



# Motion-based prediction explains the role of tracking in motion extrapolation

Mina A Khoei, Guillaume S. Masson, Laurent Perrinet

## ► To cite this version:

Mina A Khoei, Guillaume S. Masson, Laurent Perrinet. Motion-based prediction explains the role of tracking in motion extrapolation. *Journal of Physiology - Paris*, 2013, 107 (5), pp.409-420. 10.1016/j.jphysparis.2013.08.001 . hal-02387884

**HAL Id: hal-02387884**

**<https://hal.science/hal-02387884>**

Submitted on 14 Jan 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Motion-based prediction explains the role of tracking in motion extrapolation

Mina A. Khoei, Guillaume S. Masson and Laurent U. Perrinet\*

Institut de Neurosciences de la Timone, UMR 7289, CNRS / Aix-Marseille Université -  
27, Bd. Jean Moulin, 13385 Marseille Cedex 5, France

## Abstract

During normal viewing, the continuous stream of visual input is regularly interrupted, for instance by blinks of the eye. Despite these frequent blanks (that is the transient absence of a raw sensory source), the visual system is most often able to maintain a continuous representation of motion. For instance, it maintains the movement of the eye such as to stabilize the image of an object. This ability suggests the existence of a generic neural mechanism of motion extrapolation to deal with fragmented inputs. In this paper, we have modeled how the visual system may extrapolate the trajectory of an object during a blank using motion-based prediction. This implies that using a prior on the coherency of motion, the system may integrate previous motion information even in the absence of a stimulus. In order to compare with experimental results, we simulated tracking velocity responses. We found that the response of the motion integration process to a blanked trajectory pauses at the onset of the blank, but that it quickly recovers the information on the trajectory after reappearance. This is compatible with behavioral and neural observations on motion extrapolation. To understand these mechanisms, we have recorded the response of the model to a noisy stimulus. Crucially, we found that motion-based prediction acted at the global level as a gain control mechanism and that we could switch from a smooth regime to a binary tracking behavior where the dot is tracked or lost. Our results imply that a local prior implementing motion-based prediction is sufficient to explain a large range of neural and behavioral results at a more global level. We show

that the tracking behavior deteriorates for sensory noise levels higher than a certain value, where motion coherency and predictability fail to hold longer. In particular, we found that motion-based prediction leads to the emergence of a tracking behavior only when enough information from the trajectory has been accumulated. Then, during tracking, trajectory estimation is robust to blanks even in the presence of relatively high levels of noise. Moreover, we found that tracking is necessary for motion extrapolation, this calls for further experimental work exploring the role of noise in motion extrapolation.

## Keywords

motion detection; motion extrapolation; probabilistic representation; predictive coding; contrast response function; gain control

## 1 Introduction

### 1.1 Problem statement

The continuous flow of information originating from the visual world is constantly fragmented by different sources of noise, occlusions or blanks. For instance, the path of a moving object can often be transiently blocked from the observer's line of sight. However, one is still able to judge the current position of a moving object during such periods of occlusion as well as estimate its future trajectory at its reappearance. This ability to transform such fragmented sensory inputs into a correct continuous representation has been a major pressure in the evolution of visual systems because it leads to

---

\*Corresponding author, Laurent.Perrinet@univ-amu.fr

appropriate reactions matched to the physical evidences: It is vital to accurately follow the trajectory of a fleeing prey and stabilize its image onto the retina in order to catch it or, on the contrary, to escape from an approaching predator, despite the fact that it can transiently disappear from the line of sight [19]. The problem of motion occlusion is a particular case of a more general problem in neuroscience: *motion extrapolation*. In the absence of sensory input, the visual system can extrapolate the instantaneous position of a moving object from its past trajectory.

An essential clue to solve that problem is the prior knowledge that objects follow smooth, coherent trajectories. Following the first law of newtonian mechanics, the trajectory of an object is only perturbed by external forces. Since we know a priori that these forces are more likely to be small compared to the inertia of an object of relevance, the trajectory of objects in the physical world tend to follow smooth, straight trajectories. As such, the projection of these trajectories on the retinotopic space is such that the statistics of natural images also exhibit similar regularities regarding their visual trajectories. Such prior knowledge may be the basis of learning processes based on the prediction of the path of the trajectory. During transient blanking, it is most likely that such processes (along with the knowledge that the sensory input was indeed blanked and not definitively removed) are at the root of the mechanisms underlying motion extrapolation. Their behavioral consequences are well known. For instance, when a moving target disappears, smooth pursuit eye movements continue at the same velocity during the initial period of occlusion [9] and such a feat is only possible when observers have some knowledge on the path of motion [20]. Therefore, there must be some underlying neural computations but it is yet not clear how this can be done efficiently and where it is implemented in the visual system.

This perceptual phenomenon provides invaluable tools with which we may study the mechanisms of motion detection and draw inferences about the properties of underlying neural populations. First, it is involved in different sensory modalities as sensory fragmentation exists in vision but also for instance in haptic tasks (hence in the somatosensory system). Second, it is a powerful mean to distinguish between the different computational steps of the visual motion system. Object motion information is extracted

along a cascade of feedforward cortical areas, where area V1 extracts local motion information that is integrated in extra-striate middle temporal (MT) and medial superior temporal (MST) areas.

The middle temporal (MT) and medial superior temporal (MST) areas in the superior temporal sulcus (STS) process visual motion and oculomotor signals driving pursuit (see [27] for a review) and are therefore key elements in motion extrapolation. Early physiological studies in macaque monkey identified area MT as a specialized module for visual motion processing [2, 15]. This involves extracting the speed and direction of the moving object. MT neurons respond selectively to visual motion and tuned for local speed and direction of luminance features moving in their receptive fields [37]. Pack and Born [46] have shown that the temporal dynamics of motion integration can be seen from time-varying firing rates. They showed that neuronal responses quickly progress from local to global motion direction in about 100ms suggesting that such mechanisms are dynamical and progressive. These results pinpoint the key role of MT neurons in local motion analysis as well as global motion integration. However, these neurons respond only when the retinal image motion is present while MST neurons maintain their firing activity when there is no retinal image motion as during a transient image occlusion [42] or during tracking imaginary target covering the visual field outside of the receptive field currently recorded [26]. Similar sustained activity during target occlusion has been found in monkey posterior parietal cortex, and it is linked to an image motion prior to target disappearance [4]. In another study [56] have stimulated the retina of tiger salamander with a periodically flashing stimulus and have found various firing patterns when a flash is omitted. This sustained activity is known as “omitted stimulus response” (OSR) and is explained by a model based on tunable oscillators which extrapolate the response to the periodic stimulation even at times matched to the missing stimulus. OSR has also been reported in the flicker electroretinogram (ERG) of the human cone system [38].

What is the link between behavioral and neuronal signatures of motion extrapolation? Visual motion information is primarily used for gaze stabilization [27, 28, 35] and sensorimotor transformation underlying smooth pursuit eye movements [30]. The fact that sustained activity in area MST was

seen during transient occlusion of a moving target supports the notion that the two phenomena are closely related [42]. On the other hand, since motion extrapolation is also seen in lower level neuronal structures, such as the retina, this calls for a more generic computational framework. Since motion extrapolation is implemented at the scale of a single cortical area, this would suggest that such a mechanism would be implemented by a finely structured set of diffusive mechanisms. A potential candidate is naturally the dense network of lateral interactions as found in sub-cortical and cortical structures involved in sensory processing as well as sensorimotor control. However, direct evidence for such neural mechanisms is still lacking. Before proposing a solution using motion-based prediction, we will first review some existing experimental and theoretical evidences.

## 1.2 Different types of motion extrapolation

A classical way of studying motion extrapolation is by presenting a moving target that travels behind an occluder for a short period of time. A seminal study used timing estimation by asking participants to make a button press response at the time they judge the occluded target to have reached a particular point [54]. Since then, this phenomenon has been studied at various levels (behavioral or neural), across species and modalities. For instance, motion extrapolation has been under study by focusing on various specific questions in physiology or behavior. In physiology, motion extrapolation was shown to occur in retina [19] and [56] or in higher cortical areas [4]. Behaviorally, motion extrapolation was studied in the context of target catching [43], apparent motion [23] and trajectory extrapolation for occluded or disappeared stimuli [34], perceptual extrapolation of blurred visual target [17], in audio visual targets [65], role of motion extrapolation in control of eye movements [33], blurred targets and behavior humans [17]. Motion extrapolation can be carried out for lateral motion, with the target moving across the fronto-parallel plane, or for approach motion, when the object moves towards the observer [14]. Herein, we investigate visual, lateral motion extrapolation as a generic paradigm to challenge prediction algorithms.

A tightly coupled phenomenon is motion inertia,

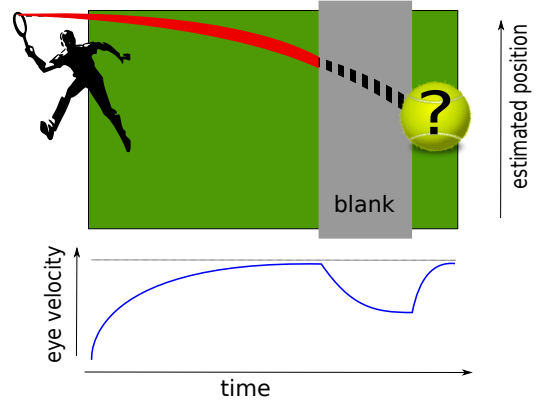


Figure 1: The problem of fragmented trajectories and motion extrapolation. As an object moves in visual space (as represented here for commodity by the red trajectory of a tennis ball in a space-time diagram with an one-dimensional space on the vertical axis), the sensory flux may be interrupted by a sudden and transient blank (as denoted by the vertical, gray area and the dashed trajectory). How can the instantaneous position of the dot be estimated at the time of reappearance? This mechanism is the basis of *motion extrapolation* and is rooted on the prior knowledge on the coherency of trajectories in natural images. We show below the typical eye velocity profile that is observed during Smooth Pursuit Eye Movements (SPEM) as a prototypical sensory response. It consists of three phases: first, a convergence of the eye velocity toward the physical speed, second, a drop of velocity during the blank and finally, a sudden catch-up of speed at reappearance [7].

which might be regarded as the perceptual equivalent of motion extrapolation for object identification. To put motion inertia in evidence, it has been shown in experiments that when one object moves and breaks into two trajectories, the trajectory that tends to be perceived as pursuing its motion is the one corresponding to the least perturbation (acceleration or curvature). Equivalently, if a moving object has been presented before, there is a strong perceptual tendency to continue seeing it in previous direction [50]. These findings also imply that the interactions between pairs of dots seen in sequence is affected by the history of their interactions, suggesting that probably the neurons responding to motion are directionally coupled in a feed forward way which facilitates the perception of unidirectional movement [3]. Assuming the existence of such a strategy, it needs to be clarified how such rules may be related to the spread of neural activity and how a neural system uses accumulated information from the trajectory of moving object in order to favor the detection of an unique, global motion. This was studied by looking at how people may extrapolate motion on a straight line [47]. One can interpret that in a Bayesian way: as a prior, motion is temporally coherent, and motion inertia is a built in strategy of the visual system to respect this prior. As such motion inertia and motion extrapolation certainly share some common mechanisms, though here, we focus on the later.

### 1.3 Experimental evidence of motion extrapolation

The neural systems controlling smooth pursuit eye movements (SPEMs) are likely to be critically dependent upon motion extrapolation, in close synergy with saccades [66]). Several studies have shown that blanking a small moving target results in a very typical temporal profile of eye velocity (see Figure 1). Eckmiller and Mackeben [16] investigated monkey smooth pursuit behavior when a moving target briefly disappeared and then reappeared. They found that monkeys are able to continue pursuing when the target disappears for up to 800 ms. Using a similar paradigm, Becker and Fuchs [7] showed that humans maintain smooth pursuit up to 4 s after the disappearance of the target. They found that the eye velocity rapidly decreased about 200 ms after target disappearance. This deceleration phase

lasted for about 280 ms and then the eye velocity stabilized at approximately 40 to 60% of the normal pursuit velocity. To develop an eye velocity related to the velocity of the target that preceded the extinction, the subjects needed to see the motion for at least 300 ms. Becker and Fuchs [7] referred to this phenomenon as predictive pursuit. This mechanism can also be at play during other open-loop responses such as anticipatory smooth tracking of a highly predictable target motion [5]. There is an ongoing debate of whether the origin of motion extrapolation is within the oculomotor control system [33] or rather occurs at the sensory level. Using event related potentials, Makin, Poliakoff, and El-Deredy [34] have suggested on electrophysiological grounds that both systems may be contributing. To tease apart the relative contribution of retinal (i.e. image-driven) and extra-retinal (i.e. eye movements-driven) in the phenomenon of motion extrapolation is out of the scope of the present study and we will restrict ourselves herein to the open-loop, image-driven pursuit behavior.

Motion extrapolation seems to be a highly adaptable mechanism. We have already suggested that such behavior may be related to the regularities observed in natural scenes. One may then wonder how this may be affected by experimental conditions such as learning or reinforcement [32]. Becker and Fuchs [7] had already examined the effect of training on predictive pursuit and reported only a modest change, indicating that such a response could be under adaptive control. Using an operant conditioning procedure, Madelain and Krauzlis [32] found that human subjects instructed to track a small spot, tend to follow it even during the absence of sensory input. The speed decreased however to a smaller plateau value and subject often performed a catch-up saccade to track the object again. Crucially, their performance increased across sessions and subjects could pursue dots up to 4 seconds after the onset of a blank after intensive learning. One important aspect for prediction to occur is that target trajectories must be regular and clear. In another study, [10] investigated used the aperture problem to probe the impact of visual motion information at target reappearance. A moving tilted bar produces a small direction bias at pursuit initiation in the direction orthogonal to the bar's orientation. They found a significant, albeit much smaller bias at target reappearance, as compared to pursuit ini-

tiation. Moreover, they put in evidence a strong difference in the amplitude of such a bias, depending on whether the blanking onset occurred in either the open- or closed-loop phase of pursuit. The tracking direction bias introduced by the aperture problem was significantly less in the late phase, suggesting that the oculo-motor system would switch from a preference for the sensory input (early phase) to an internal (motor-based) signal in the late phase. All these results raise the question of how we can model the different facets of motion extrapolation in a common framework.

#### 1.4 Existing theories on motion extrapolation

There are a variety of models proposing different mechanisms underlying motion extrapolation. A first class of models are built upon control-like models of the visuo-oculomotor system [53]. Such models were refined to specifically address the problem of motion extrapolation [13] by including additional layers in a cascade model from Goldreich, Krauzlis, and Lisberger [18]. These models may be subdivided into those where the predicted signal is based of some motor command [9] and those that specifically use the adaptation of an internal model [32]. Still, while these different behavioral models can fit some data very nicely, they lack a global explanation of the mechanisms underlying motion extrapolation.

Most of these models share a common mechanism: during blanking, information is inferred from past information using a smoothness constraint on possible trajectories. This is well formulated by smoothing the inferred velocity in control models with an internal positive feedback [29, 52, 53]. An engineering answer for such an adaptive system is a Kalman filter. It involves projecting the current estimate of the system based on the prior knowledge and correcting the predictions based on the measurement. A mix of measurement and prediction are used to estimate the current state based on their reliability reflected from their variances. Studies investigating sensory-motor transformation already suggest for a mix of measurement based signal and an internal signal based on reliability extracted from their respective uncertainties for an optimal performance in a motor task [8]. Similarly, this may be expressed in as a Kalman filter, that is in a generic Bayesian framework with a clear hypothesis [64].

Following the idea of Kalman filter and extending the work of Montagnini et al. [39], Bogadhi, Montagnini, and Masson [10] proposed a hierarchical recurrent Bayesian framework to understand both motion integration as observed in smooth pursuit and also the predictive nature of pursuit. Probabilistic inference has been successful in explaining motion perception to a variety of stimuli [63]. They are somewhat similar to some of the engineering models proposed earlier [44] but allow for a more explicit formulation of the underlying hypothesis. Such a framework accommodates uncertainty in the motion information in the measurement likelihoods [22, 60, 63] and also expectation can be represented through the prior which can alter motion perception [58]. Representing uncertainty in the measurements and prior expectation gives a simple, yet powerful framework to investigate predictive behavior of the system under investigation possibly to optimally adapt to changes in the measurements. As shown by Wuerger et al. [65] in a temporal localization task, the bias and variability show similar patterns for motion defined by vision, audition or both. Such optimal integration is consistent with a probabilistic representation of motion. The framework implements Bayesian estimation utilizing motion measurements and motion prediction. Measurements of observed input are interpreted probabilistically by a likelihood function. To detect straight trajectories with constant velocity, input motion can be temporally grouped and expressed in terms of a Bayesian generalization of a Kalman filtering [64], as standard Kalman filter models are not able to account for psychophysical data. A neural network model of described probabilistic framework shares interesting similarities with known properties of visual cortex and qualitatively accounts for psychophysical experiments on motion occluders and motion outliers. The approach from Bogadhi, Montagnini, and Masson [10] allows for a mix of prediction and measurement based on their reliability, as measured from their respective variances. The combined estimate is used to drive the pursuit response. The hierarchical framework allows to investigate the adaptive behavior of pursuit as well as the role of prediction on motion integration as observed in pursuit responses. However, this model may still be seen as an incremental refinement of previous results and does not yield a generic account on the motion extrapolation mechanism.

As we have seen, most theoretical efforts to study motion extrapolation is based on temporal coherency of motion. This assumption, as understood in a Bayesian framework, may be represented by defining a prior in the probabilistic representation of motion. This will then be integrated in the dynamical motion integration process: In a probabilistic representation of motion, posterior estimation of motion is the product of this prior and current sensory evidence (likelihood). An important question is therefore to know how to define this prior function.

### 1.5 Motion extrapolation and motion-based prediction

Yuille and Grzywacz [67] have shown that the efficiency of motion integration was highly dependent on the smoothness of the trajectory of the stimulus. Behavioral data showed that humans can detect a target dot moving in a smooth trajectory embedded in randomly moving dots, while the target dot is not distinguishable from noise in each frame separately. This challenging detection task is called outlier detection and might be explained by a network of interconnected motion detectors [62]. In such a network, every stimulated local motion detector sends a facilitatory signal to adjacent units. These are in turn stimulated and this sequence goes on, ultimately implementing a direction selective spatiotemporal integration. Signals from local motion detectors are made coherent in space and time and lower the threshold for detecting stimuli moving in smooth versus segmented trajectories [21]. In the outlier detection case, distractor dots do not move coherently enough to accumulate information while for the target dot, precision increases gradually and as a consequence, the accuracy of velocity estimation is improved. During occlusion of target motion, that is without likelihood measurements, velocity estimation is degraded and probabilities are diffused in space and time. However, the model may still have enough momentum or motion inertia to propagate estimations of target dot's position. This process will break down if the occluder gets too long but the motion inertia effect of target motion on distractors is visible [62]. As a consequence, an important aspect of this prior is a *motion-based prediction*, that is, including both the position and velocity from the trajectory of motion.

Such a prior on the temporal coherency of mo-

tion can be defined in a probabilistic framework. This was formulated theoretically by [11] but their neural network implementation lacked the precision needed to work on realistic input sequences. In our earlier work [49], we implemented efficiently such a prior to investigate different aspects of spatiotemporal motion integration. Particularly, this model focused on the aperture problem and proposed that motion-based predictive coding is sufficient to infer global motion from all local ambiguous signals. The aperture problem is a challenging problem to study integration of local motion information [12, 31, 46]. The model proposed that instead of specific mechanisms such as line-endings detectors, the gradual spatio-temporal integration of motion. It accounts for the properties of physiological and behavioral responses to the aperture problem. First, the temporal dynamics of the solution to the aperture problem and its dependence on several properties of input such as contrast or bar length can be represented. Second, end stop cells emerge from the dynamics of the model instead of having *ad hoc* rules such as line-ending detectors.

The hypothesis of independence of motion signals in neighboring parts of visual space results in the failure of feedforward models in accounting for temporal dynamic of global motion integration. In those models, local measurement of global motion is the same everywhere independent of position. In motion-based prediction, the retinotopic position of motion is an essential piece of information to be represented. By explicitly including the interdependence of local motion signals between neighboring times and positions knowing the current speed along a smooth trajectory, incoherent features are explained away, while coherent information is progressively integrated. This context-dependent, anisotropic diffusion in the probabilistic representation of motion also results in the formation of a tracking behavior favoring temporally coherent features. Herein, we will challenge such a model to account for the different properties of motion extrapolation.

### 1.6 Objectives and outline

This paper has been prepared in following order: In Section 2 we develop the same probabilistic modeling framework as the one proposed for the solution to the aperture problem [49]. Moreover, we include

details on the structure and implementation of the model but also details on the experimental and numerical aspects of the model. Then, we report in Section 3 results from experiments where we studied motion extrapolation under three different conditions for a horizontally moving dot: moving in a blanked trajectory, moving in presence of high background noise and moving in a blanked trajectory with high background noise. In the first condition, extrapolation of motion information during a blank has been studied compared to a control stimulus without blank. To stress on the role of prediction in motion extrapolation, we have done all experiments under three configuration of the model which correspond to motion estimations with and without prediction in position or velocity of stimulus. In the second condition, we have surveyed motion extrapolation by looking at states of motion tracking and its stability. In the last condition, we predict that motion extrapolation is dependent on noise and propose a behavioral experiment to test this prediction.

Finally in the discussion (Section 4), we will interpret these results in the light of current knowledge on probabilistic inference and dynamical systems and we will discuss the limitations of the current study along with suggestions for future work.

## 2 Model & methods

### 2.1 Probabilistic detection of motion

First, we define a generic probabilistic framework for studying motion integration. The translation of an object in the planar visual space at a given time is fully given by the probability distribution of its position and velocity, that is, as a distribution of our value of belief among a set of possible positions and velocities. It is usual to define motion probability at any given location. If one particular velocity is certain, its probability becomes 1 while other probabilities are 0. The more the measurement is uncertain (for instance when increasing noise), the more the distribution of probabilities will be spread around this peak. This type of representation can be successfully used to solve a large range of problems related to visual motion detection. These problems belong to the more general framework of the optimal detection of a signal perturbed by different sources

of noise and ambiguity.

In such a framework, Bayesian models make explicit the optimal integration of sensory information with prior information. These models may be decomposed in three stages. First, one defines likelihoods as a measure of belief knowing the sensory data. This likelihood is based on the definition of a generative model. Second, any prior distribution, that is, any information on the data that is known before observing it, may be combined to the likelihood distribution to compute a posterior probability using Bayes' rule. The prior defines generic knowledge on the generative model over a set of inputs, such as regularities observed in the statistics of natural images or behaviorally relevant motions. Finally, a decision can be made by optimizing a behavioral cost dependent on this posterior probability. An often used choice is to choose the belief that corresponds to the maximum a posteriori probability.

### 2.2 Luminance-based detection of motion

Such a Bayesian scheme can be applied to motion detection using a generative model of the luminance profile in the image [63]. This is first based on the luminance conservation equation. Knowing the velocity  $\vec{V} = (u, v)$ , we can assume that luminance is approximately conserved along this direction, that is, that after a small lapse  $dt$ :

$$I_{t+dt}(x + u \cdot dt, y + v \cdot dt) = I_t(x, y) + \nu_I \quad (1)$$

where we define luminance at time  $t$  by  $I_t(x, y)$  as a function of position  $x, y$  and  $\nu_I$  is the observation noise. This noise is assumed to be Gaussian with zero mean and variance  $\sigma_I^2/dt$ , that is,  $\nu_I \propto \mathcal{N}(I; 0, \sigma_I^2/dt)$ . Note that for convenience we scaled variance by  $dt$  such that the variance  $\sigma_I^2$  can be represented per unit of time, independently of the time step  $dt$ . Using the Laplacian approximation, one can derive the likelihood probability distribution  $p(I_t(x, y) | \vec{V})$  as a Gaussian distribution. In such a representation, precision is finer for a lower variance. Indeed, it is easy to show that the logarithm of  $p(I_t(x, y) | \vec{V})$  is proportional to the output of a correlation-based elementary motion sensors or equivalently to a motion-energy detector [1]. Interestingly, lower contrast motion



results in wider distributions of likelihood and thus posterior  $p(\vec{V}|I_t(x, y))$ . Therefore, contrast dynamics for a wide variety of simple motion stimuli is determined by the shape of the probability distribution (i.e. Gaussian-like distributions) and the ratio between variances of likelihood and prior distributions as was validated experimentally on behavioral data [6].

The generative model explicitly assumes a translational motion  $\vec{V}$  over the observation aperture, such as the receptive field of a motion-sensitive cell. Usually, a distributed set  $\vec{V}_t(x, y)$  of motion estimations at time  $t$  over fixed positions  $x, y$  in the visual field gives a fair approximation of a generic, complex motion that can be represented in a retinotopic map such as areas V1 and MT. This provides a field of probabilistic motion measures  $p(I_t(x, y)|\vec{V}_t(x, y))$ . To generate a global read-out from this local information, we may integrate these local probabilities over the whole visual field. Assuming independence of the local information as in [63], spatio-temporal integration is modeled at time  $T$  by Equation (1) and

$$p(\vec{V}|I_{0:T}) \propto \prod_{x,y,0 \leq t \leq T} p(I_t(x, y)|\vec{V}(x, y))p(\vec{V}) \quad (2)$$

where we write as  $I_{0:t}$  the information on luminance from time 0 to  $t$ . Such models of spatio-temporal integration can account for several nonlinear properties of motion integration such as monotonic spatial summation and contrast gain control and are successful in explaining a wide range of neurophysiological and behavioral data.

### 2.3 Motion-based predictive coding

The independence hypothesis set above formally states that the local measurement of global motion is the same everywhere, independently of the position of different motion parts. In fact, the independence hypothesis assumes that if local motion signals would be randomly shuffled in position, they would still yield the same global motion output (e.g. [40]). As shown by Watamaniuk, McKee, and Grzywacz [62], this hypothesis is particularly at stake for motions along coherent trajectories: motion as a whole is more than the sum of its parts. A first assumption is that the retinotopic position of motion is an essential piece of information to be represented. In particular, in order to achieve

fine-grained predictions, it is essential to consider that the spatial position of motion  $x, y$ , instead of being a given parameter (classically, a value on a grid), is an additional random variable for representing motion along with  $\vec{V}$ . Compared to the representation  $p(\vec{V}(x, y)|I)$  used in previous studies [11, 63], the probability distribution  $p(x, y, \vec{V}|I)$  more completely describes motion by explicitly representing its spatial position jointly with its velocity. Indeed, it is more generic as it is possible to represent any distribution  $p(\vec{V}(x, y)|I)$  with a distribution  $p(x, y, \vec{V}|I)$ , while the reverse is not true without knowing the spatial distribution of the position of motion  $p(x, y|I)$ . By doing so, we introduce an explicit representation of the segmentation of motion in visual space as an essential ingredient in motion-based predictive coding.

Here, we explore the hypothesis that we may take into account most dependence of local motion signals between neighboring times and positions by implementing a predictive dependence of successive measurements of motion along a smooth trajectory. In fact, we know *a priori* that natural scenes are predictable due to both rigidity and inertia of physical objects. Due to the projection of their motion in visual space, visual objects preferentially follow smooth trajectories. We may implement this constraint into a generative model by using the transport equation on the motion itself. Assuming for simplicity that the sensory representation is updated at discrete, regularly spaced times, then, at time  $t$ , during the small lapse  $dt$ , motion is approximately translated with respect to its velocity :

$$\begin{aligned} x_{t+dt} &= x_t + u_t \cdot dt + \nu_x \\ y_{t+dt} &= y_t + v_t \cdot dt + \nu_y \end{aligned} \quad (3)$$

$$\begin{aligned} u_{t+dt} &= \gamma \cdot u_t + \nu_u \\ v_{t+dt} &= \gamma \cdot v_t + \nu_v \end{aligned} \quad (4)$$

with

$$\nu_x, \nu_y \propto \mathcal{N}(x, y; 0, D_X \cdot dt) \quad (5)$$

$$\nu_u, \nu_v \propto \mathcal{N}(u, v; 0, (\sigma_p^{-2} + D_V^{-1})^{-1} \cdot dt) \quad (6)$$

Where  $\nu_x, \nu_y, \nu_u$  and  $\nu_v$  are random variables that blur position and velocity at each time step. These are centered Gaussians defined by their variances in position space by  $D_X \cdot dt$  and in velocity space defined as in [63] by  $(\sigma_p^{-2} + D_V^{-1})^{-1} \cdot dt$  (where

the standard deviation of the prior is defined as  $\sigma_p$ ). Note that for convenience we scaled variance by  $dt$  such that the diffusion coefficients can be represented per unit of time, independently of the time step  $dt$ . Here,  $\gamma = (1 + \frac{D_V^2}{\sigma_p^2})^{-1}$  is the damping factor introduced by the prior and  $\gamma \approx 1$  for a high value of  $\sigma_p$ . The update rule (see [48] for a derivation) assumes independence of the prior on slow speeds with respect to predictive prior on smooth trajectories. Equations (3)-(4) describe the internal model of motion, while equations (5)-(6) give a model for the noise in this model.

As a consequence, the estimated positions and velocities are slightly randomized at each frame. We controlled that this procedure led to similar results as [63] (see also [48] for an analytical solution). Note that we return to the usual formulation described in [49] when setting  $\sigma_p$  to a high value (100 was sufficient in our numerical simulations). We define  $D_X$  and  $D_V$  as the diffusion coefficients of position and velocity, respectively in a generative model of diffusion which are scaled to be represented in units of time, independently of  $dt$ . In fact, it is important to properly tune  $D_X$  and  $D_V$  since they explicitly quantify the precision of the prediction.

We may now use this generative model to integrate motion information using a Markov random chain on joint random variables  $z_t = x_t, y_t, u_t, v_t$ :

$$p(z_{t+dt}|I_{0:t}) = \int dz_t p(z_{t+dt}|z_t) \cdot p(z_t|I_{0:t}) \quad (7)$$

$$p(z_{t+dt}|I_{0:t+dt}) = p(I_{t+dt}|z_{t+dt}) \cdot p(z_{t+dt}|I_{0:t}) / p(I_{t+dt}|I_{0:t}) \quad (8)$$

To implement this recursion, we first compute  $p(I_t|z_t)$  from the observation model (Equation (1)). The predictive prior probability  $p(z_t|z_{t-dt})$  is defined by the generative model in Equation (3) and (4). Note that prediction (Equation (7)) always increases the variance by “diffusing” information. On the other hand, during estimation (Equation (8)), coherent data increases precision of the estimation while incoherent data increases the variance. This balance between diffusion and reaction will be the most important factor for the convergence of the dynamical system. Overall, these master equations, along with the definition of the prior transition  $p(z_t|z_{t-dt})$ , define our model as a dynamical system with a simple global architecture but yet with complex recurrent loops.

Unfortunately, the dimensionality of the probabilistic representation makes it impossible to implement a realistic simulation of the full dynamical system on classical computer hardware. In fact, even with a moderate quantization of the relevant representation spaces, computing integrals over hidden variables in the filtering and prediction equations (respectively Equations (7) and (8)) leads to a combinatorial explosion of parameters that is intractable with the limited memory of current sequential computers. Alternatively, if we assume that all probability distributions are Gaussian, this formulation is equivalent to Kalman filtering on joint variables. Such an implementation may be achieved using for instance a neuromorphic approximation of the equations mentioned above [11]. Indeed, one may assume that master equations are implemented by a finely tuned network of lateral and feed-back interactions. One advantage of this recursive definition in the master equations is that it gives a simple framework for the implementation of association fields. However, this implementation has the consequence of blurring predictions. To describe some nonlinear aspects of motion integration we have used particle filtering in our previous work [49] and here we will use it to tackle the motion extrapolation problem.

## 2.4 Numerical simulations

Master equations can be approximated using Sequential Monte Carlo (SMC) [49]. The SMC algorithm itself is controlled by only two parameters. The first one is the number of particles  $N$  which tunes the algorithmic complexity of the representation. In general,  $N$  should be large enough and an order of magnitude of  $N \approx 2^{10}$  was always sufficient in our simulations. In the experimental settings that defined here (moving dots), the complexity of the scene is well controlled and remains low. Control experiments have tested the behavior for several numbers of particles (from  $2^5$  to  $2^{16}$ ). We found that, except with  $N$  smaller than 100, results were always very similar. However, we kept  $N$  to this quite high value to maintain the generality of the results in the perspective of further extensions of the model. The other parameter is the threshold above which particles are resampled. We found that this parameter had little qualitative influence providing that its value is large enough to avoid staying in a

local minima. Typically, a resampling threshold of 20% was sufficient.

Once the parameters of the SMC were fixed, the only free parameters of the system were the variances used to define the likelihood and the noise model. The likelihood of sensory motion was computed using Equation (1) using the same method as Weiss, Simoncelli, and Adelson [63]. We defined space and time as the regular grid on the toroidal space to avoid border effects. Next, visual inputs were  $128 \times 128$  grayscale images on 150 frames. All dimensions were set in arbitrary units and we defined speed such that  $V = 1$  corresponds in toroidal space to the velocity of one spatial period within one temporal period that we defined arbitrarily to 100 ms biological time. Raw images were preprocessed (whitening, normalization) and we computed at each processing step the likelihood locally at each point of the particle set. This computation was dependent only upon image contrast and the width of the receptive field over which likelihood was integrated. We tested different parameters values that resulted in different motion direction or spatio-temporal resolution selectivities. For instance, a larger receptive field size gave a better estimate of velocity but a poorer precision for position, and reciprocally. Therefore, we set the receptive fields size to a value yielding to a good trade-off between precision and localization (that is 5% of the image's width in our simulations). Similarly, the contrast of the likelihood was tuned to match the average noise value in the set of images. Once fixed, these two values were kept constant across all simulations. All simulations were performed using python with modules numpy [45] and scipy (respectively version 2.6, 1.5.1 and 0.8.0) on a cluster of linux nodes. Visualization was performed using matplotlib [24]. All scripts are available upon request from the corresponding author.

## 2.5 Experimental procedure

All of our experimental conditions include a single dot moving horizontally at constant speed, albeit in different contexts. The dot could be blanked transiently during its displacement (blank condition), presented with a high level of noise (noise condition) or blanked with a high level of background noise (blank + noise condition). In the two blanking conditions, the target disappeared for a duration of 28%

of its whole spatial period. It then reappeared at the location it should be with a continuous motion and followed the same trajectory some the remaining lapse of time. Note that as  $V = 1$ , the blank duration is also 28% of a temporal period. We controlled that during the blank, the local motion energy gave an uniform likelihood for all velocities, as expected. To investigate the effects of background noise, we added to the signal one of 20 linearly increasing values of standard deviation of independent, identically distributed gaussian noise (from 0.01 to 0.20 relative to a peak signal value of 1). This noise could be added to a normal trajectory of a partially blanked one.

To quantify the efficiency of motion detection, we computed different statistics. A first representation looked at the temporal dynamics of motion distribution. For this we estimated the histograms of estimated position and velocity signals along both horizontal and vertical axis. When adding noise to the motion stimulus, we also measured contrast gain in order to quantify the tracking performance by plotting the estimated values of eye velocity as a function of contrast. All simulations were repeated for 20 trials and below we plot average values across trials. Error bars are their standard deviation across trials.

One main objective of the study was to compare our motion-based prediction (MBP) model with different controls. We defined two models as limiting cases of the MBP model. A first control model is obtained when motion estimation is made dependent upon the velocity prediction but not upon the position prediction. We called this model PV, and it corresponds to the model of Weiss, Simoncelli, and Adelson [63] as defined by Equation 2. Then, the motion-based prediction was similar but we switched off the prediction in position by setting a high value for  $D_X$  with respect to the spatial period (see equations (3)-(5)). Typically, a value of 100 was sufficient. Equations of motion's position in the generative model then were simplified to  $x_t, y_t \propto \mathcal{N}(x, y; 0, D_X \cdot t)$ .

A second control model was by making motion estimation dependent upon the position prediction but not the velocity prediction. We called it PX and set an high value for  $D_V$  with respect to the typical physical speed of stimuli (see equations (4) and (6)). Typically a value of 100 was sufficient in our simulations. The equations for the velocity

in the generative model can then be simplified to  $u_t, v_t \propto \mathcal{N}(u, v; 0, \sigma_p^2)$ . This PX model will perform an isotropic diffusion of information since it lacks any prediction on the velocity component of motion. This is similar to the isotropic diffusion model of Tlapale et al [61] when removing the contour-based information.

### 3 Results

Our goal was to investigate the role of prediction in motion extrapolation in the presence of different sources of uncertainty such as a transient disappearance of the target and high background noise. We tested the dependency of our model upon prediction versus current sensory input. Motion extrapolation for a single dot moving at constant velocity was tested in three different conditions: a transient blanking trajectory in absence of background noise; a complete trajectory with different background noise levels and a transient blanking trajectory with background noise.

#### 3.1 Extrapolation of motion information in a blanked trajectory

We first used as an input the movie of single dot translating along a straight trajectory and that is transiently blanked after a short period of visible displacement. This situation is similar to those used in physiological [4], behavioral [10] and theoretical previous studies [11]. By doing so, we can challenge the dynamics of information being accumulated along the occluded trajectory that is, in absence of sensory input. We measured the estimated positions and velocities of dot motion at time windows located just before, during and after the blank.

In Figure 2, we plot the histogram of the estimated positions and velocities obtained with the three different models: PX (motion estimation is only predictive in position, not velocity), PV (motion estimation is only predictive in velocity, not position) and MBP that is the full motion-based model where there is a predictive motion estimation both in position and velocity. Remind that PX and PV were obtained simply by choosing high values of  $D_V$  and  $D_X$ , respectively.

In Figure 2, for the earliest frames, velocity histograms first spread over a larger area but pro-

gressively fit into a narrow band centered on the physical velocity ( $u = 1$  and  $v = 0$ , see rightmost columns). This strongly suggests a convergence of the estimated motion direction towards the veridical movement of the stimulus. During the blanking period marked by vertical white dashed lines, the histograms illustrate different states. In the PX control model (upper row), velocity estimations ( $u$  and  $v$ ) are largely scattered around zero, favoring the occurrence of slow speeds. Because of the measurements, the estimations still became narrower and centered on the physical velocity of stimulus both before and after blanking. During blanking, estimated positions diffused in an isotropic manner (two leftmost columns). With both PV and MBP model configurations, the dynamics of velocity estimations paused during blanking and distributions were maintained around the last estimated values computed right before target disappearance. At stimulus reappearance, the distributions immediately resume their convergence. The estimated positions ( $x$  and  $y$ ) computed with the full MBP model exhibited a dynamics similar to velocity estimations suggesting the existence of an internal model that updates the estimations with a slow diffusion. By contrast, in the PV control model, there is no prediction to update the next stimulus position and therefore estimation histograms spread across all possible positions (see second row, left columns). This difference between PV and MBP model performance is summarized in Figure 3. We plot the estimated velocity obtained with each model (mean and standard deviation) together with the control condition where the dot was continuously visible. Clearly, when the stimulus reappeared after blanking, motion-based predictive estimation tend to converge immediately back to the control speed with a quick catch up. Such dynamics was more sluggish with the PV model (blue curve): motion integration did resume but at roughly the same slope as observed at the onset of the blank. Note that we did not plot the performance of the PX model in Figure 3, because of the very large variability of estimated velocity observed across trials (see Figure 2). Moreover, the rather small difference observed between PV and MBP models is due to the rather simplistic dot motions used in the present study. As explained above, the sensory layer of both models is made of a bank of motion energy filters which are highly efficient in locally detecting straight dot translations. Choosing

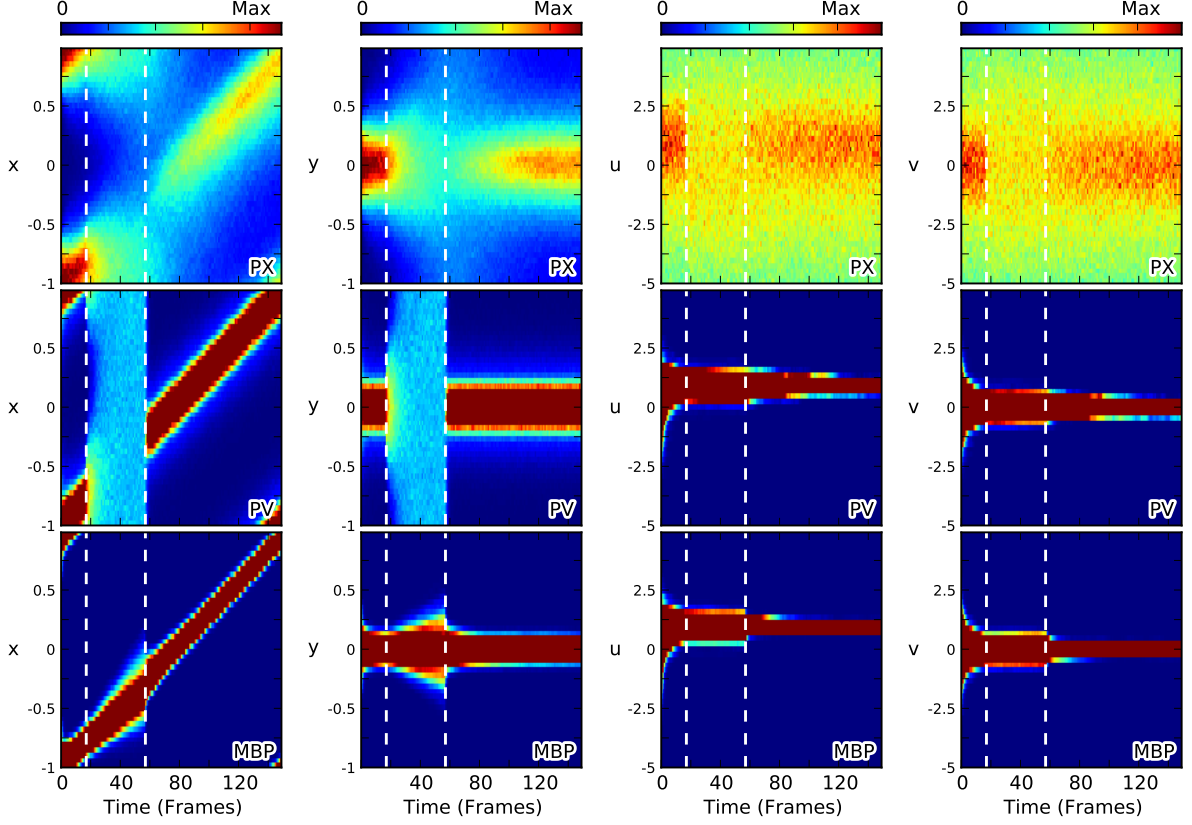


Figure 2: Histograms of estimated motion for a horizontally moving dot with a blanked trajectory under three predictive configurations of the model. Blanking period is indicated with dashed white lines and each row represents full motion estimation under the configuration denoted by the inner title. Each plot illustrates the probability distribution function of a relevant variable (vertical axis) with respect to time (horizontal axis) as in Figure 1. The color bar on top indicates the value of probability as it is estimated for each frame (one column in each image). In each configuration, the two left columns correspond to estimated positions ( $x$  and  $y$ ) while the right columns represent estimated velocities ( $u$  and  $v$ ). At the earliest frames, for all configurations, estimated variables are scattered in a rather wide area but then gradually converge to the veridical solution  $(x, y, u, v) = (1, 0, 1, 0)$ . **(First row)** PX configuration: motion estimation is only predictive in position of motion and not in velocity. **(Second row)** PV configuration: Motion estimation is only predictive in velocity of motion and not in position. **(Third row)** MBP configuration: in this configuration, motion estimation is predictive in both position and velocity of motion and predictive information is transported anisotropically using the velocity information (compare variable  $x$  with configurations PX and PV).

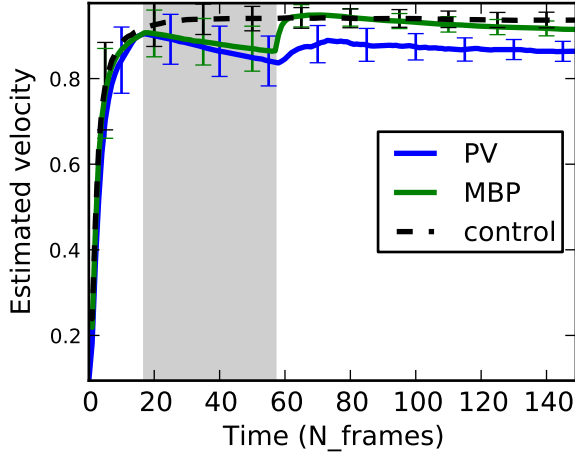


Figure 3: Estimated velocity in PV and MBP configurations of Figure (2). Estimation is measured as the response of the model to a horizontally moving dot with  $u = 1$  and a short blank in the trajectory. Blanking period has been shown with shaded area. The trace in black dashed lines represent the control condition in which trajectory of stimulus includes no blank. Error bars show standard deviation of error over 20 trials. Note the quick catch up after reappearance of stimulus in MBP configuration.

a high value for  $D_X$  and  $D_V$  in PV and PX models is then equivalent to switching to basic sensory level without prediction in position and velocity respectively. That is to say, our working hypothesis imposes a large scale coherency constraint on stimulus trajectory as reflected in the range of values for  $D_X$  and for  $D_V$ . Still, we can observe the temporal dynamics of motion estimation as already shown with more ambiguous inputs [49].

### 3.2 Predictive tracking in the presence of varying noise levels

In the next step, we tested the robustness of our model when using more realistic conditions such as low contrast (or low signal-to-noise ratio) inputs. This approach is similar to the previous psychophysical work on temporal coherency and predictability of motion [62]. Below, we report the performance of two model configurations (PV and MBP) when gradually increasing the level of background noise

to an horizontally moving dot. We first did it for a fully visible trajectory in order to estimate the contrast (or SNR) thresholds at which the tracking states of the model change [49].

We measured the estimated velocity averaged over 20 trials when the input image was corrupted by an independent and identically distributed Gaussian noise (see Figure 4). In order to first explore the role of prediction for overcoming the distracting effect of noise, we set our motion estimation routine to the PV case to minimally rely on position predictions. To do so, we chose a  $D_X$  value high enough so that the model did favor any estimation in particular. We then repeated the same experiment but with the full MBP model. We found that including motion-based prediction led to a more precise tracking than in the PV case, at both low and high levels of noise. We found a range of contrast (or SNR) in which the MBP model was still maintaining perfect tracking while the PV model was in the no tracking state. Two particular aspects shall be noticed. First, with the PV model, increasing the noise level gradually decreased the convergence rate of the motion detection process. Second, with the full MBP model, we observed a binary response mode (i.e. the dot is either tracked or not tracked). In the tracking state, the convergence rate was found to be dependent upon the level of noise, as in the PV configuration. Increasing noise up to a certain level results in a shift of the onset of the tracking state, until the model reached the no tracking state. Our results in Figure 4 are similar to the outlier detection experiment observed in psychophysics by Watamaniuk, McKee, and Grzywacz [62] where a horizontally moving dot was surrounded by many other distractor dots with random movements. This psychophysical study showed that the temporal coherency of the target dot motion rendered it detectable with a high confidence as measured by a tenfold increase of detection threshold. Our modeling results are consistent with this behavioral observation. As a consequence, we similarly found a binary tracking response in the sense that tracking is rather good up to some noise level. Therefore we have either tracking or not tracking states. Furthermore, increasing the noise level imposes a delay on emergence of tracking state which is reflected in smooth slowing of initial raising in velocity traces.

To summarize the effects of noise, we plotted the efficiency of model with and without prediction in

position (i.e. the MBP and PV models) for a range of contrasts and fitted them with the Naka-Rushton function [41] that can adequately describe the different aspects of motion integration (e.g. [48]) (see figure 5). The gain was defined as the average estimated speed and contrast as the signal-to-noise ratio. The contrast response functions were plotted for both early and late phases (as defined in figure 2) of tracking. Remind that here the dot trajectory was fully visible. In the early phase (red curve), both models have very different best-fit contrast saturation values ( $C_{50}$  of 10.35 and 27.37, respectively) and exponents ( $n$  of 7.7 and 2.19, respectively). Interestingly, one can see in the late phase (blue curve) a global increase in contrast gain for both models, as illustrated by a leftward shift of the curves. The PV model led to a change in only the contrast saturation parameter (from  $C_{50} = 7.71$  to  $n = 3.42$ ), while the full MBP model exhibited a significant change in both in half-saturation ( $C_{50} = 7.15$ ) and slope ( $n = 20.89$ ) parameters of the contrast response function. This is characteristic of the emergence of the tracking behavior in the motion-based prediction model and complements the analysis done in [49]. As a consequence, we have demonstrated here that this model is sufficient to explain some well-known static non-linear computations such as the gain control mechanism implemented by divisive normalization [55, 57]. These are essential components of neural computations and we show here that they may emerge from a predictive coding formulation instead of an explicit descriptive mechanism. In a final experiment, we will see the potential function of this tracking response.

### 3.3 Motion extrapolation in noisy blanked trajectory

In the last series on experiments, we combined the different sources of uncertainty studied above by simulating a noisy dot moving along a partially blanked trajectory. Above, we have shown that motion extrapolation requires enough accumulation of information from the observed trajectory parts for allowing the emergence of the tracking state. Moreover, we found that there is contrast threshold for reaching this tracking state. Since our goal is to investigate the temporal evolution of the information that is accumulated from the observed trajectory, by imposing two independent sources of uncertainty

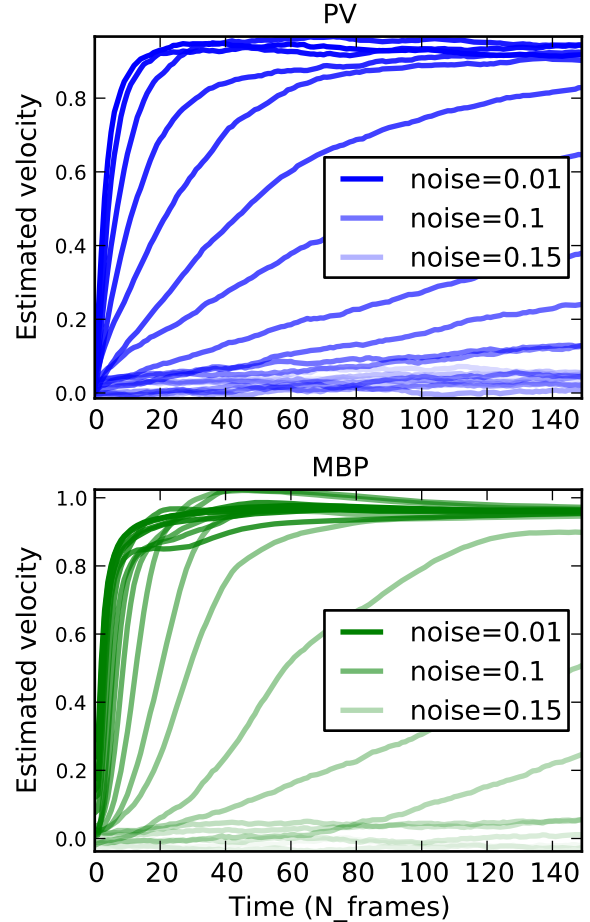


Figure 4: Estimated velocity of PV and MBP configurations averaged over 20 trials. Stimulus is a horizontally moving dot with  $u = 1$  which includes different noise values at background and no blank in trajectory. Colors from dark to light correspond to the response to the stimulus with noise levels between 0.01 to 0.2. **(Top)** Estimated velocity of **(PV)** configuration while motion estimation only benefits from predictions in velocity of stimulus. **(Bottom)** Estimated velocity of motion-based prediction **(MBP)** configuration, where estimation is predictive in both position and velocity of motion. This configuration tracks well up to approximate noise value of 0.13 and after that enters into the “no tracking” state. For PV configuration this state transfer happens at noise value 0.06. As noise increases, in both configurations we observe a slower convergence in estimated velocity and more importantly a temporal shift of the emergence of tracking.



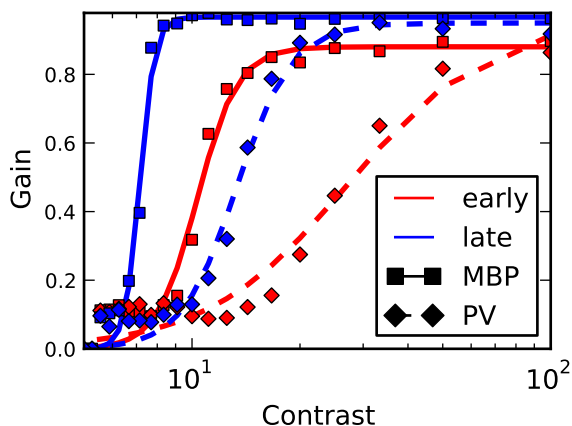


Figure 5: Response gain functions are plotted with best-fit Naka-Rushton functions [48] for both PV and MBP models. Increasing contrast produces a S-shape increase in response gain whose shape changes with both time and model configurations. Similar to the psychophysical reports by Watamaniuk, McKee, and Grzywacz [62], gain and half-saturation values increase from the early to late tracking phases. There is an increase in the slope of the contrast response curve in the late response of the motion-based prediction configuration indicating a transition from no-tracking to tracking states.

(i.e. noise and blanking) we can highlight the differences between predictive and non predictive motion estimation.

As in the previous sections, we quantified the efficiency of motion estimation by the estimated velocity of the tracking responses (see figure 6). We extend the results shown in figure 3 by now using blanked trajectory with low noise to higher levels of noise. As we mentioned before, a quick velocity catch up as illustrated in figure 3 indicates the emergence of a tracking state after stimulus reappearance. Such catch up was still visible in the presence of strong noise levels, at least up to a certain threshold. We expected a general degradation of motion extrapolation by increasing noise level and consequently a lower tracking performance, down to the no tracking state. For noises higher than contrast thresholds, no such velocity catch up was observed and the models in fact remained in the no tracking state (see figure 6). At all noise levels, incorporating position prediction as in the full MBP model revealed several differences in performance, when compared to the PV model. In particular, the MBP model was less sensitive to noise and its dynamics at intermediate signal-to-noise ratio was brisker than the PV case. Indeed, the MBP model remained able to match the stimulus trajectory after target reappearance in the presence of relatively high noise level (up to 0.11). In comparison, the PV model remained in the no tracking state for noise levels higher than 0.05.

In summary, we found that making the motion extrapolation task more difficult by mixing two uncertainty sources deteriorates the tracking response. This can be explained by an insufficiently accumulated information about dot trajectory in the noisy and blanking conditions. This is evidenced by the comparison of responses at corresponding contrasts between figure 4 and figure 6. The full MBP model takes advantage from predictions in position and velocity domains, in comparison to the PV case and can accommodate higher noise levels before losing its tracking ability. In addition, a stronger internal representation of motion is maintained during blanking in this case (see MBP estimations in Figure 2). It also more quickly converges to the true, physical motion after reappearance. These results call for similar experiments to be done psychophysically by combining these different sources of uncertainty.



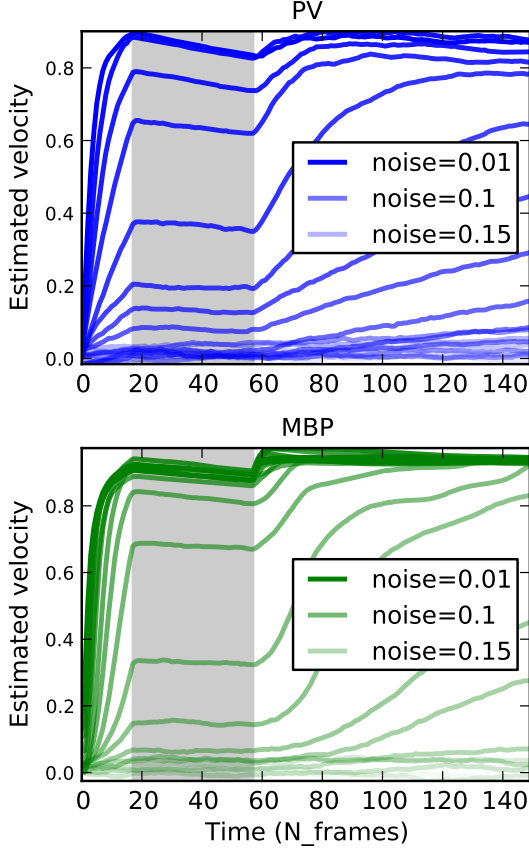


Figure 6: Motion extrapolation with sensory noise: Stimulus is a horizontally moving dot with  $u = 1$  which includes a blank as shown with shaded area. In addition there is a sensory noise and colors from dark to light correspond to noise levels increasing linearly between 0.01 to 0.2. **(Top)** Estimated velocity of model under **PV** configuration while motion estimation only benefits from predictions in velocity space. **(Bottom)** Estimated velocity of motion-based prediction (**MBP**) configuration, where estimation is predictive in both position and velocity of motion. In both configurations, increasing of noise corrupts tracking performance and after blank response converge only for noise values under a threshold and then enters to no tracking state. This threshold for PV and MBP configurations are 0.05 and 0.11 respectively. Note that the quick catch up after reappearance of stimulus never appears in PV but only in MBP in the cases in which a tracking state stabilized before blank.

## 4 Discussion

In the present study, we investigated the role of motion-based prediction [49] in motion extrapolation during target blanking, a condition frequently used in psychophysical, behavioral and neuronal studies to measure how the brain maintains an accurate representation of target motion despite large fluctuations in the input (e.g. [4, 7, 10]). Our goal was to test how the motion prediction framework described in our previous work [49] can be extended to these conditions.

First, we probed the dynamics of motion extrapolation by measuring the impact of a transient absence of the stimulus, as imposed by a short blank in trajectory of the stimulus. We found a prototypical temporal pattern characterized by a pause in the motion integration process during the blank and a quick recovery of the actual position of the dot. This model behavior was largely different when turning off the anisotropic component of motion-based prediction. In this PV incomplete model, at the end of the blank, the integration dynamics resumed at a convergence rate similar to the one observed at the initial target motion onset. This difference can be explained by the fact that the full model can maintain a nearly accurate representation of the target trajectory in both position and velocity domain. In this regard, the full MBP model is more consistent with both physiological (e.g. Assad and Maunsell [4] and Newsome and Paré [42]) and behavioral (e.g. [7, 9, 10]) observations. Interestingly, the comparison between PV, PX and the full MBP model further highlights the need of both position and velocity informations for correctly maintaining and predicting an accurate representation of target motion, an aspect that has been already introduced at the theoretical level [11, 49, 66].

An important issue was to answer to the question raised by the experimental study of Assad and Maunsell [4]. In monkeys, while MT does not represent motion during the blank, it seems that such information can be preserved in upstream cortical areas such as MST [42]. This later result is compatible with our approach, where neurons remain active during the transient disappearance of the stimulus, but it is still largely not known how and why such a dichotomy would emerge in the visual system. We demonstrate that a two layer model where motion information is primarily extracted locally before being

diffused along a particular path can provide a solution. Such architecture presents the advantage of mixing different spatial and temporal scales and can be implemented in many biological systems, from retina to cortex. Future works will be conducted to propose a biological plausible implementation of our diffusion mechanism.

To further explore the model dynamics, we tested its robustness by adding background noise in different trajectory conditions. Increasing the background noise induced at some threshold value a sharp change in the dynamics, the model shifting from tracking to no tracking states. Such sharp transition as a function of signal-to-noise ratio is consistent with behavioral studies (e.g. [59]) showing a strong nonlinear relationship between pursuit gain and contrast (see [36] for a review). Interestingly, this sharp nonlinearity of the transition between tracking and non-tracking states —and which is classically implemented by some well-known static non-linear computations such as divisive normalization [55, 57]— emerges here as a property of the dynamical system. The theoretical link between Bayesian inference and divisive normalization has been already suggested by several authors (e.g. [25]) including us ([6]). The current study emphasizes that dynamical inference as implemented here can also reproduce the temporal dynamics of normalization mechanisms through lateral interactions [51]. Further work remains to be done to validate this analogy in particular with respect to the adaptation of this non-linearity to the dynamical statistics of the input.

Our model investigates at an abstract level, the computational advantages of anisotropic diffusion of information within a probabilistic representation of motion. Previous work from Burgi, Yuille, and Grzywacz [11] has suggested that there are multiple analogies of this computing architecture with the structure of neural computation in cortical areas. They originally proposed a constructive approach to implement such motion-based prediction with neural fields. However, their implementation was limited by severe constraints on the simulation of such neural-networks on classical computers. Indeed, this parallel structure is rather not optimal for a sequential computer and necessitate a large amount of memory to achieve a sufficient precision. Hopefully, the advent of novel computational architectures (clusters, neuromorphic hardware) will

foster the precision of the implementation of such models in a more biologically realistic fashion.

A last advance of our model is its ability to reproduce the dynamics of different brain responses to transiently occluded target, from neuronal activity up to highly accurate behaviors such a voluntary pursuit eye movements. Thus, our model has the potential to unify different approaches that were previously proposed to understand motion extrapolation. For instance, recent behavioral experiments imposing a blank during the straight trajectory of a tilted line shows complementing results in the light of our own results [10]. Indeed, they show that if the object is tracked long enough and the blank is short enough, the bias that is characteristic of the aperture problem (the eye following first the direction perpendicular to the segment) disappears. This data is well fitted by a two-layer Bayesian network stacking a sensory and a motor levels. They explain motion extrapolation as a feed-back loop from the representation of the position of the eye to the sensory stage. Our model proposes that a complementary mechanism could be motion-based prediction and that the sensory representation of motion is sufficient to explain motion extrapolation. As we were careful to study the early stage of the tracking response (such that there can be no feedback from a motor stage), we predict that such systems should work in synergy and allow a more complete modeling of motion extrapolation. The main novelty of such scheme is that a simple generic framework —motion-based prediction— may explain a large range of mechanisms that are often explained by the explicit modeling of specialized computations. In our earlier work [49], we have shown that motion-based prediction is sufficient in solving the aperture problem and that the specialized mechanisms that were long supposed to be the source of this solution (texture-independent motion trackers, line-ending detectors, ...) instead emerged from the response to the model.

Many low-level classical problems such as motion extrapolation, the aperture problem or anticipation poses fundamental questions about the computational properties of large-scale networks of neurons. Moreover, their signatures can be found in many different species or neuronal architectures. They are shared by different sensory systems and can therefore be used as a way of unifying the search for generic computations using population codes. Com-

plementing our previous work on the aperture problem [49], we have shown here that the same architecture can solve another instance of low-level uncertainties, extrapolating current trajectories in the absence of sensory evidence. This study demonstrates the need to elaborate generic computational solutions that can eventually be implemented through realistic mechanisms such as divisive normalization mediated by lateral interactions.

## Acknowledgments

This work is supported by projects FACETS-ITN (EU funding, grant number 237955) and “BrainScaleS” (EU funding, grant number FP7-269921). Code to reproduce figures and supplementary material are available on the corresponding author’s website at <https://laurentperrinet.github.io/publication/khoei-13-jpp/>

## References

- [1] Edward H. Adelson and James R. Bergen. “Spatiotemporal energy models for the perception of motion”. In: *Journal of Optical Society of America, A*. 2.2 (Feb. 1985), pp. 284–99.
- [2] J. Allman, J. Kaas, and R. Lane. “The middle temporal visual area (MT) in the bush-baby, *Galago senegalensis*”. In: *Brain Research* 57.1 (July 1973), pp. 197–202.
- [3] Stuart Anstis and V. S. Ramachandran. “Visual inertia in apparent motion”. In: *Vision Research* 27.5 (Jan. 1987).
- [4] John A. Assad and John H. R. Maunsell. “Neuronal correlates of inferred motion in primate posterior parietal cortex”. In: *Nature* 373.6514 (Feb. 1995), pp. 518–521.
- [5] G. R. Barnes and P. T. Asselman. “The mechanism of prediction in human smooth pursuit eye movements.” In: *The Journal of physiology* 439 (Aug. 1991), pp. 439–461.
- [6] Frédéric V. Barthélemy et al. “Dynamics of distributed 1D and 2D motion representations for short-latency ocular following.” In: *Vision research* 48.4 (Feb. 2008), pp. 501–522.
- [7] W. Becker and A. F. Fuchs. “Prediction in the oculomotor system: smooth pursuit during transient disappearance of a visual target.” In: *Experimental brain research* 57.3 (1985), pp. 562–575.
- [8] Robert J. van Beers, Daniel M. Wolpert, and Patrick Haggard. “When Feeling Is More Important Than Seeing in Sensorimotor Adaptation”. In: *Current Biology* 12.10 (May 2002), pp. 834–837.
- [9] Simon J. Bennett and Graham R. Barnes. “Human ocular pursuit during the transient disappearance of a visual target.” In: *Journal of Neurophysiology* 90.4 (Oct. 2003), pp. 2504–2520.
- [10] Amarendra R. Bogadhi, Anna Montagnini, and Guillaume S. Masson. “Dynamic Interaction between retinal and extra-retinal signals in motion integration for smooth pursuit”. In: *submitted* (2013).
- [11] Pierre Y. Burgi, Alan L. Yuille, and Norberto M. Grzywacz. “Probabilistic Motion Estimation Based on Temporal Coherence”. In: *Neural Comput.* 12.8 (Aug. 2000), pp. 1839–67.
- [12] Eric Castet et al. “Perceived speed of moving lines depends on orientation, length, speed and luminance.” In: *Vision Research* 33.14 (1993), pp. 1921–36.
- [13] Mark M. Churchland, I-Han H. Chou, and Stephen G. Lisberger. “Evidence for object permanence in the smooth-pursuit eye movements of monkeys.” In: *Journal of neurophysiology* 90.4 (Oct. 2003), pp. 2205–2218.
- [14] Patricia R. DeLucia. “Chapter 11 Multiple sources of information influence time-to-contact judgments: Do heuristics accommodate limits in sensory and cognitive processes?” In: *Time-to-Contact*. Vol. 135. Elsevier, 2004, pp. 243–285.
- [15] R. Dubner and S. Zeki. “Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey”. In: *Brain Research* 35.2 (Dec. 1971), pp. 528–532.

- [16] R. Eckmiller and M. Mackeben. "Pursuit eye movements and their neural control in the monkey." In: *Pflügers Archiv : European journal of physiology* 377.1 (Oct. 1978), pp. 15–23.
- [17] Yu-Xi Fu, Yaosong Shen, and Yang Dan. "Motion-Induced Perceptual Extrapolation of Blurred Visual Targets". In: *The Journal of Neuroscience* 21.20 (Oct. 2001), RC172.
- [18] D. Goldreich, R. J. Krauzlis, and S. G. Lisberger. "Effect of changing feedback delay on spontaneous oscillations in smooth pursuit eye movements of monkeys." In: *Journal of neurophysiology* 67.3 (Mar. 1992), pp. 625–638.
- [19] Tim Gollisch and Markus Meister. "Eye Smarter than Scientists Believed: Neural Computations in Circuits of the Retina". In: *Neuron* 65.2 (Jan. 2010).
- [20] Erich Graf, Paul A. Warren, and Laurence T. Maloney. "Extrapolation of motion paths behind an occluder". In: *Journal of Vision* 3.9 (Oct. 2003), p. 791.
- [21] N. M. Grzywacz, S. N. Watamaniuk, and S. P. McKee. "Temporal coherence theory for the detection and measurement of visual motion." In: *Vision research* 35.22 (Nov. 1995).
- [22] James H. Hedges, Alan A. Stocker, and Eero P. Simoncelli. "Optimal inference explains the perceptual coherence of visual motion stimuli". In: *Journal of Vision* 11.6 (May 2011).
- [23] Hinze Hogendoorn, Thomas A. Carlson, and Frans A. Verstraten. "Interpolation and extrapolation on the path of apparent motion." In: *Vision research* 48.7 (Mar. 2008).
- [24] John D. Hunter. "Matplotlib: A 2D Graphics Environment". In: *Computing in Science and Engineering* 9.3 (May 2007), pp. 90–95.
- [25] Felix Hürlimann, Daniel C. Kiper, and Matteo Carandini. "Testing the Bayesian model of perceived speed." In: *Vision research* 42.19 (Sept. 2002), pp. 2253–2257.
- [26] U. J. Ilg and Peter P. Thier. "Visual Tracking Neurons in Primate Area MST Are Activated by Smooth-Pursuit Eye Movements of an "Imaginary" Target". In: *Journal of Neurophysiology* 90.3 (May 2003), pp. 1489–1502.
- [27] Uwe J. Ilg. "Slow eye movements". In: *Progress in Neurobiology* 53.3 (Oct. 1997), pp. 293–329.
- [28] K. Kawano. "Ocular tracking: behavior and neurophysiology". In: *Current Opinion in Neurobiology* 9 (1999).
- [29] R. J. Krauzlis and S. G. Lisberger. "A Control Systems Model of Smooth Pursuit Eye Movements with Realistic Emergent Properties". In: *Neural Computation* 1.1 (Mar. 1989), pp. 116–122.
- [30] S. G. Lisberger, E. J. Morris, and L. Tychsen. "Visual motion processing and sensory-motor integration for smooth pursuit eye movements." In: *Annual Review of Neuroscience* 10 (1987).
- [31] Jean Lorceau et al. "Different motion sensitive units are involved in recovering the direction of moving lines". In: *Vision Research* 33.9 (June 1992), pp. 1207–1217.
- [32] Laurent Madelain and Richard J. Krauzlis. "Effects of learning on smooth pursuit during transient disappearance of a visual target." In: *Journal of neurophysiology* 90.2 (Aug. 2003), pp. 972–982.
- [33] Alexis D. Makin and Ellen Poliakoff. "Do common systems control eye movements and motion extrapolation?" In: *The Quarterly Journal of Experimental Psychology* 64.7 (July 2011), pp. 1327–1343.
- [34] Alexis D. Makin, Ellen Poliakoff, and Wael El-Dereby. "Tracking visible and occluded targets: Changes in event related potentials during motion extrapolation". In: *Neuropsychologia* 47.4 (Mar. 2009), pp. 1128–1137.
- [35] Guillaume S. Masson, Anna Montagnini, and Uwe J. Ilg. "When the Brain Meets the Eye: Tracking Object Motion". In: *Dynamics of Visual Motion Processing*. Ed. by Uwe J. Ilg and Guillaume S. Masson. Boston, MA: Springer US, 2010. Chap. 8, pp. 161–188.
- [36] Guillaume S. Masson and Laurent U. Perinet. "The behavioral receptive field underlying motion integration for primate tracking eye movements". In: *Neuroscience & Biobehavioral Reviews* 36.1 (Jan. 2012), pp. 1–25.

- [37] J. H. Maunsell and D. C. Van Essen. “Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation.” In: *Journal of Neurophysiology* 49.5 (May 1983).
- [38] JASON J. McANANY and Kenneth r. alexander. “Is there an omitted stimulus response in the human cone flicker electroretinogram?” In: *Visual Neuroscience* 26.02 (Mar. 2009), pp. 189–194.
- [39] Anna Montagnini et al. “Bayesian modeling of dynamic motion integration”. In: *Journal of Physiology-Paris* 101.1-3 (Jan. 2007), pp. 64–77.
- [40] J. Anthony Movshon et al. “The analysis of moving visual patterns”. In: *Pattern Recognition Mechanisms*. Ed. by C. Chagas, R. Gatass, and C. Gross. Vol. 54. Rome: Vatican Press, 1985, pp. 117–151.
- [41] K. I. Naka and W. A. Rushton. “S-potentials from luminosity units in the retina of fish (Cyprinidae).” In: *Journal of Physiology* 185.3 (Aug. 1966).
- [42] W. T. Newsome and E. B. Paré. “A selective impairment of motion perception following lesions of the middle temporal visual area (MT)”. In: *Journal of Neuroscience* 8.6 (June 1988), pp. 2201–2211.
- [43] Romi Nijhawan. “Motion extrapolation in catching”. In: *Nature* 370.6487 (July 1994).
- [44] Steven J. Nowlan and Terrence J. Sejnowski. “A selection model for motion processing in area MT of primates”. In: *Journal of Neuroscience* 15 (1995), pp. 1195–1214.
- [45] T. E. Oliphant. “Python for Scientific Computing”. In: *Computing in Science and Engineering* 9.3 (May 2007), pp. 10–20.
- [46] Christopher C. Pack and Richard T. Born. “Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain”. In: *Nature* 409 (Feb. 2001), pp. 1040–1042.
- [47] M. Pavel, H. Cunningham, and V. Stone. “Extrapolation of linear motion.” In: *Vision research* 32.11 (Nov. 1992), pp. 2177–2186.
- [48] Laurent U. Perrinet and Guillaume S. Masson. “Modeling spatial integration in the ocular following response using a probabilistic framework”. In: *Journal of Physiology-Paris* 101.1-3 (Jan. 2007), pp. 46–55.
- [49] Laurent U. Perrinet and Guillaume S. Masson. “Motion-Based Prediction Is Sufficient to Solve the Aperture Problem”. In: *Neural Computation* 24.10 (Oct. 2012), pp. 2726–2750. arXiv: 1208.6471.
- [50] V. Ramachandran and S. Anstis. “Extrapolation of motion path in human visual perception”. In: *Vision Research* 23.1 (Jan. 1983), pp. 83–85.
- [51] Alexandre Reynaud, Guillaume S. Masson, and Frédéric Chavane. “Dynamics of Local Input Normalization Result from Balanced Short- and Long-Range Intracortical Interactions in Area V1”. In: *The Journal of Neuroscience* 32.36 (Sept. 2012), pp. 12558–12569.
- [52] D. A. Robinson. “Models of the saccadic eye movement control system.” In: *Kybernetik* 14.2 (Dec. 1973), pp. 71–83.
- [53] D. A. Robinson, J. L. Gordon, and S. E. Gordon. “A model of the smooth pursuit eye movement system”. In: *Biological Cybernetics* 55.1 (Oct. 1986), pp. 43–57.
- [54] D. A. Rosenbaum. “Perception and extrapolation of velocity and acceleration.” In: *Journal of experimental psychology. Human perception and performance* 1.4 (Nov. 1975), pp. 395–403.
- [55] Nicole C. Rust et al. “How MT cells analyze the motion of visual patterns”. In: *Nature Neuroscience* 9.11 (2006), pp. 1421–31.
- [56] G. Schwartz and M. J. Berry. “Sophisticated temporal pattern recognition in retinal ganglion cells.” In: *Journal of neurophysiology* 99.4 (Jan. 2008), pp. 1787–1798.
- [57] Eero P. Simoncelli and David J. Heeger. “A model of neuronal responses in visual area MT”. In: *Vision Research* 38.5 (Mar. 1998), pp. 743–761.

- [58] Grigorios Sotiropoulos, Aaron R. Seitz, and Peggy Seriès. “Changing expectations about speed alters perceived motion direction”. In: *Current Biology* 21.21 (Nov. 2011), R883–R884.
- [59] Miriam Spering et al. “Effects of contrast on smooth pursuit eye movements”. In: *Journal of Vision* 5 (2005).
- [60] Alan A. Stocker and Eero P. Simoncelli. “Noise characteristics and prior expectations in human visual speed perception”. In: *Nature Neuroscience* 9.4 (Mar. 2006), pp. 578–585.
- [61] Emilien Tlapale, Guillaume S. Masson, and Pierre Kornprobst. “Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism.” In: *Vision research* 50.17 (Aug. 2010), pp. 1676–1692.
- [62] S. N. Watamaniuk, S. P. McKee, and Norberto M. Grzywacz. “Detecting a trajectory embedded in random-direction motion noise.” In: *Vision Research* 35.1 (Jan. 1995), pp. 65–77.
- [63] Yair Weiss, Eero P. Simoncelli, and Edward H. Adelson. “Motion illusions as optimal percepts”. In: *Nature Neuroscience* 5.6 (June 2002), pp. 598–604.
- [64] Greg Welch and Gary Bishop. *An Introduction to the Kalman Filter*. Tech. rep. Chapel Hill, NC, USA, 1995.
- [65] Sophie Wuerger et al. “Motion extrapolation of auditory-visual targets”. In: *Information Fusion* 11.1 (Jan. 2010), pp. 45–50.
- [66] Jean-Jacques O. de Xivry and Philippe Lefèvre. “Saccades and pursuit: two outcomes of a single sensorimotor process”. In: *The Journal of Physiology* 584.1 (Oct. 2007), pp. 11–23.
- [67] Alan L. Yuille and Norberto M. Grzywacz. “A mathematical analysis of the motion coherence theory”. In: *International Journal of Computer Vision* 3.2 (June 1989), pp. 155–175.