Chapter 12

Visual motion processing and human tracking behavior

BibTex entry

```
@inbook{Montagnini15bicv,
    author = {Montagnini, Anna and Perrinet, Laurent U. and Masson, Guillaume S.},
    chapter = {12},
    citeulike-article-id = {13566752},
    editor = {Crist{\'{o}}bal, Gabriel and Perrinet, Laurent U. and Keil, Matthias S.},
    month = nov,
    publisher = {Wiley-VCH Verlag GmbH {\&} Co. KGaA},
    booktitle = {Biologically Inspired Computer Vision},
    keywords = {bicv-motion},
    posted-at = {2015-03-31 14:20:42},
    priority = {0},
    DOI = {10.1002/9783527680863.ch12},
    url = {http://onlinelibrary.wiley.com/doi/10.1002/9783527680863.ch12/summary},
    title = {Visual motion processing and human tracking behavior},
    year = {2015}
}
```

Contents

12 Visual motion processing and human tracking behavior **1**
12.1 Introduction 3
12.2 Pursuit initiation: facing uncertainties 5
12.2.1 Where is the noise? Motion tracking precision and accuracy 6
12.2.2 Where is the target really going? 6
12.2.3 Human smooth pursuit as dynamic readout of the neural solution to the aperture problem 7
12.3 Predicting future and on-going target motion 8
12.3.1 Anticipatory smooth tracking 8
12.3.2 If you don't see it, you can still predict (and track) it . . 10
12.4 Dynamic integration of retinal and extra-retinal motion information: computational models 13
12.4.1 A Bayesian approach for open-loop motion tracking . . . 13
12.4.2 Bayesian (or Kalman-filtering) approach for smooth pursuit: Hierarchical models 13
12.4.3 A Bayesian approach for smooth pursuit: Dealing with delays 17
12.5 Reacting, inferring, predicting: a neural workspace 19
12.6 Conclusion 23
12.6.1 Interest for computer vision 24

12.1 Introduction

We are visual animals and vision is the primary source of information about the 3D layout of our environment or the occurrence of events around us. In nature, visual motion is abundant and generated by a large set of sources, such as the movement of another animal, be it predator or prey, or our own movements. Primates possess an highly performant system for motion processing. Such system is tightly linked to ocular tracking behaviors: a combination of smooth pursuit and saccadic eye movements is engaged to stabilize the retinal image of a selected moving object within the high-acuity region of our retinas, the fovea. During smooth phases of tracking, eye velocity nearly perfectly matches the target velocity, minimizing retinal slip. Fast saccadic eye movements interrupt these slow phases to correct position errors between the retinal location of the target image and the fovea. Thus, smooth pursuit and saccadic eye movements are respectively largely controlled by velocity and position information (Rashbass, 1961). Smooth pursuit eye movements (which are the focus of this chapter)

are traditionally considered as the product of a *reflexive* sensorimotor circuitry. Two important observations support this assumption: first, it is not possible to generate smooth pursuit at will across a stationary scene, and, second, it is not possible to fully suppress pursuit in a scene consisting solely of moving targets (Kowler, 2011; Kowler et al., 2014).

In non-human and human primates, visual motion information is extracted through a cascade of cortical areas spanning the dorsal stream of the cortical visual pathways (for reviews, see (Maunsell and Newsome, 1987)). Direction and speed selective cells are found in abundance in two key areas lying at the junction between occipital and parietal lobes. Middle temporal (MT) and Medio-Superior Temporal (MST) areas can decode local motion information of a single object at multiple scales, isolate it from its visual background and reconstruct its trajectory (Bradley and Goyal, 2008; Born and Bradley, 2005). Interestingly, neuronal activities in areas MT and MST have been firmly related to the initiation and maintenance of tracking eye movements (Newsome and Wurtz, 1988). Dozens of experimental studies have shown that local direction and speed information for pursuit are encoded by MT populations (Lisberger and Movshon, 1999) and transmitted to MST populations where non retinal information are integrated to form an internal model of object motion (Newsome et al., 1988). Such model is then forwarded to both frontal areas involved in pursuit control, such as frontal eye fields (FEF) and supplementary eye fields (SEF), as well as to the brainstem and cerebellar oculomotor system (see (Krauzlis, 2004) for a review).

What can we learn about visual motion processing by investigating smooth pursuit responses to moving objects? Since the pioneering work of Hassenstein and Reichardt (Hassenstein and Reichardt, 1956), behavioural studies of tracking behavioural responses have been highly influential upon theoretical approaches to motion detection mechanisms (see Chapter `017_tim_tiedemann`, this book), as illustrated by the fly vision literature (see (Borst, 2014; Borst et al., 2010) for recent reviews) and its application in bio-inspired vision hardwares (e.g. (Harrison and Koch, 2000; Kohler et al., 2009)). Because of their strong *stimulus-driven* nature, tracking eye movements have been used similarly in primates to probe the properties of fundamental aspects of motion processing, from detection to pattern motion integration (see a series of recent reviews in (Masson, 2004; Kowler, 2011; Lisberger, 2010)) and the coupling between the active pursuit of motion and motion perception (Spering and Montagnini, 2011). One particular interest of tracking responses is that they are time-continuous, smooth, measurable movements that reflect the temporal dynamics of sensory processing in the changes of eye velocity (Masson, 2004; Masson and Perrinet, 2012). Here, we would like to focus on a particular aspect of motion processing in the context of sensorimotor transformation: uncertainty. Uncertainty arises from both random processes, such as the noise reflecting unpredictable fluctuations on the velocity signal, as well as from non stochastic processes such as ambiguity when reconstructing global motion from local information. We will show that both sensory noise and ambiguities impact the sensorimotor transformation as seen from the variability of eye movements and their course towards a steady, optimal solution.

Understanding the effects of various sources of noise and ambiguities can change our views on the two faces of the sensorimotor transformation. First, we can better understand how visual motion information is encoded in neu-

ral populations and the relationships between these population activities and behaviors (Lisberger, 2010; Osborne, 2011). Second, it may change our view about how the brain controls eye movements and open the door to new theoretical approaches based on inference rather than linear control systems. Albeit still within the framework of stimulus-driven motor control, such new point of view can help us elucidate how higher level, cognitive processes, such as prediction, can dynamically interact with sensory processing to produce a versatile, adaptive behavior by which we can catch a flying ball despite its complex and even sometimes partially occluded trajectory.

This chapter is divided in four parts. First, we will examine how noise and ambiguity both affect the pursuit initiation. In particular, we will focus on the temporal dynamics of uncertainty processing and the estimate of the optimal solution for motion tracking. Second, we will summarize how non-sensory, predictive signals can help maintaining a good performance when sensory evidences become highly unreliable and, on the contrary, when the future sensory inputs become highly predictable. Herein, we will illustrate these aspects with behavioral results gathered in both human and non-human primates. Third, we will show that these different results on visually guided and predictive smooth pursuit dynamics can be reconciled within a single Bayesian framework. Last, we will propose a biologically-plausible architecture implementing a hierarchical inference network for a closed-loop, visuomotor control of tracking eye movements.

12.2 Pursuit initiation: facing uncertainties

The ultimate goal of pursuit eye movements is to reduce the retinal slip of image motion down to nearly zero such that fine details of the moving pattern can be analyzed by spatial vision mechanisms. Overall, the pursuit system acts as a negative feedback loop where the eye velocity matches target velocity (such that the *pursuit gain* is close to 1) to cancel the image motion on the retina. However, because of the delays due to both sensory and motor processing, the initial rising phase of eye velocity, known as pursuit initiation, is *open-loop*. This means that during this short period of time (less than about 100 ms) no information about eye movements is available to the system and the eye velocity depends only on the properties of the target motion presented to the subject. This short temporal window is ideal to probe how visual motion information is processed and transformed into an eye movement command (Lisberger et al., 1987; Lisberger, 2010). It becomes possible to map the different phases of visual motion processing to the changes in initial eye velocity and therefore to dissect out the contribution of spatial and temporal mechanisms of direction and speed decoding (Masson and Perrinet, 2012). However, this picture becomes more complicated as soon as one considers more naturalistic and noisy conditions for motion tracking, whereby, for instance, the motion of a complex-shaped extended object has to be estimated, or when several objects move in different directions. We will not focus, here, on the last problem, which involves the complex and time-demanding computational task of object segmentation (Masson, 2004; Schütz et al., 2010). In the next subsections, we focus instead on the nature of the noise affecting pursuit initiation and the uncertainty related to limitations in processing spatially-localized motion information.

12.2.1 Where is the noise? Motion tracking precision and accuracy

Human motion tracking is variable across repeated trials and the possibility to use tracking behavior (at least during the initiation phase) to characterize visual motion processing across time and across all kinds of physical properties of the moving stimuli relies strongly on the assumption that oculomotor noise does not override the details of visual motion processing. In order to characterize the variability of macaques' pursuit responses to visual motion signals, Osborne et al. (2005) analyzed monkeys' pursuit eye movements during the initiation phase — here, the first 125 ms after pursuit onset. On the basis of a principal component analysis of pursuit covariance matrix, they concluded that pursuit variability was mostly due to sensory fluctuations in estimating target motion parameters such as onset time, direction and speed, accounting for around 92% of the pursuit variability. In a follow-up study, they estimated the time course of the pursuit system's sensitivity to small changes in target direction, speed and onset time (Osborne et al., 2007). This analysis was based on pursuit variability during the first 300 ms after target motion onset. Discrimination thresholds (inverse of sensitivity) decreased rapidly during open-loop pursuit and, in the case of motion direction, followed a similar time course to the one obtained from the analysis of neuronal activity in area MT (Osborne et al., 2004).

Noise inherent to the kinematic parameters of the moving target is not the only source of uncertainty for visual motion processing. A very well-known and puzzling finding in visual motion psychophysics is that the speed of low-contrast moving stimuli is most often underestimated as compared to high contrast stimuli moving with exactly the same motion properties (see chapter `009_series`). In parallel with the perceptual misjudgment, previous studies have shown that tracking quality is systematically degraded (with longer onset latencies, lower acceleration at initiation and lower pursuit gain) with low contrast stimuli (Spering et al., 2005). This reduction of motion estimation and tracking accuracy when decreasing the luminance-contrast of the moving stimulus has been interpreted as evidence in favor of the fact that, when visual information is corrupted, the motion tracking system relies more on internal models of motion, or motion *Priors* (Weiss et al., 2002).

12.2.2 Where is the target really going?

The external world is largely responsible for variability. Another major source of uncertainty when considering sensory processing is ambiguity: a single retinal image can correspond to many different physical arrangements of the objects in the environment. In the motion domain, a well-known example of such input ambiguity is called "the aperture problem". When seen through a small aperture, the motion of an elongated edge (i.e. a one-dimensional -1D- change in luminance, see Figure 12.1a, middle panel) is highly ambiguous. The same local motion signal can be generated by a infinite number of physical translations of the edge. Hans Wallach (see (Wuerger et al., 1996) for an English translation of the original publication in German) was the first psychologist to recognize this problem and to propose a solution for it. A spatial integration of motion information provided by edges with different orientations can be used

to recover the true velocity of the pattern. Moreover, two-dimensional (2D) features such as corners or line-endings (whereby luminance variations occur along two dimensions, see Figure 12.1a, right panel, for an example) can also be extracted through the same, small aperture as their motion is no longer ambiguous. Again, 1D and 2D motion signals can be integrated to reconstruct the two-dimensional velocity vector of the moving pattern. After several decades of intensive research at both physiological and behavioral levels, it remains largely unclear what computational rules are used by the brain to solve the 2D motion integration problem (see (Masson and Ilg, 2010) for a collection of review articles).

Indeed, several computational rules for motion integration have been proposed over the last 40 years (see (Bradley and Goyal, 2008) for a review). In a limited number of cases, a simple vector averaging of the velocity vectors corresponding to the different 1D edge motions can be sufficient. A more generic solution, the Intersection-of-Constraints (IOC) is a geometrical solution that can always recover the exact global velocity vector from at least two moving edges with different orientations (Fennema and Thompson, 1979; Adelson and Movshon, 1982). However, the fact that perceived direction does not always correspond to the IOC solution (for instance, for very short stimulus duration (Yo and Wilson, 1992) or when a single 1D motion signal is present (Lorenceau et al., 1993; Gorea and Lorenceau, 1991)) has supported the role of local 2D features in motion integration.

Several feedforward computational models have been proposed to implement these different rules (Wilson et al., 1992; Löffler and Orbach, 1999). All these feedforward models have the same architecture. Motion integration is seen as a two-stage computation. The first stage, corresponding to cortical area V1 in primates, extracts local motion informations through a set of oriented spatiotemporal filters. This correspond to the fact that most V1 neurons respond to the direction orthogonal to the orientation of an edge drifting across their receptive field (Albright, 1984). The local motion analyzers feed a second, integrative stage where pattern motion direction is computed. This integrative stage is thought to correspond to the extra-striate cortical area MT in primates. MT neurons have large receptive fields, they are strongly direction selective and a large fraction of them can unambiguously signal the pattern motion direction, regardless of the orientation of its 1D components (Movshon et al., 1985; Albright, 1984). Different nonlinear combinations of local 1D motion signals can be used to extract either local 2D motion cues or global 2D motion velocity vectors. Another solution proposed by Perrinet and Masson (2012) is to consider that local motion analyzers are modulated by motion coherency (Burgi et al., 2000). This theoretical model shows the emergence of similar 2D motion detectors. These two-stage frameworks can be integrated into more complex models where local motion information is diffused across some retinotopic maps.

12.2.3 Human smooth pursuit as dynamic readout of the neural solution to the aperture problem

Behavioral measures do not allow to capture the detailed temporal dynamics of the neuronal activity underlying motion estimate. However, smooth pursuit recordings do still carry the signature of the dynamic transition between the initial motion estimate dominated by the vector average of local 1D cues and

the later estimate of global object motion. In other terms, human tracking data provides a continuous (delayed and low-pass filtered) dynamic readout of the neuronal solution to the aperture problem. Experiments in our and other groups (Masson and Stone, 2002; Wallace et al., 2005; Born et al., 2006) have consistently demonstrated, in both humans and monkeys, that tracking is transiently biased at initiation toward the direction orthogonal to the moving edge (or the vector average if multiple moving edges are present), when such direction does not coincide with the global motion direction. After some time (typically $200 - 300$ ms) such bias is extinguished and the tracking direction converges to the object's global motion. In the example illustrated in Figure 12.1, a tilted bar translates horizontally, thereby carrying locally ambiguous edge-related information (middle panel of part a). A transient non-zero vertical smooth pursuit velocity (lower right panel of Figure 12.1b) reflects the initial aperture-induced bias, differently from the case where local and global motion are coherent (as for the pursuit of a horizontally moving vertical bar, see Figure 12.1b, leftmost panels).

The size of the transient directional tracking bias and the time needed for converging to the global motion solution depend on several properties of the visual moving stimulus (Wallace et al., 2005) including stimulus luminance contrast (Montagnini et al., 2007; Bogadhi et al., 2011). In section 12.4.1 we will see that this tracking dynamics is consistent with a simple Bayesian recurrent model (or equivalently a Kalman filter (Montagnini et al., 2007), which takes into account the uncertainty associated to the visual moving stimulus and combines it with prior knowledge about visual motion (see also chapter `009_series`).

12.3 Predicting future and on-going target motion

Smooth pursuit eye movements can also rely on prediction of target movement to accurately follow the target despite a possible major disruption of the sensory evidence. Prediction allows also to compensate for processing delays, an unavoidable problem of sensory-to-motor transformations.

12.3.1 Anticipatory smooth tracking

The exploitation of statistical regularities in the sensory world and/or of cognitive information ahead of a sensory event is a common trait of adaptive and efficient cognitive systems that can, on the basis of such predictive information, anticipate choices and actions. It is a well-established fact that humans cannot generate smooth eye movements at will: for instance it is impossible to smoothly track an imaginary target with the eyes, except in the special condition in which an unseen target is self-moved through the smooth displacement of one's own finger (Gauthier and Hofferer, 1976). In addition, smooth pursuit eye movements do necessarily lag unpredictable visual target motion by a (short) time delay. In spite of this, many years ago, already, it was known that, when tracking regular periodic motion, pursuit sensorimotor delay can be nulled and a perfect synchronicity between target and eye motion is possible (see (Kowler et al., 2014; Barnes, 2008) for detailed reviews). Furthermore, when the direction of motion of a moving target is known in advance (for instance because motion properties

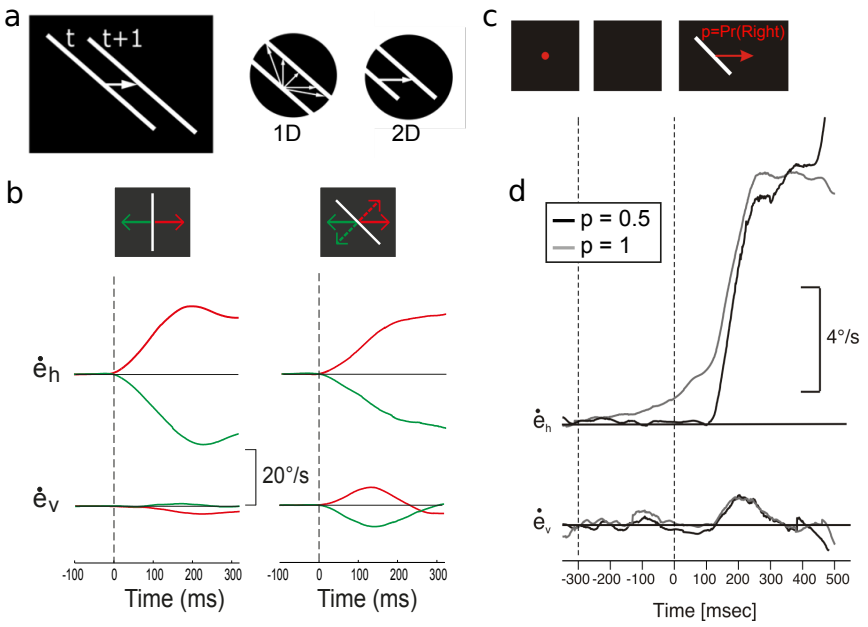


Figure 12.1: Smooth pursuit's account for the dynamic solution of motion ambiguity and motion prediction: (a) A tilted bar translating horizontally in time (left panel) carries both ambiguous 1D motion cues (middle panel), and non-ambiguous 2D motion cues (rightmost panel). (b) Example of average horizontal ($\dot{e_h}$) and vertical ($\dot{e_v}$) smooth pursuit eye velocity while tracking a vertical (left) or a tilted bar (right) translating horizontally, either to the right (red curves) or to the left (green curves). Velocity curves are aligned on smooth pursuit onset. (c) Schematic description of a trial in the experiment on anticipatory smooth pursuit: after a fixation display, a fixed duration blank precedes the motion onset of a tilted line moving rightwards (with probability p) or leftward (with probability $1 - p$). (d) Example of average horizontal ($\dot{e_h}$) and vertical ($\dot{e_v}$) smooth pursuit eye velocity in the anticipation experiment for two predictability conditions, $p = 0.5$ (unpredictable, black curves) and $p = 1$ (completely predictable, grey curves).

are the same across many repeated experimental trials), anticipatory smooth eye movements are observed in advance of the target motion onset (Montagnini et al., 2006), as illustrated in the upper panel of Figure 12.1d. Interestingly, also relatively complicated motion patterns, such as piecewise linear trajectories (Barnes and Schmid, 2002), or accelerated motion (Bennett et al., 2007) can be anticipated. Finally, probabilistic knowledge about target motion direction or speed (Montagnini et al., 2010), and even subjectively experienced regularities extracted from the previous few trials (Kowler et al., 1984) can modulate in a systematic way anticipatory smooth pursuit. Recently, several researchers have tested the role of higher-level cognitive cues for anticipatory smooth pursuit, leading to a rather diverse set of results. Although verbal or pictorial abstract cues indicating the direction of the upcoming motion seem to have a rather weak (though non inexistent) influence on anticipatory smooth pursuit, other cues are more easily and immediately interpreted and used for motion anticipation (Kowler et al., 2014). For instance, a barrier blocking one of two branches in a line-drawing illustrating an inverted-y-shaped tube, where the visual target was about to move (Kowler, 1989), leads to robust anticipatory smooth tracking in the direction of the other, unblocked branch.

12.3.2 If you don't see it, you can still predict (and track) it

While walking on a busy street downtown, it can occur that we track a moving car with our gaze and, despite it being hidden behind a truck occasionally driving in front of it, we can still closely follow its motion and have our gaze next to the car's position at its reappearance. In the lab, researchers have shown (Becker and Fuchs, 1985) that during the transient disappearance of a moving target human subjects are capable to keep tracking the hidden motion with their gaze, although with a lower gain (see Figure 12.2, upper panels). During blanking, indeed, after an initial drop, eye velocity can be steadily maintained, typically at about 70% of pre-blanking target velocity, although higher eye speed can be achieved with training (Madelain and Krauzlis, 2003). In addition, when the blank duration is fixed, an anticipatory re-acceleration of the gaze rotation is observed ahead of target reappearance (Bennett and Barnes, 2003). Extra-retinal, predictive information is clearly called into play to drive ocular tracking in the absence of a visual target. The true nature of the drive for such predictive eye velocity is still debated (see (Kowler, 2011) for a review). Previous studies have proposed that it could either be a copy of the oculomotor command (an efference copy) serving as a positive feedback (Churchland et al., 2003; Madelain and Krauzlis, 2003) or a sample of visual motion being held in working memory (Bennett and Barnes, 2003). In all cases a rather implausible "switch-like" mechanism was assumed, in order to account for the change of regime between the visual- and prediction-driven tracking.

While the phenomenology of human smooth pursuit during the transient absence of a visual target is well investigated (see for example (Bennett and Barnes, 2003; Orban de Xivry et al., 2006, 2008)), less is known about its functional characterization, and about how the extra-retinal signals implicated in motion tracking without visual input interact with retinal signals across time. In particular, as motion estimation for tracking is affected by sensory noise and computational limitations (see sections 12.2.1 and 12.2.2), do we rely on extra-

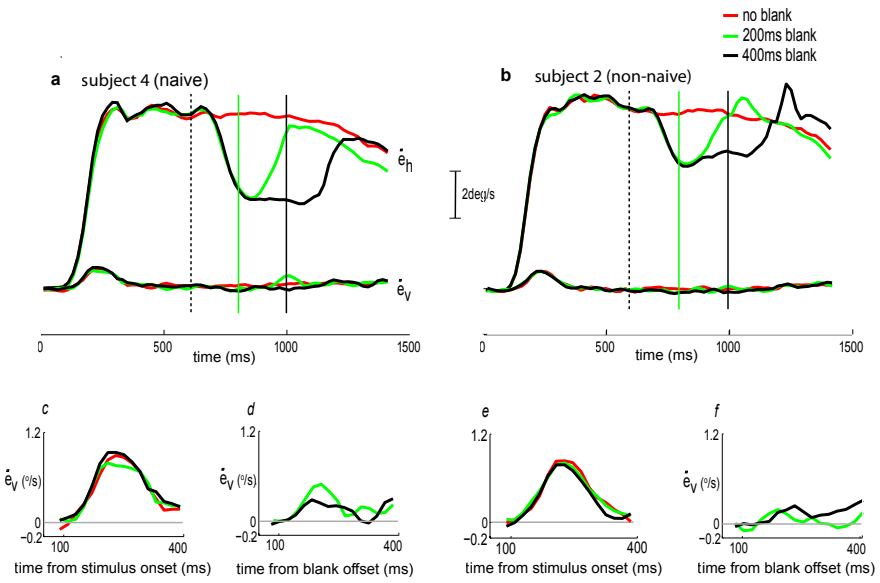


Figure 12.2: Examples of human smooth pursuit traces (one different participant on each column, a naive one on the left and a non-naive one on the right side) during horizontal motion of a tilted bar which is transiently blanked during steady-state pursuit. (a) and (b): Average horizontal ($\dot{e_h}$) and vertical ($\dot{e_v}$) eye velocity. Different blanking conditions are depicted with different colors, as from figure legend. The vertical dashed line indicated the blank onset; vertical full colored lines indicate the end of the blanking epoch for each blanking duration represented. (c) and (e): Zoom on the aperture-induced bias of vertical eye velocity at target motion onset, for all blanking conditions. (d) and (f): Zoom on the aperture-induced bias of vertical eye velocity at target reappearance after blank (time is shifted so that 0 corresponds to blank offset), for all blanking conditions.

retinal predictive information in a way that depends on sensory uncertainty? In the past two decades, the literature on optimal cue combinations in multisensory integration has provided evidence for a weighted sum of different sources of information (such as visual and auditory (Alais and Burr, 2004), or visual and haptic cues (Ernst and Banks, 2002), whereby each source is weighted according to its reliability (defined as inverse variance). Recently, Khoei et al. (2013) have shown that a predictive term on the smoothness of the trajectory is sufficient to account for the motion extrapolation part of the motion, however it lacked a mechanism to weight retinal and extra retinal information. In another recent study, we have tested the hypothesis that visual and predictive information for motion tracking are weighted according to their reliability and dynamically integrated to provide the observed oculomotor command (Bogadhi et al., 2013).

In order to do so, we have analyzed human smooth pursuit during the tracking of a horizontally-moving tilted bar that could be transiently hidden at different moments (blanking paradigm), either early, during pursuit initiation, or late, during steady-state tracking. By comparing the early and late blanking conditions we found two interesting results: first, the perturbation of pursuit velocity caused by the disappearance of the target was more dramatic for the late than the early blanking, both in terms of relative velocity drop and presence of an anticipatory acceleration before target reappearance. Second, a small, but significant, aperture-induced tracking bias (as described in section 12.2.3) was observed at target reappearance after late but not early blanking. Interestingly, these two measures (the size of the tracking velocity reduction after blank onset and the size of the aperture bias after target disappearance) turned out to be significantly correlated across subjects for the late blanking conditions.

We interpreted the ensemble of these results as an evidence in favor of dynamic optimal integration of visual and predictive information: at pursuit initiation, sensory variability is strong and predictive cues related to target- or gaze-motion dominate, leading to a relative reduction of both the effects of target blanking and of the aperture-bias. On the contrary, later on, sensory information becomes more reliable and the sudden disappearance of a visible moving target leads to a more dramatic disruption of motion tracking; coherently with this, the (re)estimation of motion is more strongly affected by the inherent ambiguity of the stimulus (a tilted bar). Finally, the observed correlation of these two quantities across different human observers indicates that the same computational mechanism (i.e. optimal dynamic integration of visual and predictive information) is scaled, at the individual level, in such a way that some people rely more strongly than others on predictive cues rather than on intrinsically noisy sensory evidence. Incidentally, in our sample of human volunteers, the expert subjects seemed to rely more on predictive information than the naive ones.

In section 12.4.2, we will illustrate a model which is based on hierarchical Bayesian inference and is capable to qualitatively capture the human behavior in our blanking paradigm. A second important question is whether and how predictive information is affected by uncertainty as well. We will start to address this question in the next section.

12.4 Dynamic integration of retinal and extra-retinal motion information: computational models

12.4.1 A Bayesian approach for open-loop motion tracking

Visual image noise and motion ambiguity, the two sources of uncertainty for motion estimate described in sections 12.2.1 and 12.2.2 can be well integrated within a Bayesian (Weiss et al., 2002; Stocker and Simoncelli, 2006; Montagnini et al., 2007; Perrinet and Masson, 2007) or, equivalently, a Kalman-filtering framework (Dimova and Denham, 2009; Orban de Xivry et al., 2013), whereby estimated motion is the solution of a dynamical statistical inference problem (Kalman, 1960). In these models, the information from different visual cues (such as local $1D$ and $2D$ motions) can be represented, as probability distributions, by their likelihood functions. Bayesian models also allow the inclusion of prior constraints related to experience, expectancy bias and all possible sources of extra-sensory information. On the ground of a statistical predominance of static or slowly moving objects in nature, the most common assumption used in models of motion perception is a preference for slow speeds, typically referred to as low-speed prior (represented in panel *a* of Figure 12.3). The effects of priors are especially salient when signal uncertainty is high (see chapter `009_series`).

The sensory likelihood functions can be derived for simple objects with the help of a few reasonable assumptions. For instance the motion cue associated with a non-ambiguous 2D feature would be approximated by a gaussian Likelihood centered on the true stimulus velocity and with a variance proportional to visual noise (e.g. inversely related to its visibility, see panel *c* in Figure 12.3). On the other hand, edge-related ambiguous information would be represented by an elongated velocity distribution parallel to the orientation of the moving edge, with an infinite variance along the edge direction reflecting the aperture ambiguity (panel *b*). Weiss and colleagues (Weiss et al., 2002) have shown that, by combining a low-speed prior with an elongated velocity likelihood distribution parallel to the orientation of the moving line, it is possible to predict the aperture-induced bias (as illustrated in Figure 12.3d). By introducing the independent contribution of the non-ambiguous 2D likelihood, as well as a recurrent optimal update of the Prior (with a feedback from the Posterior (Montagnini et al., 2007), see Figure 12.3), and cascading this recurrent network with a realistic model of smooth pursuit generation (Bogadhi et al., 2011), we have managed to reproduce the complete dynamics of the solution of the aperture problem for motion tracking, as observed in human smooth pursuit traces.

12.4.2 Bayesian (or Kalman-filtering) approach for smooth pursuit: Hierarchical models

Beyond the inferential processing of visual uncertainties, that mostly affect smooth pursuit initiation, we have seen in section 12.3.2 that predictive cues can efficiently drive motion tracking when the visual information is deteriorated or missing. This flexible control of motion tracking has traditionally been mod-

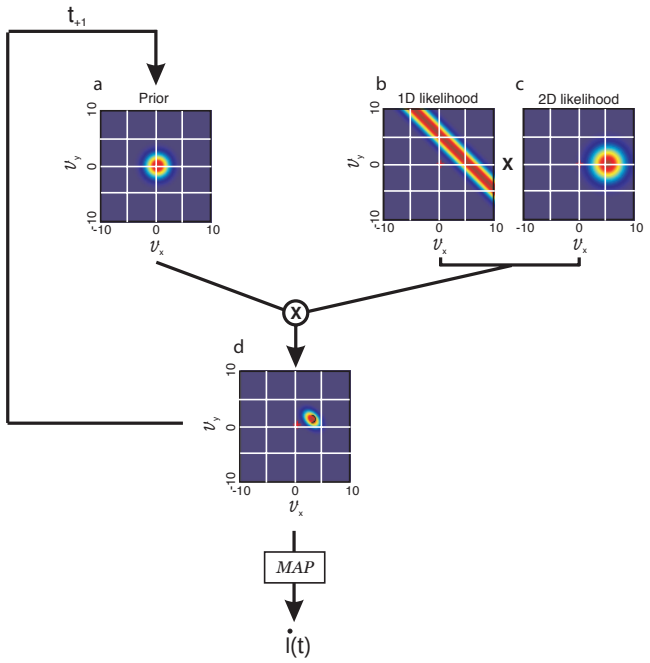


Figure 12.3: A Bayesian recurrent module for the aperture problem and its dynamic solution. Top: (*a*) the Prior and the two independent 1D (*b*) and 2D (*c*) Likelihood functions (for a tilted line moving rightwards at 5 deg/s) in the velocity space are multiplied to obtain the posterior velocity distribution (panel *d*). The inferred image motion is estimated as the velocity corresponding to the posterior maximum (MAP). Probability density functions are colour-coded, such that dark red corresponds to the highest probability and dark blue to the lowest one.

eled (Churchland et al., 2003; Bennett and Barnes, 2003; Madelain and Krauzlis, 2003; Orban de Xivry et al., 2008) in terms of two independent modules, one processing visual motion and the other maintaining an internal memory of target motion. The weak point of these classical models is that they did not provide a biologically plausible mechanism for the interaction or alternation between the two types of control: a somewhat *ad hoc* switch was usually assumed for this purpose.

Inference is very likely to occur at different spatial, temporal and neurofunctional scales. Sources of uncertainty can indeed affect different steps of the sensorimotor process in the brain. Recent work in our group (Bogadhi et al., 2013) and other groups (Orban de Xivry et al., 2013) has attempted to model several aspects of human motion tracking within a single framework, that of Bayesian inference, by postulating the existence of multiple inferential modules organized in a functional hierarchy and interacting according to the rules of optimal cue combination (Kalman, 1960; Fetsch et al., 2012). Here we outline an example of this modeling approach applied to the processing of visual ambiguous motion information under different conditions of target visibility (in the blanking experiment).

In order to explain the data summarized in section 12.3.2 for the transient blanking of a translating tilted bar, we designed a two-modules hierarchical Bayesian recurrent model, illustrated in Figure 12.4. The first module, the retinal recurrent network (panel *a*), implements the dynamic inferential process which is responsible for solving the ambiguity at pursuit initiation (see section 12.2.3) and it only differs from the model scheme in Figure 12.3 for the introduction of processing delays estimated from the literature in monkey electrophysiology (Pack and Born, 2001). The second module (panel *b*), the extra-retinal recurrent network, implements a dynamic short-term memory buffer for motion tracking. Crucially, the respective outputs of the retinal and extra-retinal recurrent modules are optimally combined in the Bayesian sense, so that their mean is weighted with their reliability (inverse variance) before the combination. By cascading the two Bayesian modules with a standard model (Goldreich et al., 1992) for the transformation of the target velocity estimate into eye velocity (panel *c*) and adding some feedback connections (also standard in the models of smooth pursuit to mimic the closed-loop phase), the model is capable, with few free parameters, to simulate motion tracking curves that resemble qualitatively the ones observed for human subjects both during visual pursuit and during blanking. Note that one single crucial free parameter, representing a scaling factor for the rapidity with which the sensory variance increases during target blank, is responsible for the main effects described in section 12.3.2 on the tracking behavior during the blank and immediately after the end of it.

Orban de Xivry et al. (2013) have proposed an integrated Kalman filter model based on two filters, the first one extracting a motion estimate from noisy visual motion input, similar to a slightly simplified version of the previously described Bayesian retinal recurrent module. The second filter (referred to as *predictive pathway*) provides a prediction for the upcoming target velocity on the basis of long term experience (i.e. from previous trials). Importantly, the implementation of a long term memory for a dynamic representation of target motion (always associated to its uncertainty) allows to reproduce the observed phenomenon of anticipatory tracking when target motion properties

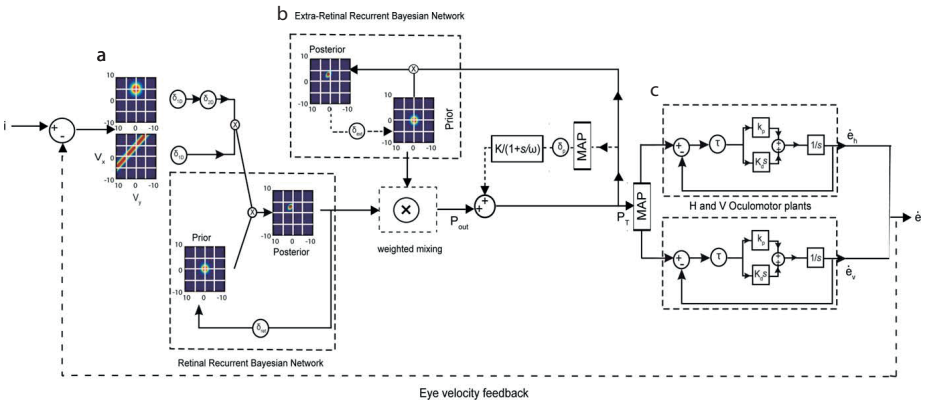


Figure 12.4: Two-stages hierarchical Bayesian model for human smooth pursuit in the blanking experiment. The retinal recurrent loop (a) is the same as in figure 12.3, with the additional inclusion of physiological delays. The posterior from the retinal recurrent loop and prior from the extra-retinal Bayesian network (b) are combined to form the post-sensory output (P_{out}). The maximum a posteriori of the probability (P_T) of target velocity in space serves as an input to both the positive feedback system as well as the oculomotor plants (c). The output of the oculomotor plant is subtracted from the target velocity to form the image's retinal velocity (physical feedback loop shown as broken line). During the transient blank when there is no target on the retina, the physical feedback loop is not functional so that the retinal recurrent block does not decode any motion. The output of the positive feedback system (shown in broken line) is added to the post-sensory output (P_{out}) only when the physical feedback loop is functional. The probability distribution of target velocity in space (P_T) is provided as an input to the extra-retinal recurrent Bayesian network where it is combined with a prior to obtain a posterior which is used to update the prior.

are repeated across trials (see section 12.3.1. However in section 12.6 we will mention some results that challenge the current integrated models of hierarchical inference for motion tracking.

12.4.3 A Bayesian approach for smooth pursuit: Dealing with delays

Recently, we considered optimal motor control and the particular problems caused by the inevitable delay between the emission of motor commands and their sensory consequences (Perrinet et al., 2014). This is a generic problem that we illustrate within the context of oculomotor control where it is particularly insightful (see for instance (Nijhawan, 2008) for a review). Although focusing on oculomotor control, the more general contribution of this work is to treat motor control as a pure inference problem. This allows us to use standard (Bayesian filtering) schemes to resolve the problem of sensorimotor delays — by absorbing them into a generative (or forward) model. A generative model is a set of parameterized equations which describe our knowledge about the dynamics of the environment. Furthermore, this principled and generic solution has some degree of biological plausibility because the resulting active (Bayesian) filtering is formally identical to predictive coding, which has become an established metaphor for neuronal message passing in the brain (see (Bastos et al., 2012) for instance). It uses oculomotor control as a vehicle to illustrate the basic idea using a series of generative models of eye movements — that address increasingly complicated aspects of oculomotor control. In short, we offer a general solution to the problem of sensorimotor delays in motor control — using established models of message passing in the brain — and demonstrate the implications of this solution in the particular setting of oculomotor control.

Specifically, we considered delays in the visuo-oculomotor loop and their implications for active inference. Active inference uses a generalization of Kalman filtering to provide Bayes optimal estimates of hidden states and action (such that our model is a particular Hidden Markov Model) in generalized coordinates of motion. Representing hidden states in generalized coordinates provides a simple way of compensating for both sensory and oculomotor delays. The efficacy of this scheme is illustrated using neuronal simulations of pursuit initiation responses, with and without compensation. We then considered an extension of the generative model to simulate smooth pursuit eye movements — in which the visuo-oculomotor system believes both the target and its centre of gaze are attracted to a (hidden) point moving in the visual field, similarly to what was proposed above in section 12.4.1. Finally, the generative model is equipped with a hierarchical structure, so that it can recognize and remember unseen (occluded) trajectories and emit anticipatory responses (see section 12.4.2).

We show in Figure 12.5 the results of this model for a two-layered hierarchical generative model. The hidden causes are informed by the dynamics of hidden states at the second level: These hidden states model underlying periodic dynamics using a simple periodic attractor that produces sinusoidal fluctuations of arbitrary amplitude or phase and a frequency that is determined by a second level hidden cause with a prior expectation of a frequency of η (in Hz). It is somewhat similar to a control system model that attempts to achieve zero-latency target tracking by fitting the trajectory to a (known) periodic signal (Bahill and McDonald, 1983). Our formulation ensures a Bayes

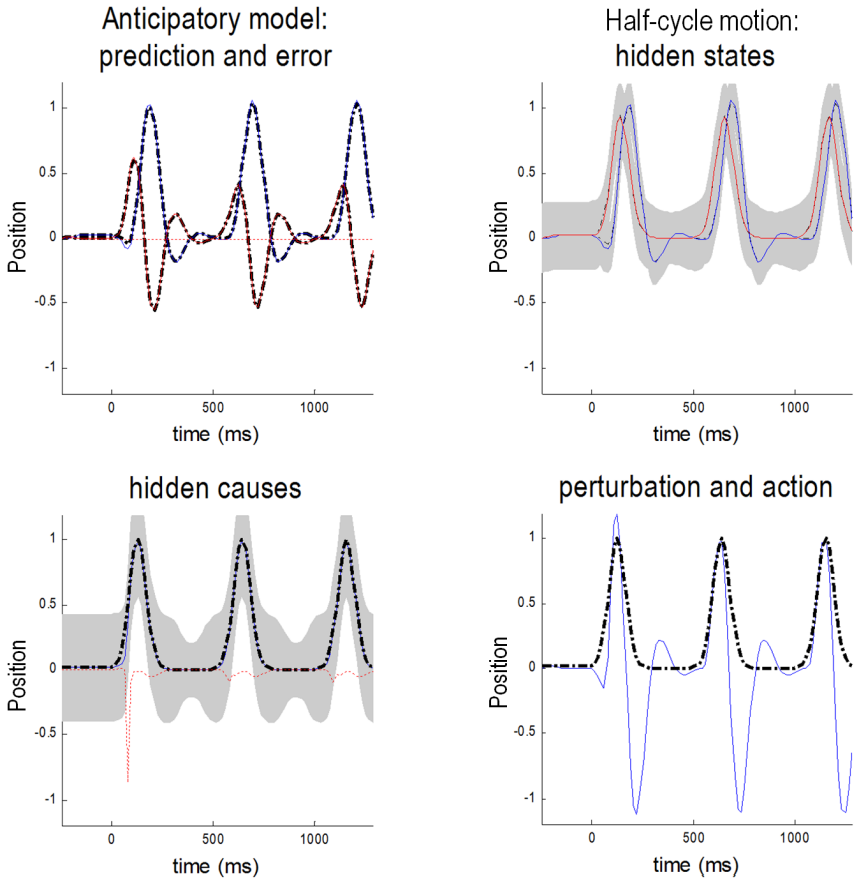


Figure 12.5: This figure reports the simulation of smooth pursuit when the target motion is hemi-sinusoidal, as would happen for a pendulum that would be stopped at each half cycle left of the vertical (broken black lines in the lower-right panel). We report the horizontal excursions of oculomotor angle in retinal space (two upper panels) and the angular position of the target in an intrinsic frame of reference (visual space, lower panels). The lower-right panel shows the true value of the displacement in visual space (broken black lines) and the action (blue line) which is responsible for oculomotor displacements. The upper left panel shows in retinal space the predicted sensory input (colored lines) and sensory prediction errors (dotted red lines) along with the true values (broken black lines). The latter is effectively the distance of the target from the centre of gaze and reports the spatial lag of the target that is being followed (solid red line). One can see clearly the initial displacement of the target that is suppressed after a few hundred milliseconds. The sensory predictions are based upon the conditional expectations of hidden oculomotor (blue line) and target (red line) angular displacements shown on the upper right. The grey regions correspond to 90% Bayesian confidence intervals and the broken lines show the true values of these hidden states. The generative model used here has been equipped with a second hierarchical level that contains hidden states, modeling latent periodic behavior of the (hidden) causes of target motion (states not shown here). The hidden cause of these displacements is shown with its conditional expectation on the lower left. The true cause and action are shown on the lower right. The action (blue line) is responsible for oculomotor displacements and is driven by the proprioceptive prediction errors.

optimal estimate of periodic motion in terms of a posterior belief about its frequency. In these simulations, we used a fixed Gaussian prior centered on the correct frequency with a period of 512 ms. This prior reproduces a typical experimental setting in which the oscillatory nature of the trajectory is known, but its amplitude and phase (onset) are unknown. Indeed, it has been shown that anticipatory responses are cofounded when randomizing the inter-cycle interval (Becker and Fuchs, 1985). In principle, we could have considered many other forms of generative model, such as models with prior beliefs about continuous acceleration (Bennett et al., 2010). With this addition, the improvement in pursuit accuracy apparent at the onset of the second cycle is consistent with what was observed empirically (Barnes et al., 2000).

This is because the model has an internal representation of latent causes of target motion that can be called upon even when these causes are not expressed explicitly in the target trajectory. These simulations speak to a straightforward and neurobiologically plausible solution to the generic problem of integrating information from different sources with different temporal delays and the particular difficulties encountered when a system — like the oculomotor system — tries to control its environment with delayed signals. Neurobiologically, the application of delay operators just means changing synaptic connection strengths to take different mixtures of generalized sensations and their prediction errors.

12.5 Reacting, inferring, predicting: a neural workspace

We have proposed herein a hierarchical inference network that can both estimate the direction and speed of a moving object despite the inherent ambiguities present in the images and predict the target trajectory from accumulated retinal and extra-retinal evidence. What could be the biologically-plausible implementation of such a hierarchy? What are its advantages for a living organism?

The fact that the pursuit system can be separated into two distinct blocks has been proposed by many others (see for recent reviews (Barnes, 2008; Lisberger, 2010; Masson et al., 2010; Fukushima et al., 2013). Such structure is rooted on the need to mix retinal and extra-retinal information to ensure stability of pursuit, as originally proposed by (Yasui and Young, 1975). Elaborations of this concept have formed the basis of a number of models based on a negative feedback control system (Robinson et al., 1986; Krauzlis and Lisberger, 1994). However, the fact that a simple efference copy feedback loop can not account for anticipatory responses during target blanking as well as for the role of expectation about future target trajectory (Barnes, 2008) or reinforcement learning during blanking (Madelain and Krauzlis, 2003) has called into question the validity of this simplistic approach (see (Fukushima et al., 2013) for a recent review). This has led to more complex models where an internal model of target motion is reconstructed from an early sampling and storage of target velocity and an efference copy of the eye's velocity signal. Several memory components have been proposed to serve different aspects of cognitive control of smooth pursuit and anticipatory responses (Barnes and Collins, 2011; Fukushima et al., 2013). The Hierarchical Bayesian model presented above (see section 12.4.2) follows the same structure with two main differences. First, the sensory pro-

cessing itself is seen as a dynamical network whereas most of the models cited in this section have oversimplified target velocity representation. Second, we collapse all the different internal blocks and loops into a single inference loop representing the perceived target motion.

We have proposed earlier that these two inference loops might be implemented by two large-scale brain networks (Masson et al., 2010). The dynamical inference loop is based on the properties of primate visual areas V1 and MT where local and global motion signals have been clearly identified, respectively (see (Masson and Ilg, 2010) for a series of recent review articles). MT neurons solve the aperture problem with a slow temporal dynamics. When presented with a set of elongated, tilted bars, their initial preferred direction matches the motion direction orthogonal to the bar orientation. From there, that preferred direction gradually rotate toward the true, 2D translation of the bar so that after about 120 ms, MT population signals the correct pattern motion direction, independently of the bar orientation (Pack and Born, 2001). Several models have proposed that such dynamics is due to reccurent interactions between V1 and MT cortical areas (e.g. (Tlapale et al., 2010; Bayerl and Neunmann, 2004; Berzhanskaya et al., 20)) and we have shown that the dynamical Bayesian model presented above give a good description of such neuronal dynamics and its perceptual and behavioural counterparts (Montagnini et al., 2007) (see also chapter `010_keil`). Such recurrent network can be upscaled to include other cortical visual areas involved in shape processing (e.g. areas V2, V3, V4) to further improve form-motion integration and select one target among several distractors or the visual background (Tlapale et al., 2010). The V1-MT loop exhibits however two fundamental properties with respect to the tracking of object motion. First, neuronal responses stop immediately when the retinal input disappears as during object blanking. Second, the loop remains largely immune to higher, cognitive inputs. For instance, the slow speed Prior used in our Bayesian model can hardly be changed by training in human observers (Montagnini et al., 2006).

In non-human primates, the medial superior temporal (MST) cortical area is essential for pursuit. It receives MT inputs about target direction and speed of visual pattern and represents the target's velocity vector. In the context of smooth pursuit control, neuronal activities in area MST show several interesting features. First, during pursuit, if the target is blanked, the neuronal activity is maintained throughout the course of the blanking. This is clearly different from area MT where neurons stop firing in the occurrence of even a brief disappearance of the target (Newsome et al., 1988). Second, both monkeys and humans can track imaginary large line-drawing targets where the central, foveal part is absent (Ilg and Thier, 1999). MST neurons, but not MT cells, can signal the motion direction of these parafoveal targets, despite the fact visual edges fall outside their receptive fields. Thus, in area MST, neurons are found whose activities are not different during pursuit of real (i.e. complete) or imaginary (i.e. parafoveal) targets (Ilg and Thier, 2003; Masson and Ilg, 2010). Lastly, many MST neurons can encode target motion veridically during eye movements in contrast to MT cells (Chukoskie and Movshon, 2008). All these evidences strongly suggest that MST neurons integrate both retinal and extra-retinal information to reconstruct the perceived motion of the target. However, despite the fact that areas MT and MST are strongly, and recurrently connected, the strong difference between MT and MST neuronal responses between pursuit seems to indicate that extra-retinal signals are not back propagated to early visual processing areas.

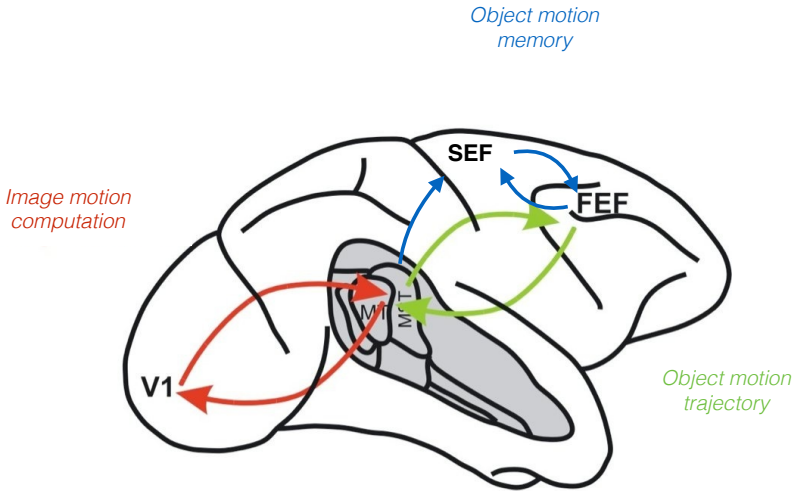


Figure 12.6: A lateral view of the macaque cortex. The neural network corresponding to our hierarchical Bayesian model of smooth pursuit is made of three main cortico-cortical loop. The first loop between primary visual cortex (V1) and the medio-temporal (MT) computes image motion and infers the optimal low-level solution for object motion direction and speed. Its main output is the medio-superior temporal (MST) area that acts as a gear between the sensory loop and the object motion computation loop. Retinal and extra-retinal signals are integrated in both MST and FEF areas. Such dynamical integration computes the perceived trajectory of the moving object and implements an online prediction that can be used on the course of a tracking eye movement to compensate for target perturbation such as transient blanking. FEF and MST areas signals are sent to the supplementary eye field (SEF) and the interconnected prefrontal cortical areas. This third loop can elaborate a motion memory of the target trajectory and is interfaced with higher cognitive processes such as cue instruction or reinforcement learning. It also implements offline predictions that can be used across trials, in particular to drive anticipatory responses to highly predictable targets.

Interestingly, several recent models have articulated their two-stage computational approach with this architecture (Pack et al., 2001; Grossberg et al., 2012) in order to model the dynamics of primate smooth-pursuit.

From there, output signals are sent in two directions. One signal reaches the brainstem oculomotor structures through the pontine nuclei, the cerebellar floccular region and the vestibular nuclei (Leigh and Zee, 2006). This cortico-subcortical pathway conveys the visual drive needed for pursuit initiation and maintenance. The second signal reaches the frontal cortex that includes the caudal parts of the frontal eye fields (FEF) and the supplementary eye field (SEF) (see (Krauzlis, 2004; Fukushima et al., 2013) for recent reviews). FEF neurons share many properties of MST cells. In particular, they integrate both retinal and extra-retinal information during pursuit so that responses remain sustained

during blanking or when simulated with imaginary targets (Ilg and Thier, 2003; Masson et al., 2010). Moreover, both FEF and MST cells show a build-up of activity during anticipatory pursuit (Ilg, 2003; Fukushima et al., 2002). Thus, FEF and MST appear to be strongly coupled to build an internal representation of target motion that can be used during steady-state tracking as well as during early phases of pursuit. Moreover, FEF area issues pursuit commands that are sent to the brainstem nucleus reticularis tegmenti pontis (NRTP) and the cerebellar vermis lobules before reaching the pursuit oculomotor structures. Several authors have proposed that such parieto-frontal loop might implement the prediction component of the pursuit responses (see (Masson et al., 2010) for review). Other have proposed to restrict its role to the computation of the perceived motion signal that drive the pursuit response (Fukushima et al., 2013) while higher signals related to prediction might be computed in more anterior areas such as SEF and prefrontal cortex (PFC).

Prediction is influenced by many cognitive inputs (cues, working memory, target selection) (Barnes, 2008). Accordingly, prediction-related neuronal responses during pursuit have been reported in area SEF (Heinen, 1995) and the caudal part of FEF (Fukushima et al., 2002). Moreover, SEF activity facilitates anticipatory pursuit responses to highly predictable targets (Missal and Heinen, 2004). The group of Fukushima have identified several subpopulations of neurons in both areas that can encode directional visual motion-memory, independently of movement preparation signals (see (Fukushima et al., 2013) for a complete review). However, FEF neurons are more often mixing predictive and motor preparation signals while SEF cells more specifically encode a visual motion memory signal. This is consistent with the fact that many neurons in the prefrontal cortex have been linked to temporal storage of sensory signals (Goldman-Rakic, 1995). Thus, a working memory of target motion might be formed in area SEF by integrating multiple inputs from parietal (MST) and prefrontal (FEF, PFC) cortical areas. Fukushima et al. (2013) proposed that a recurrent network made of these areas (MST, FEF, SEF, PFC) might signal future target motion using prediction, timing, expectation as well as experience gained over trials.

All these studies define a neuronal workspace for our hierarchical inference model, as illustrated in Figure 12.6. Two main loops seem to be at work. A first loop predicts the optimal target motion direction and speed from sensory evidence (image motion computation, in red). It uses sensory priors such as the "smooth and slow motion prior" used for both perception and pursuit initiation that are largely immune to higher influence. By doing so, the sensory loop can preserve its ability to quickly react to new sensory event and avoid the inertia of prediction systems. This loop would correspond to the reactive pathway of the pursuit model proposed by Barnes (2008); Fukushima et al. (2013). On the ground of some behavioral evidence (Montagnini et al., 2006), we believe that target motion prediction can not easily overcome the aperture problem (see also the open questions discussed in section 12.6), providing a strong indication that this sensory loop is largely not penetrable to cognitive influences. The second loop involves areas MST and FEF to compute and store online target motion by taking into account both sensory and efference copy signals (object motion computation, green). A key aspect is that MST must act as a gear to prevent predictive or memory-related signals to back propagate downstream to the sensory loop. We propose to distinguish between online prediction, involving these

two areas during on-going event such as target blanking to offline prediction. The later is based on a memory of target motion that span across trials and might be used to both trigger anticipatory responses or drive responses based on cues. It might most certainly involve a dense, recurrent prefrontal network articulated around SEF and that offers a critical interface with the cognitive processes interfering with pursuit control (Object memory loop, in blue).

This architecture presents many advantages. First, it preserves the brain to quickly react to new event as a brutal change in target motion direction. Second, it ensures maintaining pursuit in a wide variety of conditions with good stability and by constructing an internal model of target motion it allows a tight coordination between pursuit and saccades (Orban de Xivry and Lefèvre, 2004). Third, it provides an interface with higher cognitive aspects of sensorimotor transformation. Several questions remain however unsolved. Because of the strong changes seen in prediction with different behavioral context, several models, including the one presented here, postulate the existence of hard switches that can turn on or off the contribution of a particular model component. We need a better theoretical approach about decision making between these different loops. The Bayesian approach proposed here, similar to the Kalman filter models, opens the door to better understanding these transitions. It proposes that each signal (sensory evidence, motion memory, prediction) can be weighted from its reliability. Such unifying theoretical approach can then be used to design new behavioral and physiological experiments.

12.6 Conclusion

Recent experimental evidence points to the need to revise our view of the primates' smooth pursuit system. Rather than a reflexive velocity-matching negative-feedback loop, the human motion tracking system seems to be grounded on a complex set of dynamic functions that subserve a quick and accurate adaptive behavior even in visually challenging situations. By analyzing tracking eye movements produced with a simple, unique and highly-visible moving target, many of these notions could not have clearly emerged and it is now clear that testing more naturalistic visual motion contexts and carefully taking into account the sources of uncertainty at different scales is a crucial step for understanding biological motion tracking. The approach highlighted here opens the door to several open questions.

First, we have focused herein on luminance-based motion processing mechanisms. Such inputs can be well extracted by a bank of filters extracting motion energy at multiple scales. The human visual motion system is however more versatile and psychophysical studies have demonstrated that motion perception is based on cues that can be defined in many different ways. These empirical studies have led to the three-systems theory of human visual motion perception by Lu and Sperling (2001). Beside the first-order system that responds to moving luminance patterns, a second-order system responds to moving modulations of feature types (i.e. stimuli where the luminance is the same everywhere but an area of higher contrast or of flicker moves). A third-order system slowly computes the motion of marked locations in a "salience map" where locations of important visual features in the visual space (i.e. a figure) are highlighted respective to its "background" (see Chapter `013_lemeur`). The contribution of

the first-order motion to the initiation of tracking responses have been largely investigated (see (Masson and Perrinet, 2012; Lisberger, 2010) for reviews). More recent studies have pointed out that feature tracking mechanisms (i.e. a second-order system) are critical for finely adjusting this initial eye acceleration to object speed when reaching steady-state tracking velocity (Wilmer and Nakayama, 2007). Whether the third-order motion system is involved, and how in the attentional modulation of tracking, is currently under investigation by many research groups. Altogether, these psychophysical and theoretical studies point towards the need for more complex front-end layers of visuomotor models so that artificial systems would more versatile and adapted to complex, ambiguous environments.

Second, although the need of hierarchical, multi-scale inferential models is now apparent, current models will have to meet the challenge of explaining a rich and complex set of behavioral data. Just to detail an example, somewhat unexpectedly we have found that predicting 2D motion trajectories does not help solving the aperture problem (Montagnini et al., 2006). Indeed, when human subjects are exposed to repeated motion conditions for a horizontally-translating tilted bar across several experimental trials and they develop anticipatory smooth pursuit in the expected direction, still their pursuit traces reflect the aperture-induced bias at initiation, as illustrated in Figure 12.1, panel *d*. It is important to notice that the robust optimal cue combination of the output of sensory and predictive motion processing modules postulated in the previous section (Orban de Xivry et al., 2013; Bogadhi et al., 2013) are not capable, at this stage, to explain this phenomenon. Other crucial issues deserve to be better understood and modeled in this new framework, such as, for instance, the dynamic interaction between global motion estimate and object segmentation, the precise functional and behavioral relationship between smooth tracking and discrete jump-like saccadic movements, or the role of high-level cognitive cues in modulating motion tracking (Kowler et al., 2014).

12.6.1 Interest for computer vision

Machine-based motion tracking, similar to human motion tracking, needs to be robust to visual perturbations. As we have outlined above, including some form of predictive information might be extremely helpful to stabilize object tracking, in much the same way as what happens during transient target blanking for human smooth pursuit. Importantly, human tracking seems to be equipped with a more complete and advanced software for motion tracking, namely one that allows us i) to exploit the learned regularities of motion trajectories to anticipate and compensate for sensorimotor delays and consequent mismatch between gaze and target position in the event of a complex trajectory; ii) to keep tracking an object temporarily occluded by other objects; and iii) to start tracking in the dark when the future motion of an object is predictable, or even partly predictable. Some of these predictive cues might depend upon a high-level cognitive representation of kinematic rules or arbitrary stimulus-response associations. This type of apparatus is probably much too advanced to be implemented into a common automatic motion tracking device, but still it may be a source of inspiration for future advanced developments for machine vision and tracking.

Acknowledgments

The authors were supported by EC IP project FP7-269921, "BrainScaleS" and the French ANR-BSHS2-2013-006, "Speed'. They are grateful to Amarender Bogadhi, Laurent Madelain, Frederic Chavane and all the colleagues in the In-Vibe team for many enriching discussions. L.U.P wishes to thank Karl Friston and Rick Adams and the The Wellcome Trust Centre for Neuroimaging, University College London, for their essential contribution in the closed-loop delayed model. The chapter was written when G.S.M was an invited fellow from the CONICYT program, Chili.
Correspondence and requests for materials should be addressed to A.M. (email:Anna.Montagnini@univ-amu.fr).

Bibliography

C. Rashbass. The relationship between saccadic and smooth tracking eye movements. *Journal of Physiology*, 159:326–338, 1961.

Eileen Kowler. Eye movements: The past 25years. *Vision Research*, 51(13): 1457–1483, 2011. ISSN 00426989. doi: 10.1016/j.visres.2010.12.014.

Eileen Kowler, Cordelia D Aitkin, Nicholas M Ross, Elio M Santos, and Min Zhao. Davida teller award lecture 2013: the importance of prediction and anticipation in the control of smooth pursuit eye movements. *Journal of Vision*, 14(5):10, 2014. doi: 10.1167/14.5.10.

J H Maunsell and William T Newsome. Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, 10:363–401, 1987. doi: 10.1146/annurev.ne.10.030187.002051.

David C Bradley and Manu S Goyal. Velocity computation in the primate visual system. *Nature reviews. Neuroscience*, 9(9):686–95, Sep 2008. doi: 10.1038/nrn2472.

Richard T Born and David C Bradley. Structure and function of visual area MT. *Annual Review of Neuroscience*, 28:157–89, 2005. doi: 10.1146/annurev.neuro.26.041002.131052.

William T Newsome and R H Wurtz. Probing visual cortical function with discrete chemical lesions. *Trends Neurosci*, 11(9):394–400, Sep 1988.

Stephen G Lisberger and J Anthony Movshon. Visual motion analysis for pursuit eye movements in area MT of macaque monkeys. *Journal of Neuroscience*, 19(6):2224–46, Mar 1999.

William T Newsome, R H Wurtz, and H Komatsu. Relation of cortical areas MT and MST to pursuit eye movements. ii. differentiation of retinal from extraretinal inputs. *Journal of Neurophysiology*, 60(2):604–20, Aug 1988.

Richard J Krauzlis. Recasting the smooth pursuit eye movement system. *Journal of Neurophysiology*, 91(2):591–603, Feb 2004. doi: 10.1152/jn.00801.2003.

B. Hassenstein and W. Reichardt. Systemtheoretische analyze der zeit-, reihenfolgen- und vorzeichenauswer- tung bei der bewegungsperzeption des russelkafers chlorophanus. *Zeitung für Naturforschung*, 11b:513–524, 1956.

A. Borst. Fly visual course control: behavior, algorithms and circuits. *Nature reviews. Neuroscience*, 15:590–599, 2014.

A. Borst, J. Haag, and D.F. Reiff. Fly motion vision. *Annual Review of Neuroscience*, 33:49–70, 2010.

R.R. Harrison and C. Koch. A robust analog vlsi reichardt motion sensor. *Analog and Integrated Circuits and Signal Processing*, 24:213–229, 2000.

T. Kohler, F. Röchter, J.P. Lindemann, and R. Möller. Bio-inspired motion detection in a fpga-based smart camera. *Bioinspiration and Biomimetics*, 4: 015008, 2009.

Guillaume S Masson. From 1D to 2D via 3D: dynamics of surface motion segmentation for ocular tracking in primates. *Journal of Physiology-Paris*, 98 (1-3):35–52, 2004. doi: 10.1016/j.jphysparis.2004.03.017.

Stephen G Lisberger. Visual guidance of smooth-pursuit eye movements: sensation, action, and what happens in between. *Neuron*, 66(4):477–91, May 2010. doi: 10.1016/j.neuron.2010.03.027.

Miriam Spering and Anna Montagnini. Do we track what we see? common versus independent processing for motion perception and smooth pursuit eye movements: a review. *Vision Research*, 51(8):836–52, Apr 2011. doi: 10.1016/j.visres.2010.10.017.

Guillaume S. Masson and Laurent U Perrinet. The behavioral receptive field underlying motion integration for primate tracking eye movements. *Neuroscience & Biobehavioral Reviews*, 36(1):1–25, January 2012. ISSN 0149-7634. doi: 10.1016/j.neubiorev.2011.03.009. URL `http://dx.doi.org/10.1016/j.neubiorev.2011.03.009`.

Leslie C Osborne. Computation and physiology of sensory-motor processing in eye movements. *Current Opinion in Neurobiology*, 21(4):623–8, Aug 2011. doi: 10.1016/j.conb.2011.05.023.

Stephen G Lisberger, E. J. Morris, and L. Tychsen. Visual motion processing and sensory-motor integration for smooth pursuit eye movements. *Annual Review of Neuroscience*, 10, 1987. doi: 10.1146/annurev.ne.10.030187.000525. URL `http://dx.doi.org/10.1146/annurev.ne.10.030187.000525`.

Alexander C Schütz, Doris I Braun, J Anthony Movshon, and Karl R Gegenfurtner. Does the noise matter? effects of different kinematogram types on smooth pursuit eye movements and perception. *Journal of Vision*, 10(13):26, 2010. doi: 10.1167/10.13.26.

Leslie C Osborne, Stephen G Lisberger, and William Bialek. A sensory source for motor variation. *Nature*, 437(7057):412–6, Sep 2005. doi: 10.1038/nature03961.

Leslie C Osborne, Sonja S Hohl, William Bialek, and Stephen G Lisberger. Time course of precision in smooth-pursuit eye movements of monkeys. *Journal of Neuroscience*, 27(11):2987–98, Mar 2007. doi: 10.1523/JNEUROSCI.5072-06.2007.

Leslie C Osborne, William Bialek, and Stephen G Lisberger. Time course of information about motion direction in visual area MT of macaque monkeys. *Journal of Neuroscience*, 24(13):3210–22, Mar 2004. doi: 10.1523/JNEUROSCI.5305-03.2004.

Miriam Spering, D Kerzel, Doris I Braun, MJ Hawken, and KR Gegenfurtner. Effects of contrast on smooth pursuit eye movements. *Journal of Vision*, 20 (5):455–465, 2005. URL `http://w.journalofvision.org/content/5/5/6.full.pdf`.

Yair Weiss, Eero P. Simoncelli, and Edward H. Adelson. Motion illusions as optimal percepts. *Nature Neuroscience*, 5(6):598–604, June 2002. ISSN 1097-6256. doi: 10.1038/nn858. URL `http://dx.doi.org/10.1038/nn858`.

S Wuerger, R Shapley, and N Rubin. "on the visually perceived direction of motion" by hans wallach: 60 years later. *Perception*, 25(11):1317–1367, 1996. doi: 10.1068/p251317.

G S Masson and U. J. Ilg, editors. *Dynamics of visual motion processing. Neuronal, behavioral and computational approaches.* Springer Verlag, 2010.

CL Fennema and WB Thompson. Velocity determination in scenes containing several moving images. *Computer Graphics and Image Processing*, 9:301–315, 1979.

Edward H. Adelson and J Anthony Movshon. Phenomenal coherence of moving visual patterns. *Nature*, 300(5892):523–5, Dec 1982.

C Yo and H R Wilson. Perceived direction of moving two-dimensional patterns depends on duration, contrast and eccentricity. *Vision Research*, 32(1):135–47, Jan 1992.

J Lorenceau, M Shiffrar, N Wells, and Eric Castet. Different motion sensitive units are involved in recovering the direction of moving lines. *Vision Research*, 33(9):1207–17, Jun 1993.

A Gorea and J Lorenceau. Directional performances with moving plaids: component-related and plaid-related processing modes coexist. *Spatial Vision*, 5(4):231–52, 1991.

H R Wilson, V P Ferrera, and C Yo. A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience*, 9(1):79–97, Jul 1992.

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

T D Albright. Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, 52(6):1106–30, Dec 1984.

J. Anthony Movshon, Edward H. Adelson, Martin S. Gizzi, and William T Newsome. The analysis of moving visual patterns. In C. Chagas, R. Gattass, and C. Gross, editors, *Pattern Recognition Mechanisms*, volume 54, pages 117–151. Rome: Vatican Press, 1985.

Laurent U Perrinet and Guillaume S. Masson. Motion-Based prediction is sufficient to solve the aperture problem. *Neural Computation*, 24(10):2726–2750, October 2012. ISSN 1530-888X. doi: 10.1162/NECO_a_00332. URL http://dx.doi.org/10.1162/NECO_a_00332.

Pierre Y. Burgi, Alan L. Yuille, and Norberto M Grzywacz. Probabilistic motion estimation based on temporal coherence. *Neural Computation*, 12(8):1839–67, August 2000. ISSN 0899-7667. URL http://portal.acm.org/citation.cfm?id=1121336.

Guillaume S Masson and Leland S Stone. From following edges to pursuing objects. *Journal of Neurophysiology*, 88(5):2869–73, Nov 2002. doi: 10.1152/jn.00987.2001.

Julian M Wallace, Leland S Stone, and Guillaume S Masson. Object motion computation for the initiation of smooth pursuit eye movements in humans. *Journal of Neurophysiology*, 93(4):2279–93, Apr 2005. doi: 10.1152/jn.01042.2004.

Richard T Born, Christopher C Pack, Carlos R Ponce, and Si Yi. Temporal evolution of 2-dimensional direction signals used to guide eye movements. *Journal of Neurophysiology*, 95(1):284–300, Jan 2006. doi: 10.1152/jn.01329.2005.

Anna Montagnini, Pascal Mamassian, Laurent U Perrinet, Eric Castet, and Guillaume S. Masson. Bayesian modeling of dynamic motion integration. *Journal of Physiology-Paris*, 101(1-3):64–77, January 2007. ISSN 0928-4257. doi: 10.1016/j.jphysparis.2007.10.013. URL http://dx.doi.org/10.1016/j.jphysparis.2007.10.013.

Amarender R. Bogadhi, Anna Montagnini, Pascal Mamassian, Laurent U Perrinet, and Guillaume S. Masson. Pursuing motion illusions: A realistic oculomotor framework for Bayesian inference. *Vision Research*, 51(8):867–880, April 2011. ISSN 00426989. doi: 10.1016/j.visres.2010.10.021. URL http://dx.doi.org/10.1016/j.visres.2010.10.021.

G M Gauthier and J M Hofferer. Eye tracking of self-moved targets in the absence of vision. *Exp Brain Res*, 26(2):121–39, Sep 1976.

Graham R Barnes. Cognitive processes involved in smooth pursuit eye movements. *Brain Cogn*, 68(3):309–26, Dec 2008. doi: 10.1016/j.bandc.2008.08.020.

Anna Montagnini, Miriam Spering, and Guillaume S Masson. Predicting 2D target velocity cannot help 2D motion integration for smooth pursuit initiation. *Journal of Neurophysiology*, 96(6):3545–50, Dec 2006. doi: 10.1152/jn.00563.2006.

Graham R Barnes and A M Schmid. Sequence learning in human ocular smooth pursuit. *Exp Brain Res*, 144(3):322–35, Jun 2002. doi: 10.1007/s00221-002-1050-8.

Simon J Bennett, Jean-Jacques Orban de Xivry, Graham R Barnes, and Philippe Lefèvre. Target acceleration can be extracted and represented within the predictive drive to ocular pursuit. *Journal of Neurophysiology*, 98(3): 1405–14, Sep 2007. doi: 10.1152/jn.00132.2007.

Anna Montagnini, David Souto, and Guillaume S. Masson. Anticipatory eye-movements under uncertainty: a window onto the internal representation of a visuomotor prior. *Journal of Vision*, 10(7):554, 2010.

E Kowler, A J Martins, and M Pavel. The effect of expectations on slow oculomotor control–iv. anticipatory smooth eye movements depend on prior target motions. *Vision Research*, 24(3):197–210, 1984.

E Kowler. Cognitive expectations, not habits, control anticipatory smooth oculomotor pursuit. *Vision Research*, 29(9):1049–57, 1989.

W. Becker and A. F. Fuchs. Prediction in the oculomotor system: smooth pursuit during transient disappearance of a visual target. *Experimental brain research*, 57(3):562–575, 1985. ISSN 0014-4819. doi: 10.1007/BF00237843. URL `http://dx.doi.org/10.1007/BF00237843`.

Laurent Madelain and Richard J. Krauzlis. Effects of learning on smooth pursuit during transient disappearance of a visual target. *Journal of neurophysiology*, 90(2):972–982, August 2003. ISSN 0022-3077. doi: 10.1152/jn.00869.2002. URL `http://dx.doi.org/10.1152/jn.00869.2002`.

Simon J. Bennett and Graham R. Barnes. Human ocular pursuit during the transient disappearance of a visual target. *Journal of Neurophysiology*, 90(4): 2504–2520, October 2003. ISSN 0022-3077. doi: 10.1152/jn.01145.2002. URL `http://dx.doi.org/10.1152/jn.01145.2002`.

Mark M. Churchland, I-Han H. Chou, and Stephen G. Lisberger. Evidence for object permanence in the smooth-pursuit eye movements of monkeys. *Journal of neurophysiology*, 90(4):2205–2218, October 2003. ISSN 0022-3077. doi: 10.1152/jn.01056.2002. URL `http://dx.doi.org/10.1152/jn.01056.2002`.

Jean-Jacques Orban de Xivry, Simon J Bennett, Philippe Lefèvre, and Graham R Barnes. Evidence for synergy between saccades and smooth pursuit during transient target disappearance. *Journal of Neurophysiology*, 95(1): 418–27, Jan 2006. doi: 10.1152/jn.00596.2005.

Jean-Jacques Orban de Xivry, Marcus Missal, and Philippe Lefèvre. A dynamic representation of target motion drives predictive smooth pursuit during target blanking. *Journal of Vision*, 8(15):6.1–13, 2008. doi: 10.1167/8.15.6.

David Alais and David Burr. The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, 14(3):257–62, Feb 2004. doi: 10.1016/j.cub.2004.01.029.

Marc O Ernst and Martin S Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–33, Jan 2002. doi: 10.1038/415429a.

Mina Khoei, Guillaume Masson, and Laurent U Perrinet. Motion-based prediction explains the role of tracking in motion extrapolation. *Journal of Physiology-Paris*, 107(5):409–420, 2013. ISSN 09284257. doi: 10.1016/j.jphysparis.2013.08.001.

Amarender Bogadhi, Anna Montagnini, and Guillaume Masson. Dynamic interaction between retinal and extraretinal signals in motion integration for smooth pursuit. *Journal of Vision*, 13(13):5, 2013. ISSN 1534-7362. doi: 10.1167/13.13.5.

Alan A. Stocker and Eero P. Simoncelli. Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, 9(4): 578–585, March 2006. ISSN 1097-6256. doi: 10.1038/nn1669. URL `http://www.nature.com/neuro/journal/v9/n4/abs/nn1669.html`.

Laurent U Perrinet and Guillaume S. Masson. Modeling spatial integration in the ocular following response using a probabilistic framework. *Journal of Physiology-Paris*, 101(1-3):46–55, January 2007. ISSN 0928-4257. doi: 10.1016/j.jphysparis.2007.10.011. URL `http://dx.doi.org/10.1016/j.jphysparis.2007.10.011`.

Kameliya Dimova and Michael Denham. A neurally plausible model of the dynamics of motion integration in smooth eye pursuit based on recursive bayesian estimation. *Biological Cybernetics*, 100(3):185–201, Mar 2009. doi: 10.1007/s00422-009-0291-z.

Jean-Jacques Orban de Xivry, Sébastien Coppe, Gunnar Blohm, and Philippe Lefèvre. Kalman filtering naturally accounts for visually guided and predictive smooth pursuit dynamics. *Journal of Neuroscience*, 33(44):17301–13, Oct 2013. doi: 10.1523/JNEUROSCI.2321-13.2013.

Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D): 35–45, 1960.

Christopher R Fetsch, Alexandre Pouget, Gregory C DeAngelis, and Dora E Angelaki. Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience*, 15(1):146–54, Jan 2012. doi: 10.1038/nn.2983.

Christopher C Pack and Richard T Born. Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature*, 409, 2001.

D. Goldreich, Richard J Krauzlis, and Stephen G Lisberger. Effect of changing feedback delay on spontaneous oscillations in smooth pursuit eye movements of monkeys. *Journal of neurophysiology*, 67(3):625–638, March 1992. ISSN 0022-3077. URL `http://view.ncbi.nlm.nih.gov/pubmed/1578248`.

Laurent U. Perrinet, Rick A. Adams, and Karl J. Friston. Active inference, eye movements and oculomotor delays. *Biological Cybernetics*, 108(6):777–801, December 2014. ISSN 1432-0770. doi: 10.1007/s00422-014-0620-8. URL `http://dx.doi.org/10.1007/s00422-014-0620-8`.

Romi Nijhawan. Visual prediction: Psychophysics and neurophysiology of compensation for time delays. *Behavioral and Brain Sciences*, 31(02):179–198, May 2008. ISSN 1469-1825. doi: 10.1017/s0140525x08003804. URL `http://people.psych.cornell.edu/~{}jec7/pubs/nihjawan.pdf`.

Andre M. Bastos, W. Martin Usrey, Rick A. Adams, George R. Mangun, Pascal Fries, and Karl J. Friston. Canonical microcircuits for predictive coding. *Neuron*, 76(4):695–711, November 2012. ISSN 0896-6273. doi: 10.1016/j.neuron.2012.10.038. URL `http://dx.doi.org/10.1016/j.neuron.2012.10.038`.

A. T. Bahill and J. D. McDonald. Model emulates human smooth pursuit system producing zero-latency target tracking. *Biological cybernetics*, 48(3): 213–222, 1983. ISSN 0340-1200. URL `http://view.ncbi.nlm.nih.gov/pubmed/6639984`.

Simon J. Bennett, Jean-Jacques Orban de Xivry, Philippe Lefèvre, and Graham R. Barnes. Oculomotor prediction of accelerative target motion during occlusion: long-term and short-term effects. *Experimental Brain Research*, 204(4):493–504, June 2010. ISSN 1432-1106. doi: 10.1007/s00221-010-2313-4. URL `http://dx.doi.org/10.1007/s00221-010-2313-4`.

Graham R. Barnes, D. M. Barnes, and S. R. Chakraborti. Ocular pursuit responses to repeated, single-cycle sinusoids reveal behavior compatible with predictive pursuit. *Journal of Neurophysiology*, 84(5):2340–2355, November 2000. ISSN 1522-1598. URL `http://jn.physiology.org/content/84/5/2340.abstract`.

Guillaume S Masson, Anna Montagnini, and Uwe J Ilg. *Biological motion processing:*, chapter When the brain meets the eye: tracking object motion, pages 161–188. Number 8. Springer Verlag, 2010.

K Fukushima, J Fukushima, T Warabi, and Graham R Barnes. Cognitive processes involved in smooth pursuit eye movements:behavioral evidence, neural substrate and clinical correlation. *Frontiers in Systems Neuroscience*, 7(4): 1–28, 2013.

S Yasui and LR Young. Perceived visual motion as effective stimulus to pursuit eye movement system. *Science*, 190(906-908), 1975.

D. A. Robinson, J. L. Gordon, and S. E. Gordon. A model of the smooth pursuit eye movement system. *Biological Cybernetics*, 55(1):43–57, October 1986. ISSN 0340-1200. doi: 10.1007/bf00363977. URL `http://dx.doi.org/10.1007/bf00363977`.

Richard J Krauzlis and Stephen G Lisberger. A model of visually-guided smooth pursuit eye movements based on behavioral observations. *Journal of Computational Neuroscience*, 1:265–283, 1994.

Graham R Barnes and CJS Collins. The influence of cues and stimulus history on the non-linear frequency characteristics of the pursuit response to randomized target motion. *Experimental Brain Research*, 212:225–240, 2011.

Emilien Tlapale, Guillaume S. Masson, and Pierre Kornprobst. Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism. *Vision Research*, 50(17):1676–1692, August 2010. ISSN 0042-6989. doi: 10.1016/j.visres.2010.05.022. URL `http://dx.doi.org/10.1016/j.visres.2010.05.022`.

P Bayerl and H Neunmann. Disambiguating visual motion through contextual feedback modulation. *Neural Computation*, 16(2041-2066), 2004.

J Berzhanskaya, S Grossberg, and E Mingolla. Laminar cortical dynamics of visual form and motion interactions during coherent object motion perception. *Spatial Vision*, (337-395), 20.

Uwe J Ilg and PP Thier. Eye movements of rhesus monkeys directed towards imaginary targets. *Vision Research*, 39:2143–2150, 1999.

Uwe J Ilg and Peter P. Thier. Visual tracking neurons in primate area MST are activated by Smooth-Pursuit eye movements of an "imaginary" target. *Journal of Neurophysiology*, 90(3):1489–1502, May 2003. ISSN 1522-1598. doi: 10.1152/jn.00272.2003. URL `http://dx.doi.org/10.1152/jn.00272.2003`.

LJ Chukoskie and J Anthony Movshon. Modulation of visual signals in macaque MT and MST neurons during pursuit eye movements. *Journal of Neurophysiology*, 102:3225–3233, 2008.

C.C. Pack, S. Grossberg, and E. Mingolla. A neural model of smooth pursuit control and motion. *Journal of cognitive neuroscience*, 13(1):102–120, January 2001.

S. Grossberg, K. Srihasam, and D. Bullock. Neural dynamics of saccadic and smooth pursuit coordination during visual tracking of unpredictably moving targets. *Neural Networks*, 27:1–20, 2012.

R Leigh and DS Zee. *The neurology of eye movements*. Oxford University Press, New York, NY, USA, 4th edition, 2006.

Uwe J Ilg. Visual tracking neurons in area mst are activated during anticipatory pursuit eye movements. *Neuroreport*, 14:2219–2223, 2003.

K Fukushima, T Yamatobe, Y Shinmei, and J Fukushima. Predictive responses of periarcuate pursuit neurons to visual target motion. *Experimental Brain Research*, (145):104–120, 2002.

SJ Heinen. Single neuron activity in the dorsolateral frontal cortex during smooth pursuit eye movements. *Experimental Brain Research*, (104):357–361, 1995.

M Missal and SJ Heinen. Supplementary eye fields stimulation facilitates anticipatory pursuit. *Journal of Neurophysiology*, (92):1257–1262, 2004.

PS Goldman-Rakic. Cellular basis of working memory. *Neuron*, 14:477–485, 1995.

JJ Orban de Xivry and P Lefèvre. Saccades and pursuit: two outcomes of a single sensorimotor process. *Journal of Physiology*, 584:11–23, 2004.

Z-L Lu and G. Sperling. Three-systems theory of human visual motion perception: review and update. *Journal of the Optical Society of America. A*, 18: 2331–2370, 2001.

J.B. Wilmer and K. Nakayama. Two distinct visual motion mechanisms for smooth pursuit: evidence from individual differences. *Neuron*, 54:987–1000, 2007.