXPERIMENT 2: PSE and JND results.** JND differences for the Rotation (left bar) and Translation (right bar) conditions.

visual inputs whereas angular motion was found to be correctly estimated and employed in the analysis of the optic flow.

In the context of the present experiment, the first important finding is that, in order to be perceived as stable, the object has to move in the same direction as the observer. This is true both for head translation and head rotation. Interestingly, the amount of bias found here in the linear motion is similar to the bias found by other Authors for the forward motion (Wexler [2003]). Furthermore, since linear and angular movements of the head are sensed from different receptors, the (otoliths and semicircular canals - Kandel et al. [2000]), we hypothesized a different sensitivity for absolute object motion detection for the two head movements (Jaekl et al. [2004], MacNeilage et al. [2010]). Indeed, we could find a difference, in terms of perceptual stability, between translation and rotation movements. The translational movements are associated with both a more biased and less precise estimate of object's motion.

Overall, these results are in agreement with the findings of Jaekl et al (2004). One minor difference concerns the effect of yaw-rotations on the visual stability: contrary to the previous study, we could find a significant difference from the

veridical solution (to be perceived as stationary, the stimulus had to move). This may be due to a number of reasons. One simple explanation involves the different stimuli employed in the previous research: our participants judged the motion of a small object whereas in the previous research the observers were engaged in a wide-view virtual-environment. Another, equally valid, explanation concerns the different psychophysical procedures that may be corrupted by different systematic errors. Differently from Jaekl et al (2004), who used the method of adjustment, we employed the up-and-down staircase method which is less prone to hysteresis errors.

The difference between translations and rotations is particularly interesting because the perceptual judgment of object motion during self-motion requires the relative comparison between retinal and extra-retinal signals and therefore any asymmetry may suggest different underlying computations. Unfortunately, having found a bias in the perceived stability doesn't allow us to conclude which estimate (retinal, extra-retinal or both) are driving the phenomenon. In this respect, one important contribution of the present study relative to the previous ones is provided by our second experiment were passive observers were presented with the very same visual stimulation that they themselves generated during Experiment 1. This experiment allowed us to isolate the selective contribution of the extra-retinal signals, because the participants were not moving their heads (no vestibular information was available) meanwhile the visual input was kept constant. This manipulation allowed us to draw direct comparison with the first experiment. Since at the central level of the parieto-insular vestibular cortex responses to linear and angular motion were found to be equally represented (Chen et al. [2010]), we hypothesize that this difference must be searched in the early stages of signal processing. Therefore, because no systematic difference was found between the passive-translation and the passive-rotation we can conclude that the difference found in the active experiment can be ascribed to different sensory sensitivity of the peripheral vestibular system.

Chapter 5

Interaction between optic flow gradients and translational components: a parametric study

Abstract

In previous chapters we have shown that translational components of the head-centric optic flow are employed for the perceptual recovery of the tilt orientation of a planar surface: the perceived orientation of a planar surface rotating about an axis coplanar to the surface undergoes a 90° flip whenever a translational component orthogonal to the axis of rotation is added to the optic flow. This result was very robust and was found both in active and passive vision. The finding, however, was often asymmetric with respect of the direction of translation (horizontal vs. vertical). Therefore, it was important to study how and to what extent translations and the gradients of the velocity field interact. This goal was achieved by parametrically manipulating the information content conveyed by the gradients and by the translations. Two gradients (weak and strong) and three velocities (slow, medium and fast) and two directions of translations (horizontal vertical) were tested. The results showed that the gradients and the translational components of the optic flow interact to define the perceived tilt. Stronger gradients are less influenced, compared to weaker gradients, by the translational components of the head-centric optic flow. Similarly, the velocity of translation modulates the perceived surface orientation: faster speeds are associated with a stronger effect while during slower translations the effect is less pronounced. Finally, we found an asymmetry in effectiveness of vertical and horizontal translations components of the optic flow. Although this process doesn't guarantee a veridical performance, the results are consistent with the employment of a perceptual strategy that derives the tilt direction from the interaction of the gradients and the translational components of the optic flow.

5.1 Introduction

Our previous chapters showed several important facts about the perception of surface tilt direction from the Optic Flow (OF). One of the main findings concerns the role of the head-centric frame of reference for the analysis of the OF. If an OF is translating, relative to the sagittal and horizontal planes of the head, then the translation is employed for the recovery of surface tilt direction[1]. This is true even if there are no *retinal* translational components (because the eye keeps a steady fixation on the center of the stimulus where the OF vanishes). The effects of head-centric translations was present both during active and passive vision. Overall, the results were in line with the predictions of our modeling work: (i) a translation orthogonal to the direction of the gradient bias the perceptual interpretation of the surface orientation while (ii) a translation parallel to the direction of the gradient, at most, reinforces the same perceptual interpretation. However, the empirical effects of the horizontal and vertical translations were largely asymmetric and depended on the concurrent simulated OF gradient. Therefore it was important to understand how and to what extent the OF translations interacted with the gradient of the velocity field in order to specify the tilt direction.

In order to achieve this goal, we conducted a parametric study where horizontal and vertical translations interacted with horizontal and vertical gradients. The strength of the gradient (high and low) was manipulated as well as the velocity of translation (slow, medium and high velocity) and the direction of translation (horizontal and vertical). In agreement with our model we found that OF translations interact with the gradients of the velocity field for the recovery of surface tilt:

- the perception of a vertical gradient was biased by the presence of a horizontal translation; similarly the perception of a horizontal gradient was biased by the presence of a vertical translation;
- this effect was, however, asymmetric as the vertical translations were less effective in inducing a bias when compared to the horizontal translations;

[1]The tilt of a slanted surface is the projection of its normal on the fronto-parallel plane.

- interestingly, vertical translations uncover the effect of the velocities tested: faster translation induced a stronger bias relative to slower translations. This phenomenon, however, was difficult to detect in the horizontal translations because the performance was at ceiling and showed very little modulation;

- finally, the two gradients were differentially affected by the vertical translations: the stronger gradient was less affected while the weaker showed a bigger influence of the translations.

These empirical results provide support to our head-centric model. The OF translations and the gradients of the OF interacted in a complex fashion, to define the perceived tilt direction. The outcome of this process doesn't necessarily guarantee the veridical recovery of the surface's tilt.

5.2 Methods

5.2.1 Participants

Fourteen volunteers took part in the experiment: age range 20-36 (including the author and thirteen naïve observers), ten females. All had normal or corrected-to-normal vision and provided written informed consent as directed by the local Ethical Committee.

5.2.2 Apparatus

The same apparatus described in *Chapter 2* was employed in the present study. The Viewing distance was adjusted at 418.5 mm.

5.2.3 Visual stimulation

Similar to previous experiments, the stimuli employed here were simulating the projection of a small (approximately $6.8^\circ \times 6.8^\circ$) random dots planar surface. The surface was slanted and rotating either about a horizontal or vertical axis of

rotation which was passing through its center. This configuration corresponds to the following tilt directions: 0°, 90°, 180° or 270°.

Another variable manipulated in the experiment was the gradient of the velocity field (i.e. the *def*, Koenderink [1986]) generated by the rotation of the surface. The *def* was kept constant for all the duration of the trial by non-linearly rotating the surface with to the following equation $\arccos(\exp(-0.346574 + 0.5k - k(t))$. k is a constant representing the desired value of *def* and t is the normalized time. In the experiment we employed two *def* values 0.3 rad/sec and 0.5 rad/sec, this means that the initial slant values were 35° and 25° for the 0.3 and 0.5 *def* respectively. Whereas, the final slant values were 53° an 57° for the 0.3 and 0.5 *def* respectively. In both cases the instantaneous slant at half of the temporal interval was 45°.

Finally, all our displays were translating on the screen. For each stimulus frame, the translation was achieved in two steps: (i) centering the stimulus display on a simulated moving point of view (located at a constant distance of 418.5 mm) and then (ii) projecting on the screen the dots of the simulated planar surface (generalized perspective pinhole model, with the simulated point of view used as the Center Of Projection). Therefore, the motion of the object on the screen was jointly linked to the motion of the simulated camera which was translating with the characteristics specified by the Factors of the experiment (see below section 5.2.5).

5.2.4 Procedure

Participants were tested individually in a dark room. The whole procedure lasted approximately one hour. The experiment started only after 5 min of dark adaptation, when the observers head was positioned on the chinrest, at the position of 418.5 mm from the monitor screen. At the beginning of each trial, a moving red fixation mark was shown in the center of the screen and the observer was required to track the movement of the point. After two cycles of translation, when the dot reached one boundary (85 mm away from the center) and reversed its movement in the opposite direction the stimulus was displayed replacing the fixation mark. The simulated random-dot planar surface remained visible for the

entire trajectory (170 mm) until it reached the other edge. The stimulus lasted on the screen 2 sec, 1 sec or 0.77 sec according to the velocity condition tested. At that point the stimulus was replaced by a white fixation point in the center of the screen. Participants were instructed to press the keyboard buttons indicating the direction of the perceived tilt (left-right or up-down).

5.2.5 Design

The perceived tilt direction (dependent variable) was studied as a function of four factors: two Gradient Directions (Horizontal and Vertical), two Gradient Strengths (high and low, 0.5 and 0.3 respectively), two Direction of Translation (Horizontal and Vertical) and three Velocities of Translation (85 mm/sec, 170 mm/sec, 221 mm/sec). This resulted in a $2 \times 2 \times 2 \times 3 \times$ within-subjects design. Each stimulus was viewed 24 times for a total of 576 trials.

5.3 Results

Figure 5.1 and Figure 5.2 show the proportions of the "horizontal" responses as a function of the different levels of the velocity tested. Both in **Panel a** and **Panel b**, it is immediately apparent the interaction between the Horizontal and Vertical Translation (empty and filled bars, respectively). In **Panel a** the Vertical Gradient of the velocity field was perceived as (i) a horizontally tilted surface when the surface was undergoing a Horizontal Translation or (ii) a vertically tilted surface if the OF Translation was along the vertical axis. On this occasion, it seems that the three levels of the velocity tested were not different, although this might have happened because the participant's performance was very close to the maximum and minimum of the scale.

A similar pattern is found in **Panel b** where the stimulus consisted in a Horizontal gradient. This stimulus was perceived as (i) a horizontally oriented surface when it was translating laterally while (ii) it was seen as a vertically oriented surface if it was translating vertically. Interestingly, faster stimuli led to a more robust perception of a vertically oriented surface when compared to slower stimuli. In this case, the stronger Gradient of the velocity field (figure 5.2, **Panel**

b), was also less affected by the translation compared to the weaker Gradient of the velocity field (figure 5.1, **Panel b**).

The overall modulation due to the different velocities tested was, however, very weak. This effect, or lack of thereof, might be due to the relative strength of Translational velocities compared with the strength of the gradient. Another important point concerns the differential effect of Horizontal and Vertical Translations and how these translations are combined with the gradients of the OF. The Horizontal Translation, when combined with a Vertical Gradient was very effective in inducing a bias. In contrast, when the Vertical Translation was associated with a Horizontal Gradient, the bias was less pronounced. This asymmetry may reflect very basic physiological mechanisms of motion detection (Raymond [1994]).

The results provide a further demonstration that Translational components of the OF are indeed employed for the perceptual recovery of the surface tilt. This finding is not consistent with the idea of a retinotopic OF, in fact, during the translations, the eye keeps a steady fixation on the center of the stimulus where the OF vanishes (there are no translational components). The results are instead consistent with a framework based on the head-centric OF. This framework indicates that the Translational components interact with the gradients of the velocity field to specify the orientation of the surface. This interaction however, doesn't necessarily guarantee a veridical performance.

Having described the results we can now review the statistical modeling part. Just like we did in previous chapters (*Chapter 2 - 3*), we propose that the perceived tilt is derived through the computation of the angle θ by combining the gradients and the translational components of the head-centric OF. More formally, the binary outcomes (horizontal vs. vertical tilt) were modeled by the following probit function (ϕ) with a center (μ) and scale (σ) parameter:

$$y \quad = \phi_{\mu\sigma}(\theta_{w_{1\ldots n}}(a_1, a_4, a_2, a_6)) \tag{5.1}$$

The continuous predictor was the variable θ which, in turn, takes as input variables the translational components (a_1, a_4) as well as the gradients of the OF

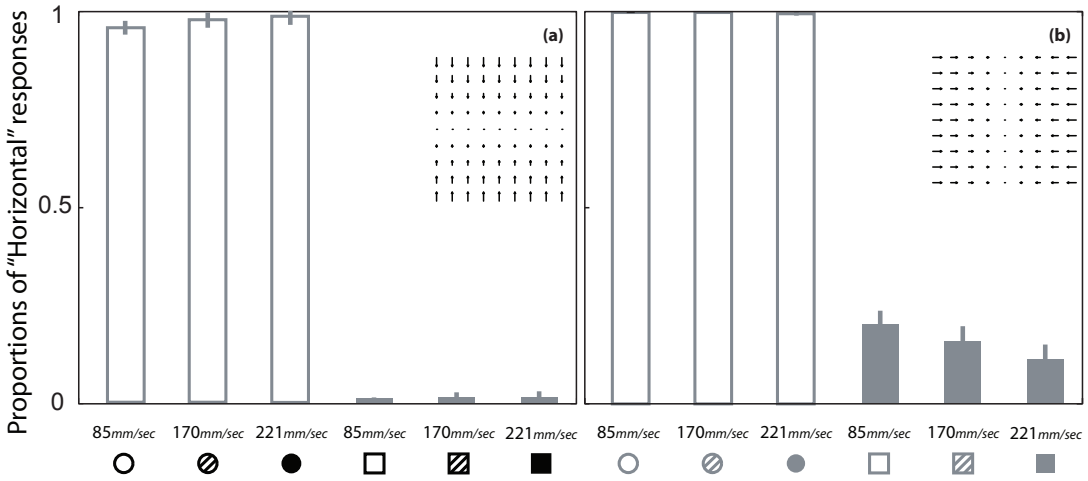

Figure 5.1: **The results for the "weak" Gradient.** The proportions of "horizontal" responses is plotted as a function of the OF Translation factor both for the vertical (**Panel a**) and horizontal (**Panel b**) gradient of the velocity field. The empty, the striped and the filled symbols identify the three-levels of the Translation Velocity factor; the gradient of the OF is identified with colors (black-vertical, grey-horizontal). Finally, the circles are associated with the Horizontal Translation and the squares with the Vertical Translation; see figure 5.3.

(a_2, a_6). The angle θ accepts a series of parameters ($w_{1\ldots n}$) associated with the numerator and the denominator of the equation. With this parameterization in mind, we then defined a set of nested non-linear regression models to fit the empirical data. Following the procedure of model simplification we found and evaluated the best model to extract the significance (in term of AIC and Transformed Likelihood Ratio test, Akaike [1974], Sugiura [1978], Hurvich and Tsai [1989], Kingdom and Prins [2010]). The best model contained three parameters: the center and scale of the cumulative distribution and one extra-parameter[1]. This regression model outperformed the null model ($\chi^2_{(2)} = 61.33, p < 0.001$). Interestingly, this time, we had to retain the parameter μ, that is the center of the distribution. In the previous chapters this parameter was always kept fixed at the value 45 °. This, fact was likely due to the interaction between the Direction of Translation and the Direction of the Gradient: it seems that the Horizontal Translation was more effective in biasing the Vertical Gradient respect to a Ver-

[1] associated with the numerator of the equation of θ.

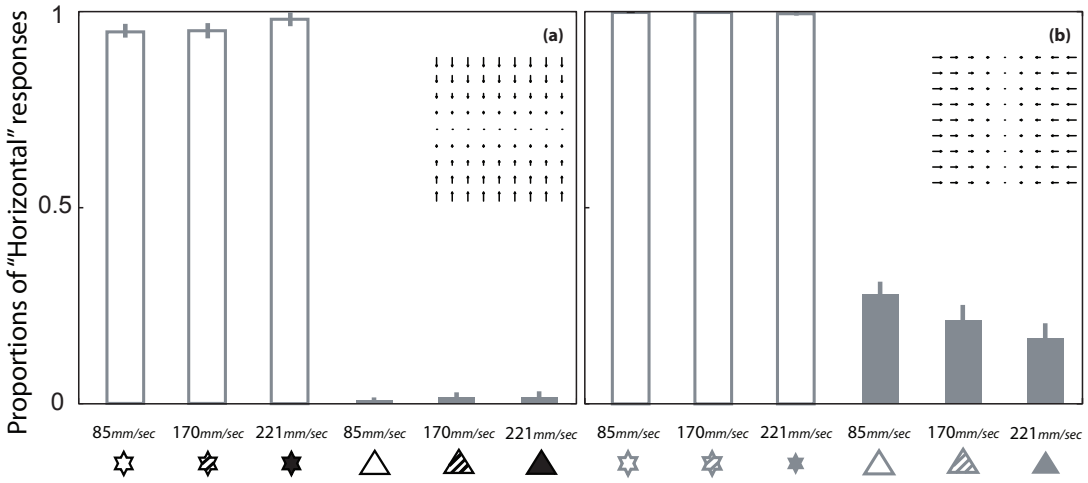

Figure 5.2: **The results for the "strong" Gradient.** The proportions of "horizontal" responses is plotted as a function of the OF Translation factor both for the vertical (**Panel a**) and horizontal (**Panel b**) gradient of the velocity field. The empty, the striped and the filled symbols identify the three-levels of the Translation Velocity factor; the gradient of the OF is identified with colors (black-vertical, grey-horizontal). Finally, the stars are associated with the Horizontal Translation and the trianlges with the Vertical Translation; see figure 5.3.

tical Translation combined with an Horizontal Gradient. The prediction of our model can be inspected in figure 5.3, **Panel a**, where the Observed results are plotted as a function of the Predicted ones. The different symbols are associated with the different levels of the factors manipulated in the present experiment. The black and grey colors are associated with the Vertical and Horizontal Gradient respectively; circles and stars indentify respectively Horizontal Translations for the weak and strong Gradient while squares and triangles are associated with the Vertical Translation for the weak and strong Gradient respectively. Finally, empty symbols are associated with the velocity 85 mm/sec, stripes symbols with 170 mm/sec, and filled with 221 mm/sec.

5.4 Discussion

The measurements of the perceived tilt direction from passively observed OFs has produced three important results. First, the study confirms the role of the head-

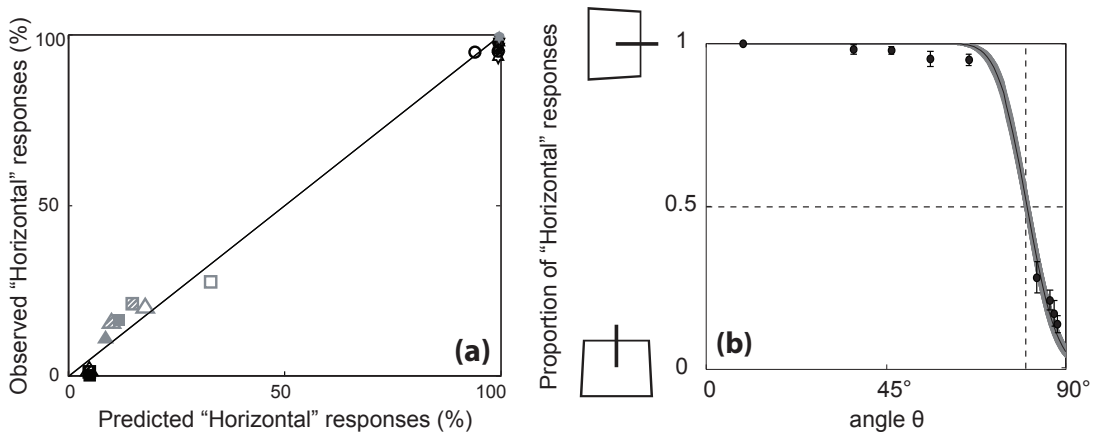

Figure 5.3: **Model's Predictions.** The Predicted versus Observed proportions are plotted in **Panel a** (see the text for the symbol codes). **Panel b** shows the psychometric function that relates the angle θ with the perceived surface tilt. Light lines indicate the uncertainty in the regression.

centric OF translations in the perceptual recovery of surface tilt. This effect was found in all the conditions tested. Second, this phenomenon was modulated by the relative strength of the velocity gradient and the translation velocity. Finally, the influence of the vertical translations was less effective when compared to the horizontal translations.

The overall effect of the head-centric translations on the perceptual interpretation of OF stimuli is consistent with our model and contrasts with the classical view of a retinotopic frame of reference. The strategy employed by the observers doesn't necessarily guarantee the recovery of the veridical $3D$ structure; nevertheless is biologically plausible as it is consistent with our behavioral data (Domini et al. [2012]).

When the stimulus was translating along the horizontal axis, the performance of our participants was very close to the extreme of the scale (i.e. 100%), and therefore very little modulation of the factors employed here was found. However, when the performance was far from the extreme, a very interesting picture emerged. The perceptual strategy employed by the observers was flexible and depended on the simultaneous information conveyed by the velocity gradients and the concurrent translations. The stronger velocity gradient ($0.5 rad/sec$) was

found to be less affected by the three velocities tested while the weaker velocity gradient ($0.3 rad/sec$) was more subject to the influence of the translations. This finding is consistent with the idea that the velocity gradient is of paramount importance for the recovery of the surface orientation from the OF (Domini and Caudek [1999], Domini and Caudek [2003a], Domini and Caudek [2003b]). A similar effect was found when was manipulated the information content of the translation velocity. Faster translations induced a stronger bias relative to slower translations. Overall, these findings suggest a complex balance between the velocity gradients and the head-centric translations.

Finally, we report that the effect of translations was overall asymmetric as the vertical translations were less effective when compared with the horizontal translations. These findings are consistent with the directional anisotropy of motion detection witnessed by Raymond (1994): the sensitivity for motion detection was higher for stimuli that were moving horizontally than for stimuli moving vertically. Therefore the asymmetry reported in our experiment might be due to basic physiological mechanisms sub serving different functions.

Chapter 6

Summary

In order to effectively navigate in a complex environment, a biological system requires to encodes and process multiple spatial information. The visual sense, for instance, provides a rich source of information about the $3D$ layout of the environment as well as the direction of self motion. Visual information, however, is often ambiguous in terms of its perceptual interpretation and it therefore requires the combination of multiple sensory signals such as vestibular and proprioceptive for the case of the moving observer. The manuscript approached this problem by looking at how we perceive ambiguous geometrical shapes when we are moving in the environment. We asked which reference frame was employed for the analysis of the visual inputs. Also, we asked, how and to what extent, vestibular/proprioceptive signals are integrated with the visual information.

The perception of $3D$ shapes has a long tradition in psychophysics. In a series of seminal papers Wexler and coworkers (Wexler et al. [2001b], Wexler et al. [2001a], Wexler [2003], Wexler and van Boxtel [2005]), found that the perception of affine structures from ambiguous Optic-Flows (OF) is improved by the presence of extra-retinal signals. More recent studies, however, hinder the biological plausibility of that interpretation. These studies, focusing on the perception of the Euclidean structure, demonstrated that, when carefully controlled, the visual information *alone* explains most of the data from the active-vision experiments with little or no contribution of the extra-retinal signals (Caudek et al. [2011], Fantoni et al. [2010], Domini and Caudek [2003a], Fantoni et al. [2012]).

The present work enters this debate by re-examining the previous studies on

the affine structure perception. Contrary to what has been proposed previously, we advocate the idea that when the relevant input variables (i.e. head-centric OF, see below) are carefully controlled, also affine properties cannot be recovered in a veridical fashion.

Our approach is based on two assumptions and the experiments conducted here were devised to test these assumptions as well as the prediction of the model. The model is based on a head-centric, as opposed to retinotopic, representation of the OF. To understand this difference, consider the case of a moving distal object and a stationary observer. According to the retinotopic approach, the eye keeps a steady fixation on the object and therefore there are no translational components in the retinal image. In contrast, if the reference frame is centered on the head, the object has induced a translational component. Translational components were neglected in previous experiments while, according to our model, these components are employed during the analysis of the OF. The second assumption concerns the role of extra-retinal signals. We assume that only angular ego-motion signals are employed in the perceptual analysis while linear ego-motion information is ignored.

In *Chapter* 2 the experiment was carefully designed to disentangle the contributions of two variables that were confounded in previous versions of the experiment: (i) the linear ego-motion signals and (ii) the translations components of the OF. To achieve this goal, our participants were asked to judge the surface tilt in two conditions: (i) when a random-dot planar surface rotated about a stationary axis or vice versa (ii) when the axis of rotation was tethered to a coordinate system centered on the observers head, so as to eliminate the translational components of the head-centric OF. The novel result of the present investigation was that the advantage of active over passive vision, previously attributed to the extra-retinal signals, was found to be mainly associated with the translational components of the OF and not with availability of the linear ego-motion signals. This result clearly supports the role of the head-centric as opposed to the retinotopic OF.

In this experiment we could find a minor contribution of the extra-retinal signals. This leaves open the possibility that the visual system has access and may use extra-retinal information for the perceptual analysis of the OF. Therefore in the next step (*Chapter* 3) it was important to understand if the OF translation

induced by the angular or linear motion of the head differently affects the perceptual interpretation of surface tilt. In order to achieve this goal, the perceived tilt was measured under different conditions. In one case, the subjects were laterally moving and the translations of the head-centric OF were either absent or present (depending on whether the object was or was not kept aligned on the line of sight). In another condition, the same presence/absence of the head-centric OF translations was manipulated while the participants were rotating their head along the horizontal plane (i.e. yaw rotation). The pattern of subject's responses was markedly different for linear vs. angular ego-motion signals. The interpretation of the results is consistent with the view that: (i) linear ego-motion signals are not taken into account in the computation of the OF while (ii) angular ego-motion signals are precisely estimated.

This difference between angular and linear ego-motion signals was investigated further in *Chapter* 4. In particular, we calculated the amount of visual shift that can be tolerated for an object to be perceived stationary, when the observers are: (i) laterally translating or (ii) laterally rotating the head. By keeping constant the visual stimulation in the two conditions, we were able to directly compare the performance of the observers when they were translating or rotating the head. We confirmed that angular movements are associated with less biased and more precise estimates of object motion, whereas linear translation are biased and are associated with a more noisy representation of object motion.

Importantly, in the final study, (*Chapter* 5) we could find a small, but consistent evidence, in favor of the interaction between the translations of the head-centric OF and the gradients of the velocity field. The perception of the tilt direction was modulated by a flexible combination of these two variables.

The picture that emerges from this collection of experiments suggests that active and passive observers employ specific perceptual strategies to extract a $3D$ interpretation from motion patterns. This interpretation seems to be based on a head-centric representation of the OF. The primary ecological relevance of such representation may be associated with tasks that involve the navigation in complex environments. In this case a head-centric flow immediately signals the motion of the head in space without any further coordinate transformations (Beintema and van den Berg [1998]); in fact, computing the head-centric transla-

tions is an important step for the perception of heading direction (Royden et al. [1992], van den Berg [1992]).

Also, our $3D$ interpretation of motion patterns seems to rely only on a small subset of extra-retinal signals, namely the rotations of the head. Under normal circumstances, we rarely move linearly; in contrast we smoothly rotate the head and orient our body to attend a specific portion of the space around us. Following this insight, previous studies suggested that we are more accustomed to rotations of the head rather than linear translations (Jaekl et al. [2004]). We embrace this view in our model.

Finally, we found that the linear ego-motion signals are mostly ignored in the perceptual interpretation of the OF. This finding is consistent with a series of studies (Caudek et al. [2011], Fantoni et al. [2010], Fantoni et al. [2012]) although it is in contrast with the major belief in the literature (Wexler et al. [2001b], Wexler et al. [2001a], Wexler [2003], Wexler and van Boxtel [2005]). Our interpretation is that, the translation of the observer is estimated from the head-centric OF and it is not directly estimated from the linear ego-motion signals Domini et al. [2012]. This perceptual strategy doesn't necessarily guarantee the recovery of the veridical $3D$ structure. However, it may constitute the most efficient choice for the human observers.

Appendix A

.1 Overview

The retinal image transformation produced by the relative motion of the observer and a planar surface, can be described in terms of $2D$ motion of discrete texture elements belonging to the $3D$ surface. The so generated motion pattern, also called Optic Flow (OF), is usually represented as an instantaneous velocity field, in which each vector corresponds to the optical motion of a point in the environment. Figure 1 describes the geometrical preliminaries for the analysis of the OF. The orientation of the planar surface P of coordinates (X, Y, Z) in $3D$ space can be described in terms of its slant (σ) and tilt (τ). Slant is defined as the angle between the line of sight and the normal to the surface (i.e. slant is equal to $0\,^\circ$ if the surface lies perpendicular to the line of sight). Tilt is defined as the angle between the projection onto the image plane of the normal to the surface and the *x-axis*.
Under the orientation of the axis specified in Figure 1 , the spatial depth gradients are:

$$\begin{aligned} \mathbf{g_x} &= \tan(\sigma)\cos(\tau) \\ \mathbf{g_y} &= \tan(\sigma)\sin(\tau) \end{aligned} \tag{1}$$

The $3D$ motion of P can be decomposed in Rotations ω_R around the eye of coordinates $(\omega_{R_x}, \omega_{R_y}, \omega_{R_z})$, and Translations ω_T of coordinates $(\omega_{T_x}, \omega_{T_y}, \omega_{T_z})$. Finally, the vector t of coordinates $(\omega_{T_x}, \omega_{T_y})$ represents the direction of the frontal

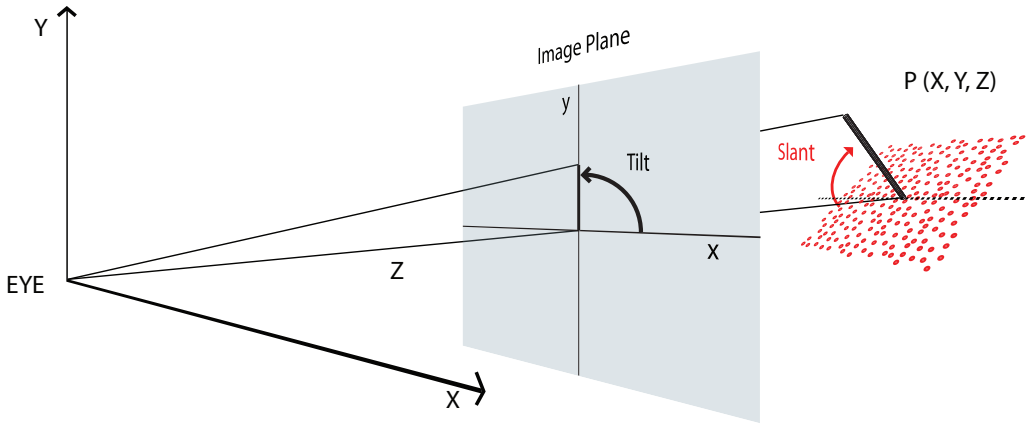

Figure 1: **The coordinate system for the analysis of the Optic Flow.** The nodal point of the eye is represented as the center of projection (EYE) corresponding to the origin of the (X, Y, Z) coordinate system, with the $Z\text{-}axis$ coinciding with the line of sight. The (x, y) plane, is the image plane which is coplanar to the (X, Y) plane and centered on the $Zaxis$. The orientation of the plane P is described by two parameters: the slant (red angle between the line of sight and the normal to the surface) and the tilt (which is the angle defined by the projection of the plane's normal onto the image plane, and the x-axis on the projection plane itself - in this example the tilt is 90 °).

translation; t describes with the $x\text{-}axis$ on the image plane the angle θ so that:

$$\begin{aligned} \omega_{T_x} &= t \cdot cos(\theta) \\ \omega_{T_y} &= t \cdot sin(\theta) \end{aligned} \tag{2}$$

If the image plane is located at a unit distance from the origin along the $Z-axis$, the retinal position of point belonging to the planar surface is given by:

$$(x, y) = (X/Z, Y/Z) \tag{3}$$

and the equation of the planar surface P becomes:

$$P = (1/Z)\mathbf{g_x}X + (1/Z)\mathbf{g_y}Y + (1/Z) \tag{4}$$

Using this coordinate system, Longuet-Higgins and Prazdny (1980 and Longuet-Higgins [1984]) provided a formal analysis of the specific patterns of optical transformations that are produced by different types of rigid body motions. Following the rigidity assumption (Ullman [1979]) Longuet-Higgins and Prazdny (1980 and Longuet-Higgins [1984]) provided the equations that relate a plane orientation and $3D$ motion with the local OF. Although in the original formulation the OF was expressed by second order polynomials, it must be said that in the small neighborhood of the origin of the image plane (x, y) the contribution of the second order terms of the OF becomes negligible. Therefore, in the following we will present a simplification of the perspective projection called *affine scheme* which does not take into account the second spatial derivative terms:

$$\begin{aligned} U_X(x,y) &= a_1 + a_2 x + a_3 y \\ V_Y(x,y) &= a_4 + a_5 x + a_6 y \end{aligned} \tag{5}$$

The motion of the projected texture elements of the surface on the image plane generates a velocity field vector. Ux and Vy describe the instantaneous velocities of any given texture elements along the horizontal and vertical direction respectively. These two instantaneous velocities are modulated by six coefficients ($a1$ to $a6$) that depend on the $3D$ structure and relative motion between the object and the point of view (see also Zhong et al. [2006]):

$$\begin{aligned} a_1 &= \omega_{T_x} + \omega_{R_y} \\ a_2 &= -\omega_{T_x} g_x - \omega_{T_z} \\ a_3 &= -\omega_{T_x} g_y - \omega_{R_z} \\ a_4 &= \omega_{T_y} - \omega_{R_x} \\ a_5 &= -\omega_{T_y} g_x + \omega_{R_z} \\ a_6 &= -\omega_{T_y} g_y - \omega_{T_z} \end{aligned} \tag{6}$$

An example will clarify the contribution of each one of these coefficients. Having in mind the geometry and the setting described in Figure 1, we can imagine

different types of object and observer relative motion. Figure 2 isolates the contribution of each one of these coefficients to the overall OF.

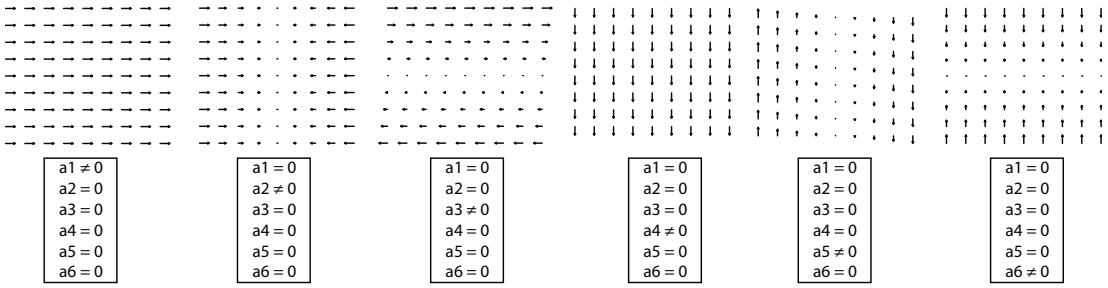

Figure 2: **The velocity field under different coefficient values.** The contribution of each parameter is explored in six different examples. a_1 and a_4 describe pure translations of the image, a_2 and a_6 describe pure gradients of compression/expansion whereas a_3 and a_5 are the shear components.

It should be noted that recovering the $3D$ motion and structure of from the OF yields to an infinite number of possible solutions (i.e. an ill-posed inverse problem) an infinite combination of $3D$ geometry and motion are compatible with the same OF. Indeed, there are 6 equations and 8 unknowns, namely the two gradients g_x and g_y and the two vectors of translation ω_T and rotation ω_R. Koenderink and van Doorn (1975) showed that the local instantaneous flow of continuous velocity field can be decomposed into four elementary transformations (see Figure 3): a rigid image rotation (*curl*), an isotropic expansion or contraction (*div*), and two components of shear along vertical (def_1) and oblique axes (def_2). These four geometrical transformations are related to the coefficients of the affine scheme as follows:

$$\begin{aligned} curl &= a_3 - a_5 \\ div &= a_2 + a_6 \\ def_1 &= a_2 - a_6 \\ def_2 &= a_3 + a_5 \end{aligned} \tag{7}$$

Each of these components of the first order OF is specified by a scalar field. The *curl* gives an indication of how the field swirls (how much the image Z-rotates) in

the vicinity of that point. The divergence (*div*) tells us, at each point, how much the field diverges away from that point. The two shear components quantify the amount of shape change and can be summarized by a unique value called *def* (computed as follows: $\sqrt{def_1^2 + def_2^2}$). The *def* is a differential operator yielding a scalar field that tells us how much the field contracts in one direction and simultaneously expands in the orthogonal directions.

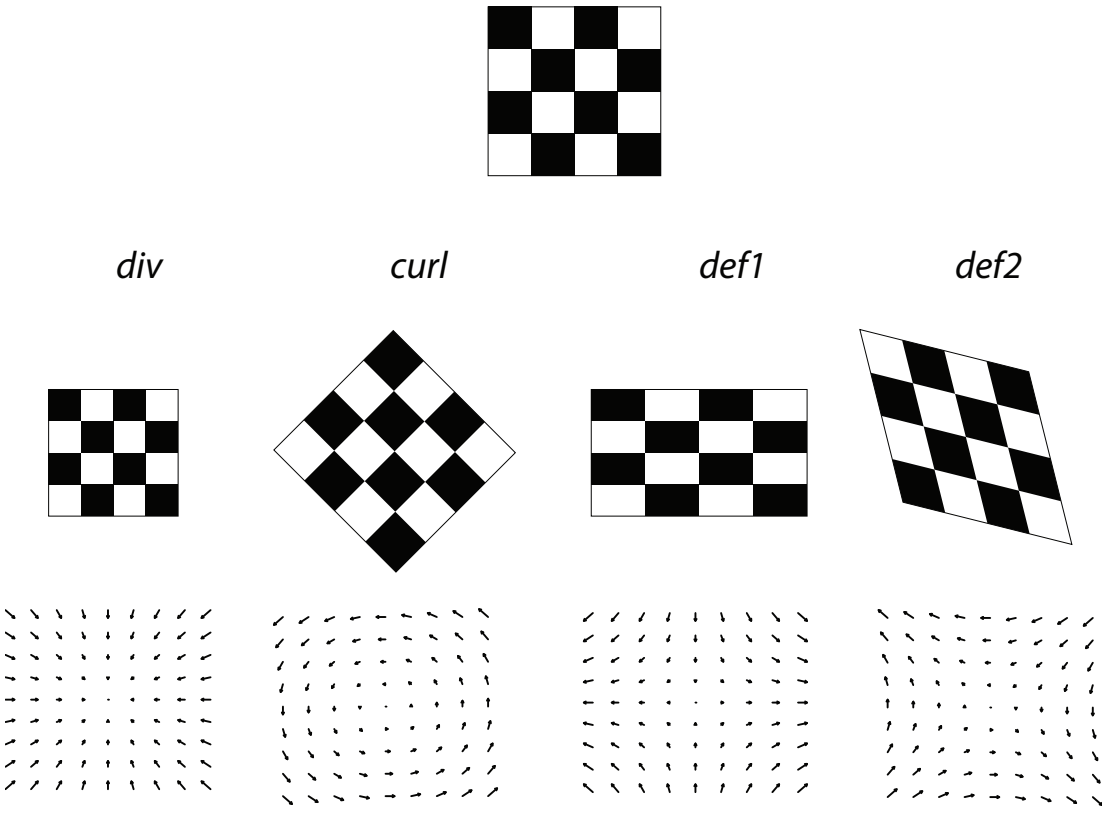

Figure 3: **Geometrical elementary transformations.** The checkboard (top row) undergoes four elementary transformations (middle row): *div*, *curl*, def_1, def_2. Each image transformation is summarized by the velocity field (bottom row). Adapted from Koenderink (1986).

.2 Recovering the tilt direction in the affine scheme

In the following we demonstrate the relationship between the tilt direction (τ) and the direction of the observer's translation (θ) on the fronto parallel plane (x, y). The system of the affine scheme of equation 6 after substituting with the equation 2 and equation 7 is equivalent to the system:

$$\begin{aligned} def_1 = & \quad -t \cdot tan(\sigma)cos(\theta + \tau) \\ def_2 = & \quad -t \cdot tan(\sigma)sin(\theta + \tau) \\ div = & -t \cdot tan(\sigma)cos(\theta - \tau) - 2\omega_{T_z} \\ curl = & \quad t \cdot tan(\sigma)sin(\theta - \tau) - 2\omega_{R_z} \end{aligned} \tag{8}$$

The first two equation 8 imply that $t \cdot tan(\sigma) = \sqrt{def_2^2 + def_1^2}$ (where $t \cdot tan(\sigma) = def$, Domini and Caudek [1999]).

Now let's consider an OF characterized by a gradient of pure horizontal expansion (or compression). This corresponds to the example in Figure 2 (second panel from the left). Since all the coefficients of the affine scheme are equal to zero with the only exception of the a_2, it follows that: (i) $def_2 = 0$, (ii) $def_1 = a_2$ and (iii) $def = def_1 = a_2$. Now, substituting def_1, def_2 and def in the first two equation 8,we are left with the system:

$$\begin{aligned} cos(\theta + \tau) = & \quad -1 \\ sin(\theta + \tau) = & \quad 0 \end{aligned} \tag{9}$$

which means that $\theta = -\tau$ or $\pi - \tau$. In other words, the direction of the frontal translation (θ) of the observer is informative about the tilt direction (τ) of the surface. θ can also be expressed (equation 2) as the $\arctan(\omega_{T_y}/\omega_{T_x})$. From equation 6 we re-write ω_{T_x} as $a_1 - \omega_{R_y}$ and ω_{T_y} as $a_4 + \omega_{R_x}$. This implies that:

$$\theta \quad = \arctan\left[\frac{(a_4 + \omega_{R_x})}{(a_1 - \omega_{R_y})}\right] = -\tau \ \ or \ \ (\pi - \tau) \tag{10}$$

Therefore, θ (and thus the tilt τ) can be estimated (up to 180° reflection) by

knowing the translational components of the OF (a_1 and a_4) and the rotation of the head about the vertical and horizontal axis (ω_{R_y} and ω_{R_x}). This formulation does not include the gradients of the velocity field, that we already know are very important for the recovery of surface shape and motion (Todd and Bressan [1990], Domini and Caudek [1999], Domini and Caudek [2003a], Fantoni et al. [2010], Caudek et al. [2011], Fantoni et al. [2012]). Moreover in the previous example, when all the coefficients are zero except a_2, it would not be possible to compute the angle θ. So, equation 10 has been extended to include the velocity gradients. The idea is very simple and refers to the concept of local motion energy (Morrone and Owens [1987]). In the case of a pure horizontal compression/expansion OF, just like our current example, the OF vanishes at the origin (the velocity field is zero). However, consistent with our idea of head-centric OF (see below), if the velocity is pooled outside the small neighborhood of the origin, then the velocity field has a horizontal translatory components. This intuition is formalized by the following equation:

$$\theta \quad = \arctan\left[\frac{|a_4+\omega_{R_x}|+a_6}{|a_1-\omega_{R_y}|+a_2}\right] = -\tau \;\; or \;\; (\pi-\tau) \tag{11}$$

In summary, our model states that the surface's tilt direction can be recovered by computing the angle θ. The angle θ in turn can be estimated knowing the translational components of the OF (a_1 and a_4), the gradients of the velocity field (a_2 and a_6) and the angular motion of the head about the horizontal and vertical axes (ω_{Rx} and ω_{Ry}). It is important to stress two points about this process. First, contrary to previous suggestions, it doesn't include the analysis of the translations of the head (ω_T). Second, the output of this analysis doesn't necessarily guarantee the veridical derivation of the surface tilt. This model allows us to make clear and counterintuitive predictions about the perceived tilt direction from the self generated OF; these predictions cannot be reconciled with current theories described earlier. In the following we will examine several typical experimental situations and we will try to explain the contribution of the present model.

.3 The probabilistic interpretation of the OF

Before we examine the model we will frame the problem in probabilistic terms. More specifically, the interpretation of the three-dimensional shapes from the OF can be described in terms of a Bayesian model that takes into account: (I) the OF and (II) the relative motion between the observer and the surface. Within the Bayesian framework, each variable is described by a probability distribution with parameters specifying the center and the scale of the distribution. This approach allows us to incorporate different assumptions or beliefs about the variable object of study. At any given point in time the observer receives the OF specified by the system of equation 6. The optic flow is generated by the relative motion (translations and rotations, respectively ω_T and ω_R) between the observer and a planar surface of structure $\mathbf{G}$ (i.e. expressed in depth gradients along the horizontal and vertical direction respectively, g_x and g_y). A principle of our approach is that the visual system doesn't take into account the linear motion of the head for the perceptual interpretation of the OF. In contrast the rotations of the head enter the computations for the extraction of shapes from the OF. These assumptions are based on previous observations (Caudek et al. [2011]) and are now translated into probabilistic terms. ω_R is represented as a Gaussian distribution $P(\omega_R|\widehat{\omega_R})$ sharply peaked on the true value of the head rotation. Instead, ω_T is represented as a non-informative Gaussian centered on zero and widely spread across many possible value of the parameters space. This difference between ω_R and ω_T outlines the first difference between our approach and previous modeling work (Colas et al. [2007]) that assumed for both translations and rotations signals a sharply peaked distribution. Another important aspect that differentiate our model from previous is that the analysis of the OF is conducted in a head-centric frame of reference. This assumption is incorporated in the way in which the coefficients of the OF are calculated and is related to the computation of the angle θ as in equation 11. Indeed, in a head-centric frame of reference, each displacement relative to the sagittal and horizontal planes of the head will result in a modification of the coefficients a_1 and a_4 and in turn of the angle θ. This was not included in previous work which were based on a retinotopic analysis of the OF. In a retinotopic scheme, the eye compensates for the relative movements

of the objects and therefore deletes the translational components of the OF (i.e. a_1 and a_4 are zero).

Finally, our model further assumes that the OF specified by the coefficients of equation 6 is also corrupted by some Gaussian noise. Furthermore, a flat prior knowledge about the plausible $3D$ structures is incorporated for the g_x and g_y nodes.

Having described the constituents of our approach we can now formulate the problem more formally. The goal of a visual system (artificial or biological) is to estimate the three-dimensional layout, $\mathbf{G}$, from $\mathbf{a}$[1] and ω_T[2] and ω_R[3]. Thanks to Bayes rule this is equivalent to calculating the following posterior distribution:

$$P\left(g_x, g_y|\mathbf{a}\right) \propto P\left(\mathbf{a}|g_x, g_y\right) P\left(g_x\right) P\left(g_y\right) \tag{12}$$

where the likelihood is obtained by integrating over all possible values of ω_T and ω_R:

$$P\left(\mathbf{a}|g_x, g_y\right) = \int_{\omega_R} \int_{\omega_T} P\left(\mathbf{a}|g_x, g_y, \omega_R, \omega_T\right) P\left(\omega_R\right) P\left(\omega_T\right) \mathrm{d}\omega_R \mathrm{d}\omega_T \tag{13}$$

and the priors are the independent probabilities of $P(g_x)$ and $P(g_y)$.

In the following, this probabilistic problem is solved using a numerical approach. The analysis is conducted with JAGS, a freeware software that allows to build flexible Gibbs Samplers and perform Bayesian analysis using Markov Chain Monte Carlo simulations. The solution is approximated by sampling from the target distributions, that is to say the Bayesian posterior probability of equation 12.

Now, in order to fully understand the model and its predictions let's examine a series of examples. The first instance that we assess is the general case of passive vision. Consider again the case of a pure horizontal gradient of the velocity field (Figure 2 second panel from the left). This OF is ambiguous about the depth sign (i.e. the tilt direction) because of the invariants of the first order OF: there are

[1] namely: a_1, a_2, a_3, a_4, a_5, a_6 (equation 6)

[2] ω_{Tx},ω_{Ty},ω_{Tz}

[3] ω_{Rx},ω_{Ry},ω_{Rz}

two equally likely and opposite perceptual solutions. Now, what happens if this OF is also translating vertically on the image plane (that is to say, a translatory component that is orthogonal to the direction of the gradient)? From the classical standpoint, the eyes are always center and aligned on the center of the image, where the OF vanishes and therefore this manipulation shouldn't be effective. In contrast, our model suggests that the translation of the image is taken into account in the perceptual interpretation of the OF as the relevant input to the visual system is the head-centric OF (figure 4, **Panel b**). More formally we can write:

$$\omega_{Ry} = \omega_{Rx} = a_1 = 0$$
$$a_4 \neq 0$$
$$a_2 \neq 0$$
$$a_4 \gg a_2$$
$$\theta = \arctan\left[\frac{|a_4+\omega_{Rx}|+a_6}{|a_1-\omega_{Ry}|+a_2}\right] = \arctan\left[\frac{a_4}{a_2}\right] = 90\,^\circ \ or \ 270\,^\circ$$

Experimentally, for the passive observer, θ can be manipulated by adding a translatory component to the gradient of the OF. If the direction of translation is not orthogonal (as in the previous example) but parallel to the direction of the gradient of the velocity field, we don't expect any change in the perceived tilt. In the current example in fact both the gradient and the angle θ specify a horizontal direction of the tilt τ (from equation 11 we have $\theta = \arctan\left[\frac{0}{a_1+a_2}\right] = 0\,^\circ \ or \ 180\,^\circ$). The same results are obtained if we consider a gradient of pure vertical expansion (or contraction) and we add a translatory component to the OF. From, equation 11 we predict either: (i) a $90\,^\circ$ inversion in the perceived tilt when the translation is *orthogonal* to the direction of the gradient (i.e. horizontal) or (ii) the same perceptual solution when the direction of translation is *parallel* to the direction of the gradient.

Let's now turn our attention to the case of the active observer. In this context there are at least four relevant combinations that can be studied and that can be used to uncover the tenets of our model. In the following we will always refer to a stimulus with a pure vertical gradient of velocity (Figure 2 last panel on the

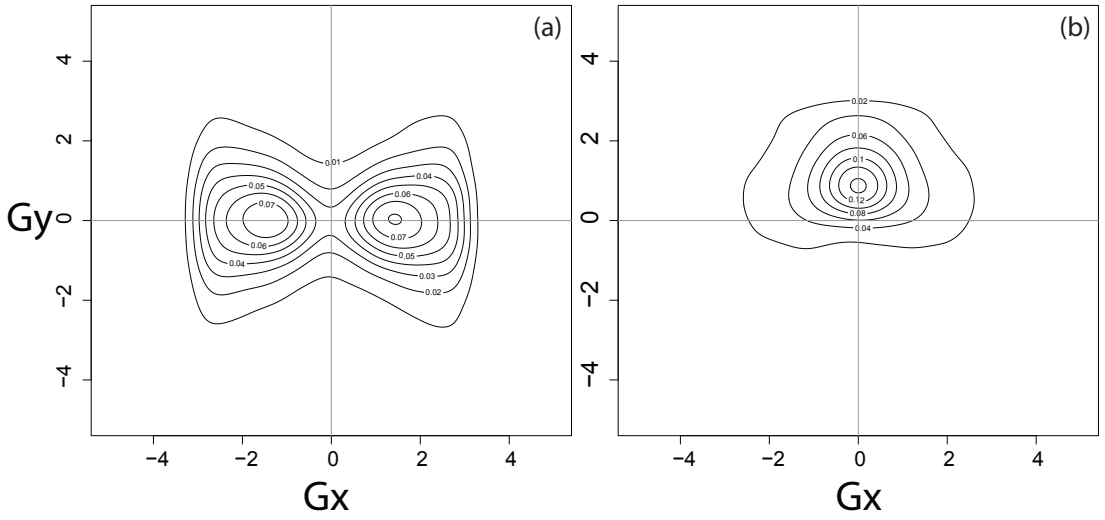

Figure 4: **Output of the model.** The isodensity curves identify the regions of the parameters space (g_x and g_y) that most likely generated the OF experienced by the observer. **Panel a** The OF is ambiguous with two possible tilt directions. **Panel b** shows what our model predicts in case a translatory components is added to the OF. In this specific case the direction of the OF translation is *orthogonal* to the direction of the horizontal gradient of the velocity field. Both the gradients and the direction of translation of the OF are used to derive the most likely $3D$ structure.

right).

First of all, consider the case of an observer that is translating laterally (from side to side). We will make the further assumption that the observer is not rotating the head and therefore ω_{R_y} and ω_{R_x} are both equal to zero. Now, consider the following experimental conditions:

- in the first scenario, the image is tethered to the line of sight of the observer, so that even if the observer is moving laterally, the sagittal plane of the head is always centered on the center of the image. In this particular scenario there are no translational components of the head-centric OF since there's no relative motion on the horizontal direction;

- in contrast, in the second scenario, if the image is stationary a relative motion is induced by the motion of the observer.

According to the main approach in this field, the two conditions shouldn't be dif-

ferent because the observer keeps a steady fixation on the target and incorporates the knowledge of body motion (thanks to the vestibular and proprioceptive systems) into the analysis of the OF. However, the knowledge about the ego-motion might be not available or it might be not employed by the visual system (Fantoni et al. [2010], Caudek et al. [2011]), Fantoni et al. [2012])) in the interpretation of the structure from motion. Indeed our model embraces this last view and considers the signals from the linear translation of the head not informative and poorly reliable. In terms of the Bayesian probabilistic model, this is translated into flat Gaussian distributions over the parameter space, with mean of 0 and large variability. Moreover, the model makes use of the translational components of the OF. In this scenario, the model make the counterintuitive predictions that (i) the perceived tilt direction stays the same when the image is fixed in a head-centric coordinate system but moving in the earth-centered reference frame (first scenario) and (ii) the perceptual solution is biased by the ego-motion when the stimulus is rotating about a *stationary* axis of rotation (second scenario). Formally, this can be seen again by looking at the direction of the angle θ which is computed as follows $\arctan\left[\frac{a_6}{0}\right]$; wheras in the second case θ is given by the formula $\arctan\left[\frac{a_6}{a_1}\right]$ (if a_1 is sufficiently strong relative to a_6, that is $a_1 \gg a_6$, a horizontal perceived solution becomes more and more likely).

A completely different picture is found when we manipulate the rotations of the head. Again, for simplicity, we limit ourselves to the case of horizontal rotations of the head (which means that ω_{R_x} and a_4 are both equal to 0) and we will assume the stimulus is an OF of a pure vertical expansion (or contraction) gradient. Our assumption (following some preliminary findigs from Jeakl et al 2004 and our *Chapter 4*), is that angular motion of the head is sensed with high precision and is also employed for the perceptual interpretation of the OF. In terms of probabilistic modeling this means that the distributions that are representing the head rotation are sharply peaked the true values of $\omega_R = \widehat{\omega_R}$. Now, let's go back to the two previous scenarios. In the first scenario, when the stimulus object is tethered on the line of sight of the subject, and therefore there are no translational components in the OF, the angle θ specify a horizontal tilt direction provided that the numerator of the equation 11 is sufficiently larger than

then denominator ($\omega_{R_y} \gg a_6$). In the second case instead the object is rotating about a stationary axis of rotation and therefore we can easily see that a_1 and ω_{R_y} cancel each other out. Therefore, this time contrary to what happens during linear head translation, no effect of the translatory component is expected.

In summary, we propose a new approach to model the perceptual strategies adopted by active and passive observers to derive the surface's tilt orientation. Our model is based on two main ideas: (i) the analysis of the OF is conducted on a head-centric frame of reference and (ii) the computation of the OF is achieved by taking into account the sensed rotations of the head and disregarding the linear translations. Within the context of this approach the tilt direction is derived by computing the direction of self motion, θ, by a joint analysis of the translational and the gradients components of the OF together with the sensed rotations of the head. This process doesn't necessarily guarantee a veridical interpretation of the OF.

References

H Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, 1974. 28, 46, 72

JA Beintema and AV van den Berg. Heading detection using motion templates and eye velocity gain fields. *Vision Research*, 38:2155–2179, 1998. 33, 78

AJ Benson. *The vestibular sensory system. In: Barlow HB, Mollon JD (ed) The senses.* Cambridge University Press, Cambridge, 1982. 49, 52

FJ Calabro, S Soto-Faraco, and LM Vaina. Acoustic facilitation of object movement detection during self-motion. *Proceedings of the Royal Society-Biological Science*, 22:2840–2847, 2011. 52

C Caudek, C Fantoni, and F Domini. Bayesian modeling of perceived surface slant from actively-generated and passively-observed optic flow. *PLOS One*, 6(4), 2011. 4, 6, 10, 12, 18, 31, 32, 39, 49, 52, 54, 76, 79, 86, 87, 91

A Chen, GC DeAngelis, and DE Angelaki. Macaque parieto-insular vestibular cortex: responses to self-motion and optic flow. *The Journal of Neuroscience*, 30:3022–3042, 2010. 53, 64

F Colas, J Droulez, M Wexler, and P Bessire. A unified probabilistic model of the perception of three-dimensional structure from optic flow. *Biological Cybernetics*, 97:461–477, 2007. 12, 30, 87

V Cornilleau-Peres and J Droulez. The visual perception of three-dimensional shape from self-motion and object motion. *Vision Research*, 34:2331–2336, 1994. 33

V Cornilleau-Peres, M Wexler, J Droulez, E Marin, C Miege, and B Bourdoncle. Visual perception of planar orientation: Dominance of static depth cues over motion cues. *Vision Research*, 42:1403–1412, 2002. 12, 30, 31

G DeAngelis and D Angelaki. Visualvestibular integration for self-motion perception. In *The Neural Bases of Multisensory Processes*. Boca Raton (FL): CRC Press, 2012. 3, 12, 52

TM Dijkstra, V Cornilleau-Peres, CC Gielen, and J Droulez. Perception of three-dimensional shape from ego- and object-motion: Comparison between small- and large-field stimuli. *Vision Research*, 35:453–462, 1995. 31

F Domini and C Caudek. Perceiving surface slant from deformation of optic flow. *Trends in Cognitive Sciences*, 25:426–444, 1999. 5, 75, 85, 86

F Domini and C Caudek. 3-d structure perceived from dynamic information: a new theory. *Trends in Cognitive Sciences*, 7:444–449, 2003a. 4, 5, 12, 31, 32, 33, 49, 75, 76, 86

F Domini and C Caudek. Recovering slant and angular velocity from a linear velocity field: Modeling and psychophysics. *Vision Research*, 43:1753–1764, 2003b. 33, 75

F Domini, C Fantoni, C Caudek, and G Mancuso. The generic linear motion assumption for the interpretation of the optic flow. *Journal of Vision*, 12(9): 1199, 2012. 36, 49, 74, 79

C Fantoni, C Caudek, and F Domini. Systematic distortions of perceived planar surface motion in active vision. *Journal of Vision*, 10:1–20, 2010. 4, 5, 6, 10, 12, 18, 31, 32, 39, 49, 52, 54, 76, 79, 86, 91

C Fantoni, C Caudek, and F Domini. Perceived surface slant is systematically biased in the actively-generated optic flow. *PLoS ONE*, 7(3):e33911, 2012. 4, 5, 12, 18, 31, 32, 39, 49, 52, 54, 76, 79, 86, 91

J Gibson. In *The Perception of the Visual World*. Riverside Press, Cambridge, England, 1950. 12, 38

J Goossens, SP Dukelow, RS Menon, T Vilis, and AV van den Berg. Representation of head-centric flow in the human motion complex. *The journal of Neuroscience*, 26:5616–5627, 2006. 33

CM Hurvich and CL Tsai. Regression and time series model selection in small samples. *Biometrika*, 76:297–307, 1989. 28, 46, 72

PM Jaekl, MR Jenkin, and LR Harris. Perceiving a stable world during active rotational and translational head movements. *Experimental Brain Research*, 163(3):388–399, 2004. 38, 49, 53, 57, 63, 64, 79, 91

ER Kandel, JH Schwartz, and TM Jessell. In *Principles of neural science*. London: Prentice-Hall, 2000. 52, 53, 63

FAA Kingdom and N Prins. In *Psychophysics: A Practical Introduction*. Academic Press: an imprint of Elsevier, London, 2010. 28, 46, 72

JJ Koenderink. Optic flow. *Vision Research*, 26:161–180, 1986. xi, 69, 84

JJ Koenderink and AJ van Doorn. Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta*, 22:773–791, 1975. 12, 38, 83

JJ Koenderink and AJ van Doorn. How an ambulant observer can construct a model of the environment from the geometrical structure of the visual inflow. *Kybernetik*, 77:224–247, 1978. 12, 38

O Kolev, T Mergner, H Kimmig, and W Becker. Detection thresholds for object motion and self-motion during vestibular and visuo-motor stimulation. *Brain Research Bulletin*, 40:467–470, 1996. 49, 52

M Lappe, F Bremmer, and AV van den Berg. Perception of self motion from visual flow. *Trends in Cognitive Sciences*, 3:329–336, 1999. 52

HC Longuet-Higgins. The visual ambiguity of a moving plane. *Proceedings of the Royal Society of London*, 223:165–175, 1984. 12, 38, 82

HC Longuet-Higgins and K Prazdny. The interpretation of a moving retinal image. *Proceedings of the Royal Society of London*, 208:385–397, 1980. 12, 38, 82

PR MacNeilage, MS Banks, GC DeAngelis, and DE Angelaki. Vestibular heading discrimination and sensitivity to linear acceleration in head and world coordinates. *The Journal of Neuroscience*, 30:9084–9094, 2010. 49, 53, 63

M Morrone and R Owens. Feature detection from local energy. *Pattern Recognition Letters*, 6:303–313, 1987. 86

JE Raymond. Directional anisotropy of motion sensitivity across the visual field. *Vision Research*, 34:1029–1037, 1994. 53, 71, 75

FP Redlick, LR Harris, and MR Jenkin. Humans can use optic flow to estimate distance of travel. *Vision Research*, 41:213–219, 2001. 52

CS Royden and EM Connors. The detection of moving objects by moving observers. *Vision Research*, 50:1014–1024, 2010. 52

CS Royden, MS Banks, and JA Crowell. The perception of heading during eye movements. *Nature*, 360:583–585, 1992. 33, 79

SK Rushton and PA Warren. Moving observers, relative retinal motion and the detection of object movement. *Current Biology*, 26:542–543, 2005. 52

WA Simpson. Optic flow and depth perception. *Spatial Vision*, 37:35–75, 1993. 2

N Sugiura. Further analysis of the data by akaikes information criterion and the finite corrections. *Communications in statistics:Theory and Methods*, 7:13–26, 1978. 28, 46, 72

JT Todd and P Bressan. The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception and Psychophysics*, 48:419–430, 1990. 86

S Ullman. In *The interpretation of visual motion*. The MIT Press, 1979. 12, 30, 38, 82

LM Vaina, SA Beardsley, and S Rushton. In *Optic flow and beyond*. Kluwer Academic, New York, 2004. 52

JJ van Boxtel, M Wexler, and J Droulez. Perception of plane orientation from self-generated and passively observed optic flow. *Journal of Vision*, 3(5):318–332, 2003. 4, 6, 30, 31

AV van den Berg. Robustness of perception of heading from optic flow. *Vision Research*, 32:1285–1296, 1992. 33, 79

H Wallach. Perceiving a stable environment when one moves. *Annual Review of Psychology*, 38:1–27, 1987. 52

H Wallach and DN O'Connell. The kinetic depth effect. *Journal of Experimental Psychology*, 45:205–217, 1953. 2

PA Warren and SK Rushton. Optic flow processing for the assessment of object movement during ego movement. *Current Biology*, 19:1555–1560, 2009. 52

W Warren. Optic flow. In *The Visual Neurosciences*. The MIT Press, 2003. 1

M Wexler. Voluntary head movement and allocentric perception of space. *Psychological Science*, 14:340–346, 2003. 4, 12, 30, 63, 76, 79

M Wexler. Anticipating the three-dimensional consequences of eye movements. *Proceedings of the National Academy of Sciences of the United States of America*, 102:1246–1251, 2005. 12, 30

M Wexler and JJ van Boxtel. Depth perception by the active observer. *Trends in Cognitive Science*, 9:431–438, 2005. 4, 76, 79

M Wexler, I Lamouret, and J Droulez. The stationarity hypothesis: An allocentric criterion in visual perception. *Vision Research*, 41:3023–3037, 2001a. v, 4, 5, 6, 10, 13, 14, 24, 30, 32, 76, 79

M Wexler, F Panerai, I Lamouret, and J Droulez. Self-motion and the perception of stationary objects. *Nature*, 409:85–88, 2001b. 4, 5, 6, 30, 76, 79

FA Wichmann and NJ Hill. The psychometric function: I. fitting, sampling and goodness-of-fit. *Perception and Psychophysics*, 63(8):1293–1313, 2001a. 57, 58

FA Wichmann and NJ Hill. The psychometric function: Ii. bootstrap-based confidence intervals and sampling. *Perception and Psychophysics*, 63(8):1314–1329, 2001b. 57, 58

A Yuille and D Kersten. Vision as a bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, 10:301–308, 2006. 3

H Zhong, V Cornilleau-Peres, LF Cheong, GM Yeow, and J Droulez. The visual perception of plane tilt from motion in small field and large field: Psychophysics and theory. *Vision Research*, 46:3494–3513, 2006. 82