

Aix-Marseille Université

École Doctorale Sciences de la Vie et de la Santé

THÈSE

Présentée et publiquement soutenue par:

Mina A.Khoei

pour obtenir le titre de Docteur en Sciences d'Aix-Marseille Université
Spécialité : Neurosciences

Motion-based position coding in the visual system: a computational study

Équipe Inference in Visual Behaviour (InViBe)

Institut de Neurosciences de la Timone, UMR 7289, CNRS / Aix-Marseille Université

Jury composé de Messieurs :	
Dr Stéphane Viollet (Président)	ISM, Marseille, France
Dr Emmanuel Daucé (Examinateur)	INS, Marseille, France
Dr Alan Johnston (Rapporteur)	UCL, London, United Kingdom
Dr Laurent Madelain (Rapporteur)	Université Lille III, Lille, France
Dr Anders Lansner (Examinateur)	KTH, Stockholm, Sweden
Dr Laurent Perrinet (Directeur)	INT-CNRS, Marseille, France
Dr Guillaume Masson (Directeur)	INT-CNRS, Marseille, France

Thesis Document - July 2014

Abstract

Coding the position of moving objects is an essential ability of the visual system in fulfilling precise and robust tracking tasks. In daily vision, low contrast conditions introduce sensory noise and transient interruptions or occlusions may temporally block the sensory flow. Moreover, due to neural delays, the visual system has a retarded access to motion information and needs to compensate for the induced positional error. This thesis is focused upon this question: How does the visual system efficiently encode the position of moving objects, despite various sources of uncertainty?

This study deploys the hypothesis that the visual systems uses prior knowledge on the temporal coherency of motion (Burgi, Yuille, and Grzywacz, 2000; Yuille and Grzywacz, 1989). We implemented this prior by extending the modeling framework previously proposed to explain the aperture problem (Perrinet and Masson, 2012), so-called motion-based prediction (MBP). This latter model is a Bayesian motion estimation framework implemented by particle filtering. Based on that, we have introduced a theory on motion-based position coding, to investigate how neural mechanisms encoding the instantaneous position of moving objects might be affected by motion.

Results obtained in this thesis may be summarized in three contributions:

- First, we have modeled tracking of low contrast and interrupted trajectories, by stressing on the velocity-dependent internal representation of motion. We have proposed that the sustained neural activity during temporal absence of stimulus, reported by experimental studies, can be explained as the dynamical integration of velocity information in position coding (Khoei, Masson, and Perrinet, 2013).

- Second, we have modeled how the visual system may compensate for the positional error caused by neural delays. To this aim, we have used an extended version of the MBP model, the so-called diagonal model. This model could explain the misperception of the earliest part of the trajectory of a moving dot (also known as the Fröhlich effect). It could also explain the trajectory-dependent anticipatory response recorded from neural populations in the early visual cortex.

- In the third part of results, we have explored motion-induced position shifts and in particular the flash lag effect (FLE) as a well surveyed visual illusion. We have reproduced

FLE and its various experimental aspects. Our account for FLE is mainly based on motion extrapolation theory, but is also compatible with a mechanism of position persistence.

These theoretical insights suggest that motion-based position coding might be a generic neural computation among all stages of the visual system. This mechanism might partially compensate the accumulative and restrictive effects of neural delays in position coding. As a specific case we have implemented the diagonal MBP model in a network of spiking neurons. Results reproduced the anticipatory response of neural populations in the primary visual cortex of macaque monkey (Kaplan et al., 2014). Our results imply that an efficient and robust position coding might be highly dependent on trajectory integration and that it constitutes a key neural signature to study the more general problem of predictive coding in sensory areas.

Keywords: *Motion-Based Position Coding; Probabilistic Representation Of Motion; Motion Coherency; Bayesian Model; Particle Filter; Predictive Coding; Motion Extrapolation; Motion Anticipation; Neural Delays, Motion-Based Position Shifts; Flash Lag Effect*

Résumé

Le codage de la position des objets alors qu'ils sont en mouvement est une capacité essentielle du système visuel dans l'accomplissement précises et robuste de tâches de suivi . Dans le système visuel, des conditions de faible contraste introduisent de fait un bruit sensoriel. Aussi, les interruptions transitoires comme les occlusions peuvent temporairement bloquer le flux sensoriel. En outre, en raison des délais axonaux des neurones, le système visuel doit compenser l'erreur de position induite par les délais. Cette thèse est centralisée sur cette question: comment est-ce que le système visuel peut coder efficacement la position des objets en mouvement, en dépit des diverses sources d'incertitude?

Cette étude déploie une hypothèse sur la connaissance *a priori* de la cohérence temporelle du mouvement (Burgi, Yuille, and Grzywacz, 2000; Yuille and Grzywacz, 1989). Nous avons ici étendu le cadre de modélisation précédemment proposé pour expliquer le problème de l'ouverture (Perrinet and Masson, 2012). Ce dernier est un cadre d'estimation de mouvement Bayésien mis en œuvre par un filtrage particulaire, que l'on appelle la prévision basé sur le mouvement (MBP). Sur cette base, nous avons introduit une théorie du codage de position basée sur le mouvement. En particulier, nous avons étudié comment les mécanismes neuronaux codant la position instantanée de l'objet en mouvement pourraient être affectés par le signal de mouvement le long d'une trajectoire.

Les résultats obtenus dans cette thèse peuvent être résumér en trois catégories:

- Tout d'abord, nous avons modélisé le suivi à faible contraste et suivant des trajectoires interrompues, en insistant sur la représentation interne dépendant de la vitesse de mouvement. Nous avons proposé que l'activité neuronale soutenue pendant l'absence de stimulus temporel qui est rapportée par les études expérimentales peut être réalisée grâce à l'intégration de l'information de vitesse.

- Deuxièmement, nous avons modélisé la manière dont le système visuel peut compenser l'erreur de position en dépit des délais neuronaux. Dans ce but, nous avons utilisé une version étendue du modèle MBP, le soi-disant modèle diagonal. L'utilisation d'un tel modèle nous a permis d'expliquer la perception erronée de la première partie de la trajectoire (dite Fröhlich effet), ainsi que la réponse anticipée à une trajectoire enregistrée partir de populations de neurones dans le cortex visuel précoce.

- Dans la troisième partie des résultats, nous avons exploré les changements de position induits par un mouvement. En particulier, nous nous sommes concentrés sur l'Effet de Saut de Flash (FLE) comme une illusion visuelle de référence. Nous avons reproduit le FLE et ses différents aspects expérimentaux. Notre modélisation du FLE est principalement basée sur l'extrapolation du mouvement mais est également compatible avec le mécanismes de persistance de la position.

Les connaissances th éoriques de cette thèse suggèrent que le codage de la position basé sur le mouvement peut constituer un calcul neuronal générique parmi toutes les étapes du système visuel. Cela peut en partie compenser les effets cumulatifs des délais neuronaux dans le codage de la position. Comme un cas particulier, nous avons mis en uvre le modèle de MBP diagonal dans un réseau de neurones impulsionnels et avons reproduit la réponse anticipée de populations de neurones dans l'aire cortical V1 (Kaplan et al., 2014). Nos résultats indiquent qu'un codage en position efficace et robuste peut tre fortement dépendant de l'intégration le long de la trajectoire.

Mots-clés: *Représentation probabiliste du mouvement; Cohérence de trajectoire; Modèle Bayésien; Filtre particulaire; Codage prédictif; Délais neuronaux, Effet de Saut de Flash*

Acknowledgements

This Ph.D thesis contains research results undertaken in the team InViBe, Institute de Neuroscience de La Timone (INT), CNRS and Aix Marseille University. The research was in the framework of FACETS-ITN project (EU funding, grant number 237955) and I also benefited from the funding of BrainScaleS project (EU funding, grant number FP7-269921) for the final year.

As any other early stage researcher I have gone through a challenging trip, with many ups and downs, to learn a vast body of principals and methods and to focus on my questions of interest. Fortunately I was not alone in this path and herein I would like to express my gratitude to a number of people. Of course many of them are Dr, but I am not going to include their titles:

Definitely, Laurent as my supervisor is the most influential person in my research and I deeply appreciate his patience in daily mentorship and availability for discussions. I learned research alphabets and principals of computational neuroscience from him and I have been fully benefited from his great expertise in mathematics and modeling. He is full of brilliant ideas and always eager to discuss them in an original framework. I am sure that in my future career I will profit even more from research equipments gained from him.

Furthermore, I would like to thank Guillaume, for many fruitful discussions and very helpful and critical feedbacks on my work. Since earliest months of my thesis work I learned a lot from his concluding comments on seminar or journal club sessions, rising from his very clear scientific mind.

Beside my supervisors, my sincere thanks also goes to thesis jury members, who kindly accepted to review and evaluate my work with their valuable research experience and expertise.

During my thesis work, thanks to my generous scholarship, I profited from two short visits in partner laboratories which were great opportunities for learning and interactions: Daniel Shulz and Valerie Ego-Stengel (CNRS-UNIC-Gif sur Yvette) warmly welcomed me in their lab, with their data on barrel cortex and some nice discussions on the predictive coding in rodent sensory system.

Anders Lansner (KTH, Stockholm) kindly hosted me in his lab, where I learned principles of spiking neural networks and spent hours of valuable interaction with him and Bernhard Kaplan, to implement our theory on spiking neural networks.

I also appreciate Bjoern Kindler for being such a wonderful project administrator and Taarabte Iherti and Joelle Forestier for administrations of INT.

Some other members of InVibe team were also supporting me in different ways:

Fredo was always ready to answer my questions, with his very up to date literature knowledge,

Anna was always there, kind and open to any discussions about Psycophysics and Bayesian models,

Amar and Andrew have been my excellent friends, providing me with scientific discussions and also supporting me emotionally during my hard times,

Seb, Julie and Fred were always very kind and welcoming to my various requests,

Giacomo, Claudio, Wahiba, Sandrine and Sara supported me with their nice attitudes and inspiring discussions,

Eugene deserves a sincere appreciation, as a brilliant scientist and a great friend who always listened to all my various complaints.

I also would like to express my gratitude to S.M Hashemi Golpayegani and F.Towhidkhah, my former teachers in Tehran Polytechnic, who guided me to Neuroscience research with their wonderful courses during my undergraduate studies.

And finally I am indebted to my family for their permanent love, encouragement and support:

My dad, with his immense scientific enthusiasm and a library full of books in Biology brought me up a curious person,

My mom, as my first example of an independent and goal directed woman in a hard society,

And my brother, who believes a lot in me.

Contents

List of Figures

Part I

General introduction

Chapter 1

Motion-based position coding: experimental evidences

Abstract

Processing and sensing the position of moving objects is one of the most complex tasks of the visual system. From evolutionary point of view, it is essential for the survival of animals, to be able to move and react successfully against predators or detect the correct position of moving preys. The locality of moving targets can be estimated only with an accurate motion processing and the related neural mechanisms can be studied from various point of views. In this chapter, we will first review the basic knowledge on motion processing of visual system and then we will focus on motion-based position coding of moving objects by referring to existing experimental studies in the field.

1.1 Introduction

Many tasks in daily life are based on correct localization of moving objects, for instance, actions like catching a moving target or avoiding it in dangerous situations. Dealing with these situations needs a high degree of accuracy, specially in the case of high speed motions. Indeed, retinotopic organization of the visual cortex makes it possible to estimate instantaneous location of responding neural population to a moving stimulus with known properties. Obviously, response properties will depend to the contextual information like detailed physical properties of the stimulus, type of motion etc.

The influence of visual motion on object localization has been under study from various aspects, but without a conclusive agreement. There is relatively clear knowledge about topographical structure of visual cortex and vast amount of literature report characteristics of tuning curves of neural populations in V1 and MT. Despite this, the type of utilized stimuli to survey details of positional coding is most often simple and limited. This can be an advantage which keeps experimental conditions rather under control. On the other hand, analyzing neural response to these basic stimuli types is very useful and can be generalized to some other more complicated motion types with a good degree of approximation.

Dependence of position coding of moving target to properties of its motion can be studied with different experimental frameworks, like measuring reaction times of subjects in locating a certain moving target, or exploring difference between position coding of objects moving with different motion characteristics. All these studies share a core question: knowing that the arrival of stimulus in a particular position stimulates corresponding neural population, then how dynamics of neural responses may be modulated by characteristics of the motion? This type of questions can be addressed by studying ***motion-based position coding*** in the visual system.

Studying positional coding of moving objects is highly related with questions about neural transfer delays and their conflicting consequence in accurate motion localization. In fact, understanding neural mechanisms responsible for position coding also allows to better understand the issue of delay compensation.

In this chapter, we will first briefly review the anatomy and function of earliest visual areas involved in detection of motion and position coding. Then we will go through experimental evidences about modulation of position coding by motion of the stimulus.

1.2 Motion representation in the visual system

For the visual system, motion of objects can be regarded as a systematic displacement of distribution of light in the visual scene. This displacement is detected as early as retina and then is followed by a chain of stimulation in all dedicated cortical areas. Therefore, initial motion signals from the world encode slightest changes in the light equilibrium of a scene, and this will be followed with various complex neural mechanisms in known compartments of the motion processing path.

These mechanisms are origins of our motion perception and also sources of appropriate neural commands to eyes to locate the target. Findings in anatomy of motion processing path, from retina to higher areas, report a highly organized structure with specialized detectors. In this hierarchy, each detector is sensitive to a determined part of the scene as its receptive field, and is involved only in processing of a very particular aspect of moving object inside the receptive field, like speed or direction.

Retina, early estimation of motion information: Visual processing machinery begins in the retina at the moment when light hits photoreceptors. Specialized rod and cone photoreceptors extract relevant information from the light signal and activate some chemical reactions in bipolar cells. Via bipolar cells, stimulation reaches ganglion cells and is converted to the electrical neural message. Ganglion cells occupy the inner most part of the retina and integration of their axons forms the optic nerve, which transfers visual information from retina to thalamus and LGN (lateral geniculate nucleus) and then to the cortex.

Retinal ganglion cells are divided into two different functional types: first M cells which have bigger cell bodies, larger receptive fields and quick response characteristics even in low contrast conditions, and second P cells, with smaller size of body and receptive field and slower response

dynamics, and with sensitivity to high stimulus contrast.

In the particular case of moving objects, specific neural activities responsible for detecting and coding of motion start at retina by activation of direction selective ganglion cells (DSGCs). These cells have been found by Barlow and Levick (1965), just after discovery of direction selective units in the primary visual cortex. They have a preferred direction for moving objects in their receptive fields, and when stimulated with that direction they emit their strongest electrical discharge. The whole population of DSGCs altogether covers all possible motion directions and therefore for any motion direction in the visual field, there will be enough responding cells.

LGN, relay between retina and early visual cortex: The optic nerve then transfers motion information to LGN. It is a part of brain thalamus and relays visual information from eye to the early visual areas. The LGN includes six layers, numbered from 1 to 6 and 1 is the inner most layer. It has a retinotopic organization which means adjacent points in the retina are projected to adjacent points in LGN and all six layers of LGN have this organized structure. These layers are distinct and organized followingly: M type ganglion cells are projected to two inner layers and P cells are projected to layers number 3 to 6. Accordingly, mentioned layers of LGN are called Magnocellular (ventral) and Parvocellular (dorsal) layers. Motion detection is done mostly by M type ganglion cells and Magnocellular layers. These two neural pathways are kept distinct by getting mapped into separate areas in V1.

V1, accurate decomposition of visual motion: V1 (early or primary visual cortex) which is also known as striate cortex is the first destination of the visual signal after being gated and modulated by LGN. During last decades a vast amount of research effort has been dedicated to understanding its structure and function, as well as the analysis of its responses to various moving stimuli. V1 has a retinotopically organized structure composed of highly specialized sensory detectors, specially in the detection of orientation and direction. Furthermore, all detector units are densely connected to others inside the area, via a complex network of lateral interactions.

As a fundamental finding Hubel and Wiesel (1968) published the results of their experimental study on the functional architecture of monkey striate cortex, in continuation of their previous work on cat (Hubel and Wiesel, 1962). This work served as the most seminal research on the visual cortex and opened huge insights to the systematic visual processing by the brain. They conducted their experiments on lightly anesthetized macaque monkey by stimulation of retina with light spots and patterns. With extracellular single cell recordings, they categorized the cells based on their response properties: simple, complex and hyper complex cells .

The proportion of simple cells has been reported to be around 10% of whole recorded population, and they have strong sensitivity to the orientation of presented stimulus in their receptive fields. Their receptive field has an elongated shape with central ON region and two surrounding OFF areas. Complex cells are the most common cell types in V1 and do not have

this separated excitatory and inhibitory areas in their receptive fields (See Fig 1.1).

As a major finding they characterized two main axes for the organization of the cortex: one is tangential to the surface of the cortex (pia) and corresponds to different selectivities (Retinotopy), the other is perpendicular axis to the pia forming cortical columns. In a layered structure, cells aggregated in a column do have common sensitivity to a particular property of stimulus, like its orientation. Though depending on their position in the column they are sensitive to a range of other aspects of the stimulation. Therefore hyper-columns in V1 are formed by finely structured column with orientation selectivity gradually varied around a central point. By dedicating a continuous color code to the range of stimulus orientations, one can show this pinwheel-like organization in the visual cortex (See Fig 1.2).

The role of V1 in motion processing has been widely studied. V1 receives motion signals from LGN and projects them to specialized detectors at retinotopic positions of cortex in presence of a rich network of excitatory and inhibitory connections from surrounding areas, as well as feedbacks from higher areas like MT.

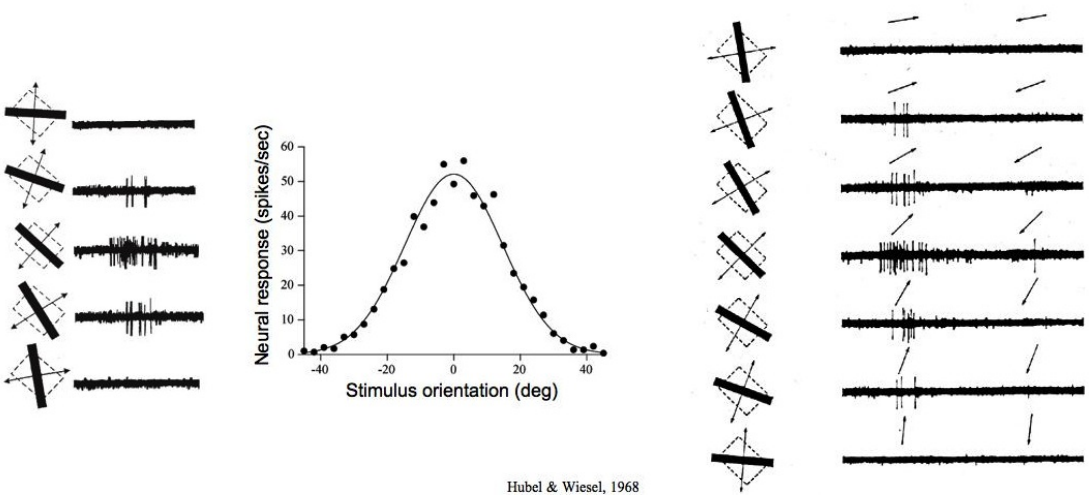

Figure 1.1: **Seminal results of Hubel and Wiesel**, figure adopted from (Hubel and Wiesel, 1968), **Left)** orientation selectivity in V1 cells, the recorded cell shows the most sensitivity to the orientation presented in the middle of the column. The sensitivity of cells to a range of orientations is varied smoothly, as shown by the curve. **Right)** direction selectivity in V1 cells, the recorded cell has the maximum sensitivity to the stimulus presented in the middle of the column.

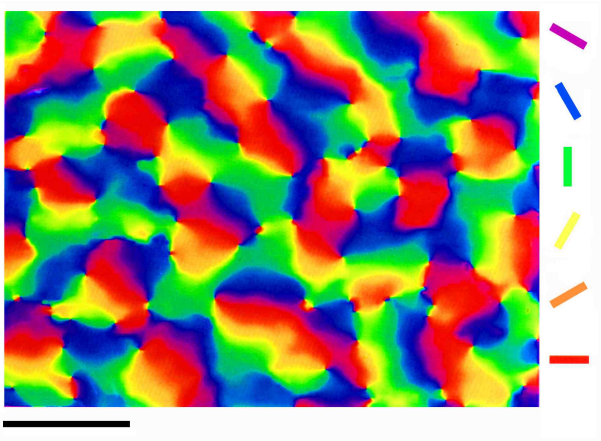

Figure 1.2: **Pinwheel structure of orientation selective functional columns in V1**, figure adopted from (Blasdel and Salama, 1986). Cells with similar orientation selectivity are aggregated in functional columns and a continuous map of orientations forms a pinwheel. The black bar represents $1mm$ of V1.

Motion processing in extra striate cortex: MT is the first area in extra striate visual cortex and it is known to have a key role in motion processing with high selectivity of its cells to speed and direction of moving objects. There are also various studies which emphasize the functionality of MT in integration of local motion signals arriving from V1. Aperture problem is a well known issue addressing questions on details of motion integration process by visual system, and also differentiates exclusive functions of MT. Early motion signals are ambiguous and non conclusive about direction of a moving object and this problem is solved in 100 ms by area MT, where cells show coherent directional response coding global motion (Pack and Born, 2001a). Area MT+ is known to have a role in predicting the location of of moving objects (Maus, Fischer, and Whitney, 2013; Maus et al., 2013).

1.3 Neural delays and predictive position coding

Neural delays are natural consequence of signal transfer from retina to cortical areas, plus sensory processings (Nijhawan, 2002). This leads to expectancy of having a response latency, equal to or greater than the delay. In theory, neural delays impose a time gap between the stimulation of a specific position in the visual field and activation of the reciprocal neural population in the cortex. Then, the most central question which would arise is that, whether they cause a perceptional lag in the localization of moving objects, or the veridical position is inferred by some compensatory mechanisms?

The significance of perceptual lag in localization of moving objects is almost denied by simple behavioral evidences. The primate visual system is efficient enough in fulfilling subtle tracking tasks and generating accurate and on time motor actions when needed. A classical example in this context is a tennis player, which efficiently locates and hits the ball. These subtle motor actions suggest the access of motor system to accurate sensory information from the visual

scene. Otherwise a late action is expected to catch the target at the place it was about 100 ms ago not now. In our daily vision, we always see the actual position of the target, or a well enough approximation of it, and not its position in 100 ms ago. All these evidences reinforce the hypothesis about existence of some compensatory mechanisms which at least moderates the drawbacks of late arrival of motion information to the cortex, if they don't cancel it out completely.

Studying neural delays and the way that visual system may deal with them can provide insights into the positional coding of moving objects in the brain. Clearly, retarded localization of stationary stimuli does not have such a survival value for various species, but these delays would play a crucial role in processing of moving objects. Hence, inferring the actual position of stimulus from its previous path is regarded as a compensatory mechanism. The main debates have been pushed toward the details of those possible mechanisms and also the candidate cortical area that may implement that. Some studies suggest that it is mainly the motor system that builds a well enough approximation of the actual position of stimului (Kerzel and Gegenfurtner, 2003; Nijhawan et al., 2004) and some others have put more effort on surveying the sensory origins of corrective computations (See next sections). Altogether, it seems fair to believe that both sensory and motor systems have their own contributions in those specific computations.

Positional coding has been central question of many studies (Krekelberg, 2001; Krekelberg and Lappe, 2000; Krekelberg and Lappe, 2001; Krekelberg et al., 2000). It is bounded by surveying the 'where' pathway in the visual system, while neural delays highlight the importance of the 'when' pathway as well. A fairly understood structure and function of retinotopic map, provides predictions of neural activities in target cortical positions. Although, it leaves questions on numerous ambiguous aspects, associated with processing of more specific visual stimuli. One of these questions is a more detailed correspondence between position of stimulus in a smooth trajectory and the localization of neural activity in the cortex.

The relationship between neural delays and latency of response in positional coding of stimulus and associated neural mechanisms have been under careful study for last two decades. Some researches have reported shorter delays for moving objects than stationary ones (Whitney, Cavanagh, and Murakami, 2000; Whitney and Murakami, 1998; Whitney, Murakami, and Cavanagh, 2000). Though, latency of response is function of some stimulus properties like its luminance (Purushothaman et al., 1998), as increasing the luminance of moving stimulus is accompanied by shorter latencies in response (Pulfrich, 1922). Reduced neural processing time, compensatory processes and anticipatory mechanisms associated with encoding of moving objects also have been extensively discussed in the literature.

In this context, the Flash lag effect (FLE) as a visual illusion has shown great potential in highlighting some aspects of those questions. In the next section we will review current experimental and theoretical knowledge on FLE.

1.4 The Flash lag effect

It has been longly believed that position coding of an object is independent from motion processing, until some visual illusions provided evidences against this view. These illusions suggest that having the same retinal position for two objects in the visual field does not necessarily guarantee equal perceived position for both. Indeed, positional coding is likely to be affected by some contextual information, like contrast or being in motion or static.

Most studies focused on the sensory origin of compensatory mechanisms are closely tied with flash lag effect. FLE is a visual illusion which has been under study for a couple of decades (MacKay, 1958; Nijhawan, 1994). It reveals a difference between localization of smoothly moving stimuli versus stationary flashed ones. There are few different experimental setups of FLE but all of them share the basic demonstration: a coherently moving stimulus is traveling through a straight or circular path and somewhere in the trajectory a stationary flashed stimulus appears at the exact same position of moving one (See Fig 1.3).

Physical setup of this simple experiment assures that when the stationary stimulus is flashed both stimuli are at the exact same vertical location. While subjects would report a displacement between two stimuli and moving one would be perceived significantly ahead. FLE does not prove if we see the moving stimulus on the real or lagged position, but reveals the possibility of quicker or at least different processing strategies for smoothly moving objects.

It is hard to deduce whether we see the moving stimulus with zero lag, but at least FLE provides a strong evidence that moving objects are perceived more precisely with respect to their real position. In an elegant experimental study on population response of visual cortex in cat, Jancke et al. (2004) suggest that there are unavoidable transfer delays, but these delays are shorter for smoothly moving objects (See Fig 1.5). As such, the representation of a smooth trajectory develops a certain expectancy for approaching stimulation in positions ahead of current location, while it would not be the case for a stationary flashing stimulus.

Flash drag effect is another illusory example which demonstrates how under certain conditions, positional coding of stationary object can be influenced and dragged in the direction of contextual motion (Maus, Fischer, and Whitney, 2013).

FLE was first observed by MacKay (1958) and then for more than three decades did not receive much research attention. Nijhawan restarted working on the similar questions with a challenging experimental setup and reported clear difference in the perception of moving and strobing stimuli (Nijhawan, 1994). In his setup, a bar is moving angularly and two other strobing bars appear aligned with that temporally, but they are perceived displaced. (See Fig 1.3).

Since then, various studies have surveyed this effect and multiple theories have been proposed to account for that. Here we briefly review the most relevant studies:

Motion extrapolation and predictive localization of moving stimulus: Nijhawan proposed a compensatory mechanism in early visual system, which extrapolates the future position of predictably moving stimulus and potentially cancels out the effect of neural delays,

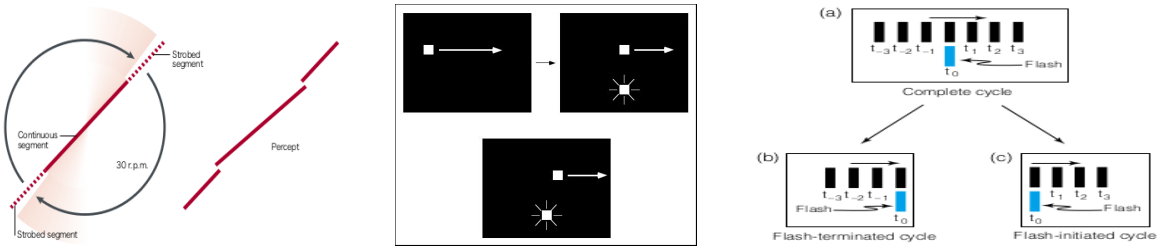

Figure 1.3: **Different demonstrations of flash lag effect, Left)** FLE in angular motion, adapted from (Schlag and Schlag-Rey, 2002), this set up originally studied by Nijhawan (Nijhawan, 1994). A central bar is moving with angular speed 30 r.p.m and two other strobing bars temporally appear along with that. Most of subjects report a perceptual displacement between central bar and strobing ones as it is depicted in right. **Middle)** FLE in linear motion, stimulus is moving coherently and when it reaches to the middle of trajectory a stationary flashed square appears parallel to that. Lower box illustrates the perceptual lag reported by subjects. **Right)** breaking the flash lag experiment into two half cycles: flash initiated cycle and flash terminated cycle, figure is adopted from (Nijhawan, 2002)

while there is no such a mechanism for flashing stimuli (Nijhawan, 1994). Results of this study also suggested that brain may use different mechanisms in position coding of moving and flashing objects, based on having a predictable trajectory behind or not. It also raised a new question on the intrinsically different response latencies for two stimulus types: does the light information of flashing stimulus arrive in the brain later than corresponding information for coherently moving one? Or is there a facilitatory mechanism only for moving stimuli which reduces this delay toward zero?

It sounds more correct to have the transfer delays in the same range for both stimuli, as they are sent through the same physical path. Also one can decompose a smoothly moving object as a densely packed trajectory of flashing stimuli, which their light information are sent in the row from retina to the cortical areas. On the other hand, considering evolution of neural system to be compatible with processing of natural scenes, developing a neural mechanism to overcome neural transfer delays in processing of predictable trajectories represents more survival impact. Thus, it makes sense to dedicate some supplementary mechanisms for quicker processing of moving stimuli.

FLE also raises a philosophical question about the origin of response to each stimulus: is the neural response to the flash really lagging behind? Or we perceive the moving stimuli in an extrapolated position which is always spatially ahead? In the other words, is our percept from the position of moving object matched to its actual position or even slightly ahead of it?

Different processing delays for moving and flashing stimuli: After proposition of motion extrapolation as a candidate mechanism to correct positional error of moving objects, the competency of this approach to account for various implementations of FLE has been under question. For example it has been shown that perceived lags would be shorter for higher frequency of flashing stimuli (Krekelberg, 2001), while this can not be explained with

extrapolation mechanism.

Another useful effort to understand the nature of lags is to break the trajectory of moving stimuli into two half cycles, including trajectories before and after appearance of flash (See Fig 1.3). These trajectory pieces are called flash terminated and flash initiated cycles (Nijhawan, 2002). Studying FLE for two half cycles showed that the no lag is perceived in flash terminated cycle, while the lag observed in flash initiated cycle is comparable to the one of standard FLE setup.

The result observed in the flash terminated half cycle is at first thought in contradiction with what one would expect from a motion extrapolation hypothesis. No matter if the stimulus is in the end of trajectory, extrapolated position still needs to be ahead and produce a perceptual overshot at the termination point. Results from another experiment conducted by Whitney and Murakami (1998) show that the motion extrapolation idea is not compatible also with the cases in which the trajectory is abruptly reversed at some point. They rather suggested that neural delays are likely to be shorter for moving stimuli than stationary ones. The motion extrapolation hypothesis is also unable to explain why a lag is observed at flash initiated cycle where the moving object is at very early trajectory and still not very distinguished from stationary one.

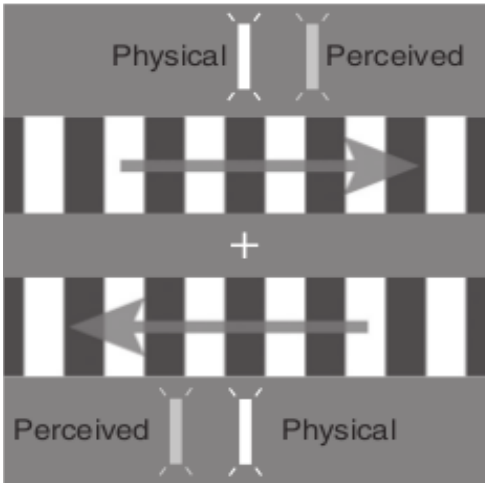

Figure 1.4: **Demonstration of Flash drag effect**, figure adopted from (Maus, Fischer, and Whitney, 2013): a stationary object aligned with a moving pattern is perceived dragged in direction of pattern.

Temporal averaging and postdiction: Another theory proposed for FLE (Krekelberg and Lappe, 2000) suggests a temporal averaging mechanism for the position of stimuli in a window of 500 ms. According to this theory, position information are updated every 500 ms by an averaged position during this period. Clearly, position of stationary flash would remain constant during this period but for the moving stimulus a half distance positional lead will be perceived. This approach does not light up the role of delays in motion processing of moving and stationary stimuli. Also according to motion integration dynamics, 500 ms is a unrealistically

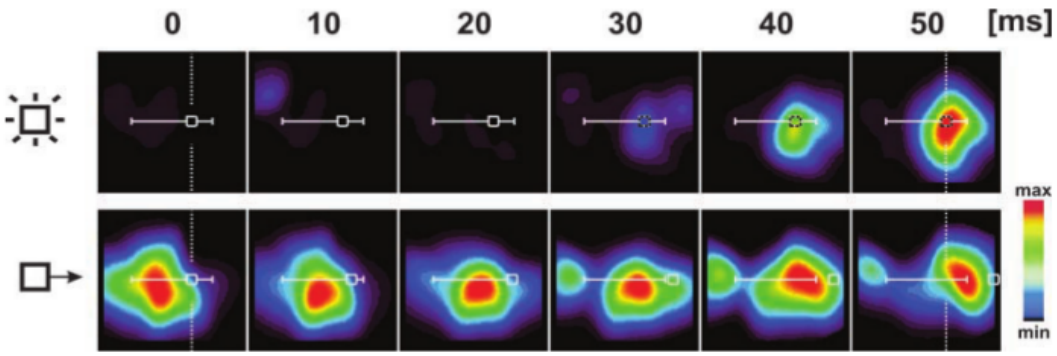

Figure 1.5: **Electrophysiological counterpart of FLE:** recorded spiking activity in cat neural populations of primary visual cortex to study positional coding of moving stimulus versus flashing one, figure adopted from (Jancke et al., 2004)

long averaging window.

Eagleman and Sejnowski (2000) suggested that the visual system is likely to implement a sensory integration mechanism called postdiction. In this mechanism, visual system collects sensory information from the past and future to deduce about current position of the stimulus. In the other words, when receptive field of a specific neural population is stimulated by a moving or flashing stimulus, the corresponding neural activity would be slightly delayed, in order to collect some information from receptive fields that are stimulated just after the current one.

According to the postdiction hypothesis, to interpret received sensory data at current time, visual processing is looking ahead to future and "postdicts" the coding of position. Therefore, every flash would act as a reset for motion integration process. Postdiction model explains why no lag is observed in flash terminated cycle but fails to hold an account for the standard FLE.

Interpretation of FLE with postdiction model comes with allocation of an integration window centered at t_0 (time of flash). Then considering t_0 as the integration restart time leads to a shorter integration window for moving stimulus which starts at t_0. Theoretically, the perceived speed for moving object must be twice as its physical speed, while such a speed incrementation has not been reported by subjects (Hick, 1950).

1.5 Motion extrapolation

The main shortage of motion extrapolation in explaining the FLE is its predicted overshoot at the flash terminated cycle, which does not happen in reality. In later work by Nijhawan (2002) this issue is discussed and the strengths of motion extrapolation as positional error corrector is compared with other above mentioned mechanisms. Focussing on motion onset and motion offset, reveals relatively higher importance of accurate motion offset than motion onset. In the animal world, when the goal is catching a flying prey, positional error at motion onset would not result in failure, as long as an efficient correction is applied somewhere later in the trajectory. For example, when a fly occasionally stops for a very short while, without an accurate localization,

a predator would lose the opportunity of hunting it.

On the other hand, Fröhlich effect (Fröhlich, 1923) which is regarded to share some explanation with FLE, also illustrates that for a moving object with high speed the very early positions in the trajectory are perceptually missed. Therefore, the position reported by subjects as motion onset is significantly shifted toward the middle of the trajectory. Considering Fröhlich effect versus correct perception of motion offset suggests that visual system invests more on the motion termination processing than motion onset.

Nijhawan (2002) suggested that the positional overshoot predicted by motion extrapolation can be canceled out by a special mechanism for registering motion termination. Therefore, motion extrapolation still remains a competent candidate mechanism for FLE. Also, some other studies on FLE have focussed on the separate analysis of moving object and stationary flash stimulus and have reported results in favor of motion extrapolation and in contradiction with the other proposed mechanisms. Fu, Shen, and Dan (2001) have conducted psychophysical experiments to study the motion processing of two dots moving in opposite directions which stop for 100 ms at the same position and then disappear. Their results show a perceptual lag between the position of these dots, which is highly dependent on the blurring and velocity of dots. Report of the lag in this experiment, without any flashing stimulus, questions the role of flash as a reset for motion integration process. Consequently, competency of temporal averaging and postdiction accounts for FLE is also questioned.

Here, we keep motion extrapolation as a demonstration of motion-based position coding and in the next section we will review experimental evidences in favor of motion-based position coding.

1.6 Modulation of position code by motion

Anticipation of moving stimuli by Retina: In an elegant study on ganglion cells of salamander and rabbit Berry et al. (1999) have examined the contribution of retina in FLE and compensation of photoreceptor transduction delays. Stimulating the eye with a moving bar, they have surveyed the neural image on the retina composed of moving wave of spiking activity in population of ganglion cells. Their result shows anticipatory response of ganglion cells to a smoothly moving stimulus, versus response to a stationary flash. According to the recorded neural image, the hill of firing activity over retina is spatially ahead of the leading edge of bar and this is an evidence for motion extrapolation at retinal level.

This anticipatory response shows retina's contribution to motion perception and also accounts for FLE, though existence of similar mechanisms in cortex is also known. Motion extrapolation in retina is limited to a certain range of speed and contrast of stimulus and can be explained with nonlinear contrast gain control in M type ganglion cells and also properties of spatially extended receptive fields. As an alternative view to position coding questions, one can elaborate the systematic sweep of retina by the neural image of moving stimuli. This approach can be also extended to exploring the smooth trajectory of neural activities in two dimensional cortical

sheets.

As another experimental evidence, Schwartz and Berry (2008) have reported extrapolation in retinal ganglion cells as an anticipatory mechanism to correct positional error. According to these studies, the early compensation of delays in retina can have particular facilitation outcomes.

Motion based position coding in the early visual cortex: In an extensive study, Whitney et al. (2003) have explored possible correlations between cortical representation of position and perceived position. They have conducted fMRI and psychophysical experiments and reported that necessarily there is no correlation between BOLD signal of represented position in primary visual cortex and perceived position.

Jancke et al. (2004) have investigated FLE by recording spiking activity in neural populations of primary visual cortex of cat (See Fig 1.5). As an electrophysiological counterpart to the FLE, they have elaborated position presentation of stationary flashing stimuli versus coherently moving ones and have illustrated a similar positional lag. Particularly, they have discussed about pre-activation mechanisms that prepare the future positions in the trajectory.

Receptive field shift in the area V1 in response to moving stimuli: Neural response to moving stimuli is thought to result from an interplay between intrinsic characteristics of recorded neurons and dynamic setup of synaptic connection. Highly specialized detectors in the visual cortex strongly respond to the relevant stimuli but the dynamics of their activity has been known to be partially dependent to the more detailed properties of the stimulus. In a single unit recording setup on anesthetized cat, Fu et al. (2004) have studied motion induced modulations in receptive fields of the primary visual cortex. Their experimental and modeling studies is consistent with the earlier fMRI study of V1 by Whitney et al. (2003) suggesting a shift in receptive fields in a direction opposite to the motion direction (See Fig 1.7).

Asymmetry in receptive field profiles is more evident when they are compared for motions in two opposite directions like rightward and leftward. This shift can also be representative of the anticipatory activity implemented by long range synaptic interactions, after accumulation of enough sensory evidence on the trajectory of motion.

Spatially asymmetric response to the moving patterns in visual cortex: Some research efforts to clarify details of retinotopic position coding have addresses questions about the information content of population response in visual cortex, while getting stimulated by moving patterns. In most of experimental studies it is assumed that the peak of response codes the position of objects, but Whitney and Bressler (2007) have shown that in addition to peak, the pattern of response, like its asymmetry, also contains some meaningful information which contributes in retinotopical position coding. In a fMRI experimental design, they have developed a new technic based on spatial pattern of BOLD signal, to discriminate position codes. The technic is able to discriminate positional responses to moving Gabor patterns at points of

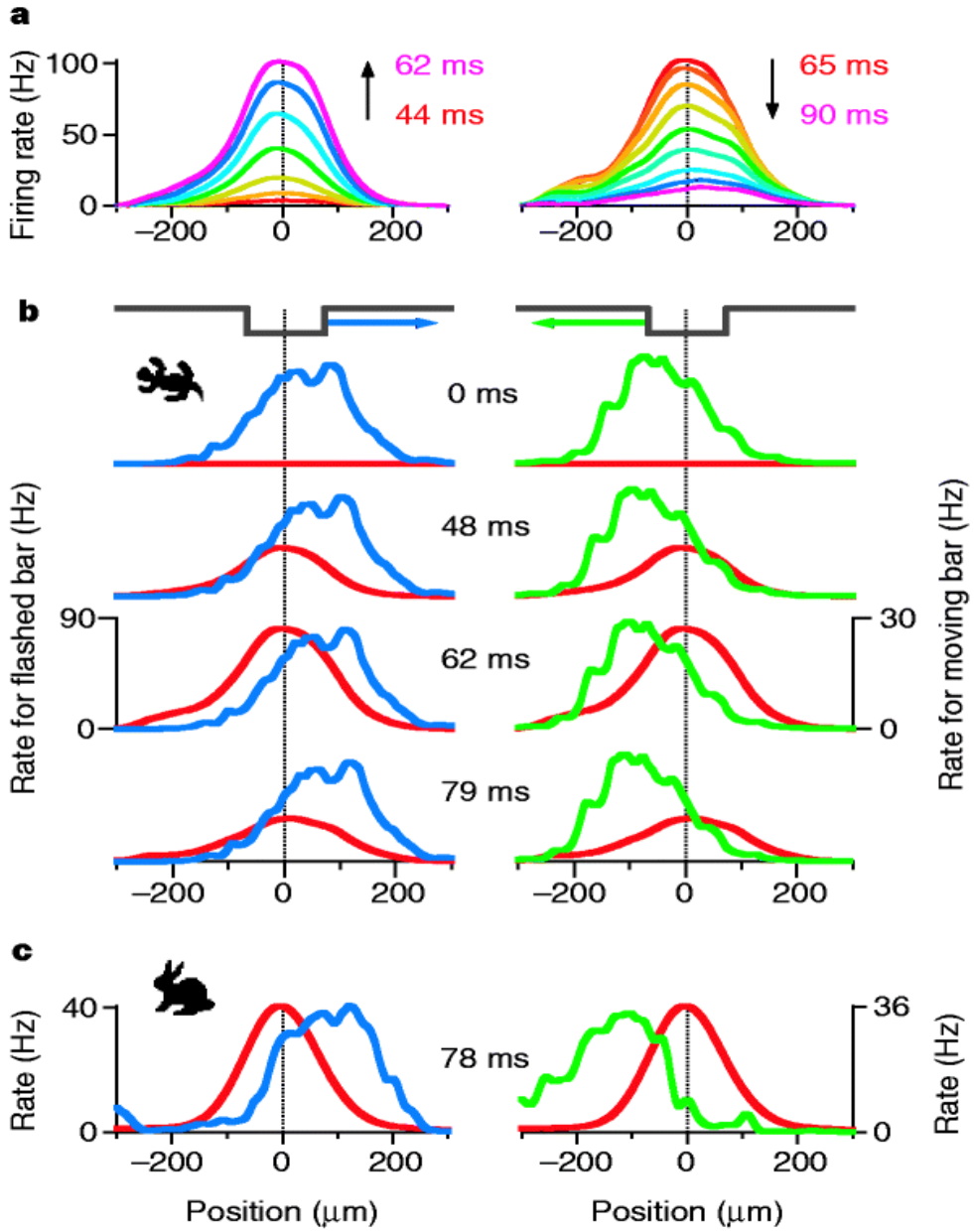

Figure 1.6: **Flash lag effect in retina of salamander and rabbit**, figure adopted from (Berry et al., 1999): Population response of retinal ganglion cells to moving and flashed bar, **(a)** illustrates gradual increase and then decrease of firing rate in response to the flashing stimuli and in **(b)** response of same cells to the rightward and leftward moving bars is shown along with development of response to flashing one. Figures in **(c)** show that when response to the flash reaches to its maximum, response to the moving stimulus is ahead in the direction of motion.

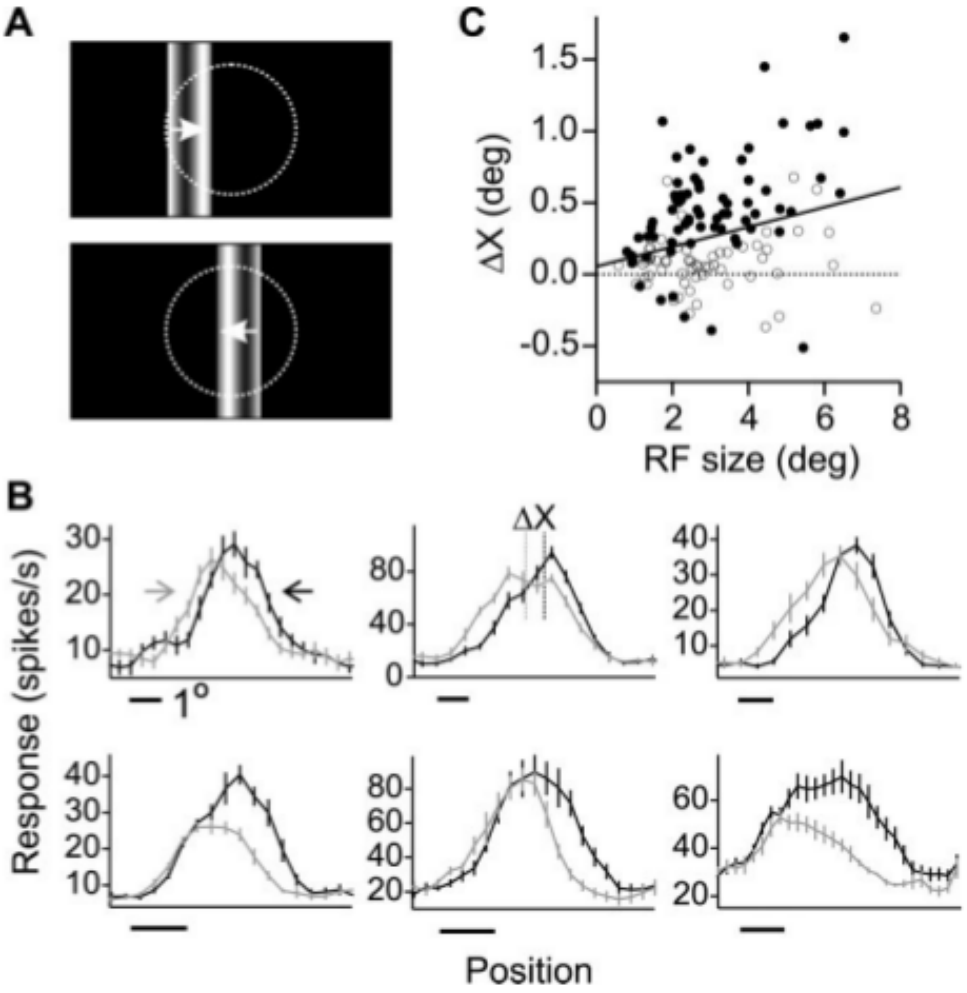

Figure 1.7: **Receptive field shift in V1 caused by motion of stimulus**, figure adopted from (Fu et al., 2004): **B)** illustrates response of six neurons in cat visual cortex to the stimulus shown in **A)**. Cell responses to the stimulus moving rightward and leftward are shown in gray and black respectively, while the preferred direction of cells is leftward. ΔX is the measured shift in center of receptive field and **C)** summarizes the range of shift versus receptive field size. Filled circles correspond to the cells with significant shift

their trajectory with tiny positional difference. This result suggests that retinotopic position coding is modulated by visual motion and this effect may be implemented by long range neural mechanisms.

Position codes of moving objects at different parts of their trajectory has different spatial pattern and this may be an evidence for existence of a large scale code for trajectory motion. In such a code, depending on how far the object is from the onset of motion, spatial profile of response would be affected by long range modulations. As existence of a relatively precise position coding in human pulvinar area has been reported (Fischer and Whitney, 2009), thus early visual cortex may compute a more fine grained position map as well as coding the relative position in the trajectory of moving objects.

Predictive position coding in V1 and V3A areas: Some experimental data have reported motion induced mislocalization of moving objects at the end of their trajectory. This perceptual effect, happens with no retinal input and may originate from cortically extrapolated position responses, that is a forward shift in cortical representation of moving objects. To study this effect, Maus et al. (2010) conducted an experiment with a moving stimulus in two conditions: outward condition in which motion starts with lowest contrast and then contrast is gradually increased and trajectory ends at highest contrast, and second, inward condition where stimulus starts with highest contrast and gradually fades out at the end of trajectory.

They reported that in the inward condition where motion fades out at the end of trajectory, the stimulus is perceived to disappear at positions that correspond to contrasts lower than the luminance detection threshold.

They also assessed recorded responses in the retinotopic areas matched to extended part of trajectory of fading stimulus and found that area V3A predictively codes for the position of moving objects. These results also support the idea that V3A participates in anticipatory position coding for moving objects by using information from trajectory, while the feedback to V1 may also influence retinotopic position representation in V1. This study does not report any significant predictive component in the position representation in V1. This is perhaps because V1 has a more fine grained retinotopic map and also it is more driven by retinal input.

This insight is also consistent with another hypothesis: position coding of a moving object can be distributed at different stages of cortical representation, depending on the relative position of stimulus in the trajectory. As such, V3A may keep anticipating, where V1 codes the motion offset, with complementary mechanisms to correct perceptual overshoots and overcome prediction errors.

Motion dependent presentation of space in areas MT and MT+: Area MT+ of primates has been known for its coarse retinotopic organization and many studies focused on its contribution to motion perecption. To explore the contribution of MT+ and V1/V2 areas in positional lead of moving objects in FLE, Maus et al. (2013) disrupted the neural activity in these areas by means of TMS (Transmagnetic Stimulation) and have measured the degree of

perceptual lead for moving object. They have reported a considerable decrease in FLE by TMS interruption of MT+ and no significant effect by conducting the same procedure in areas V1 and V2. Their results suggest that neural mechanisms responsible for position presentation of moving objects are mainly implemented in the area MT+.
In a fMRI study, Maus, Fischer, and Whitney (2013) also emphasized that position coding mechanisms in MT+ area are affected by motion based phenomena. They have reported that receptive field properties of neurons can change with moving stimuli. These findings are consistent with anticipatory position coding view, suggesting that in the area MT+ population receptive fields and spatial coding are modulated by motion.

Another fMRI study in human area MT (d'Avossa et al., 2007) reported modulation of BOLD responses by gaze direction.

1.7 Processing of trajectory motion

Positional extrapolation of motion trajectory is the most debated account for FLE. Indeed, it provides very intuitive insights to the motion processing constraints of brain. There are some sparse studies which have conducted relevant experiments and have offered evidences that some motion-based phenomena govern position coding of trajectory motion.

Detection of trajectory motion by V1 and MT neurons: Mikami et al. (1986) have explored spatiotemporal range of directional interactions in V1 and MT by stroboscopic flashing stimuli. They presented stroboscopic flashing stimuli with varying spatial and temporal intervals to measure the limits of directional selectivity in neural populations of V1 and MT. They report that spatial interval limit for MT neurons is much larger than the same limit for V1 neurons, while the temporal interval limit is the same for both V1 and MT populations.

The main finding of this study is the ability of direction selective cells in MT to code higher speeds by integration of motion information over a bigger distance at the same time. Their results indicate that MT neurons can interpolate between successively flashed stimuli and process it as a trajectory, whereas V1 neurons stay irresponsive for this kind of motion.

Temporal recruitment along the trajectory of moving objects: Sensitivity of position coding to the history of motion is not unexpected, as motion integration process itself needs 100 ms to be completed. To explore positional judgments by temporal recruitment of trajectory information and extending the study of Nijhawan (1994), Krekelberg and Lappe (1999) conducted an angular demonstration of FLE. The stimulus includes three inner dots in the middle of a rotating bar and two occasionally flashing dots at each side. Flashing dot appeared along with central part but subjects reported an angular lag between the two parts of the stimulus (See Fig 1.8).

The experiment is repeated with different temporal visibility durations for the flashing part and also over a range of angular speeds. Their results suggest that, longer visible trajectory

enhances position perception of stimulus. The similar effect is observed in the dependence of lag with respect to speed of the stimulus. In standard FLE experiment, the lag would increase with the speed of moving stimulus, but this effect can be modulated by presenting flashing stimuli over a longer spatial range. Consistently with the motion extrapolation hypothesis, they propose that integrating along the trajectory of motion, including speed information, improves position judgements, it is as well a potential mechanism for correction of neural transfer delays .

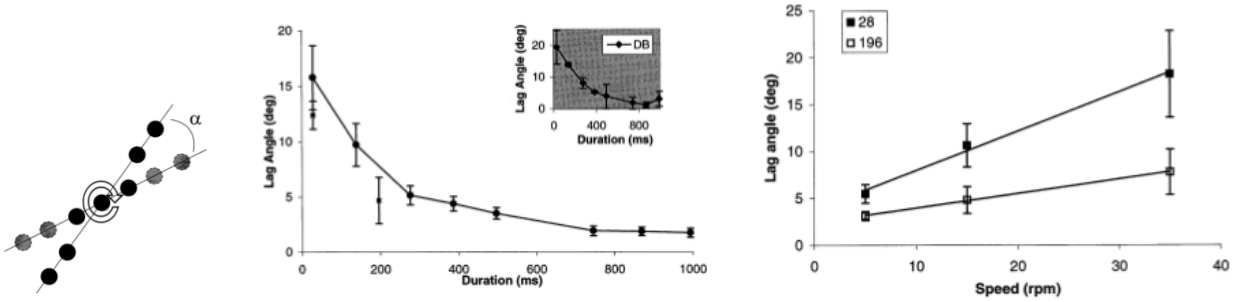

Figure 1.8: **Angular counterpart of flash lag effect**, figure adopted from (Krekelberg and Lappe, 1999). **Left)** A rotating bar includes three central and fixed dots and two flashing dots in each side of the bar. The actual and perceived positions respectively are shown by gray and black and α is lag angle. **Middle)** α falls down to zero by increasing the trajectory length of flashing dots. The main figures correspond to the data averaged over ten participants, while inset is the result of a single subject and in one trial. **Right)** α increases with increase in speed of rotation, the figure shows it for two different trajectory duration.

Motion anticipation and facilitatory neural mechanism in processing of smooth motions: Focusing on delay compensatory mechanisms, motion extrapolation can be discussed from some other aspects like trajectory anticipation or pre-activation of neural populations. Pre-activation is development of response in neural populations which their receptive field is very likely to be stimulated with respect to the history of trajectory. The classical receptive field (CRF) of a neuron or a neural population is defined as a portion of visual field which its stimulation leads to a strong neural activity in neuron or population. Knowing the retinotopic map of visual cortex and for a given motion trajectory, one can predict which populations are going to be activated.

Going further from basic knowledge on well studied neural populations in the visual cortex, during last few decades there has been intense debates on more detailed sensory mechanisms associated with the connection and interaction of different neural populations in a known area. Different efforts to explore modulatory effect of stimulated populations on each other and on non stimulated ones has brought some new terms in categorization of receptive fields. As most studied ones, the terms of nCRF, eCRF and PRF respectively refer to none classical, extra classical and population receptive fields. We will review relevant works for each of these categories.

Modulatory effect of coherent trajectory on motion processing: Given the extensive

lateral, feedback and feedforward connections in the visual cortex, any study neglecting them and their modulatory role in the response of each cell would end up in non-realistic conclusions. Therefore, as suggested by some previous studies, most cells integrate the signals from their CRF with the contextual sensory information from a more extended neighborhood, called the non-classical receptive field (so called nCRF).

Guo et al. (2007) have recorded the activity of cells in V1 to study their relative sensitivity to the stimuli inside their CRF, versus the more distant stimuli. They have shown different modulations in response depending on the motion trajectory. This study suggests that each neuron in V1 codes the information in its CRF, but importantly also contributes in coding of larger scale spatiotemporal structure of the whole scene.

This study is compatible with Bayesian inference definition. In this context, network connections of visual cortex can be assumed as an implementation of prior information on the statistical regularities of physical world. Then direct input in the CRF of cell would be equivalent of the current likelihood and the posterior inference on spatiotemporal structure of neural response is affected by both (priors and liklihoods). With a simplified view, one can consider that different motion trajectories activate different priors (as they travel trough different network connections), therefore they do support arrival of a stimulus into a specific position, by developing an anticipatory response before arrival of stimulus.

From this point of view, pre-activation of neural population is also discussable: if a stimulus is moving smoothly between positions **A** and **B**, by progression along the path, neural populations associated to the CRF at position **B** accumulate more evidence to predict arrival of stimulus at their CRF. This gradual preparation appears as a rising phase in the response of cells, before arrival of stimulus to their CRF and we refer to it as trajectory-dependent anticipatory response (Benvenuti et al., 2011).

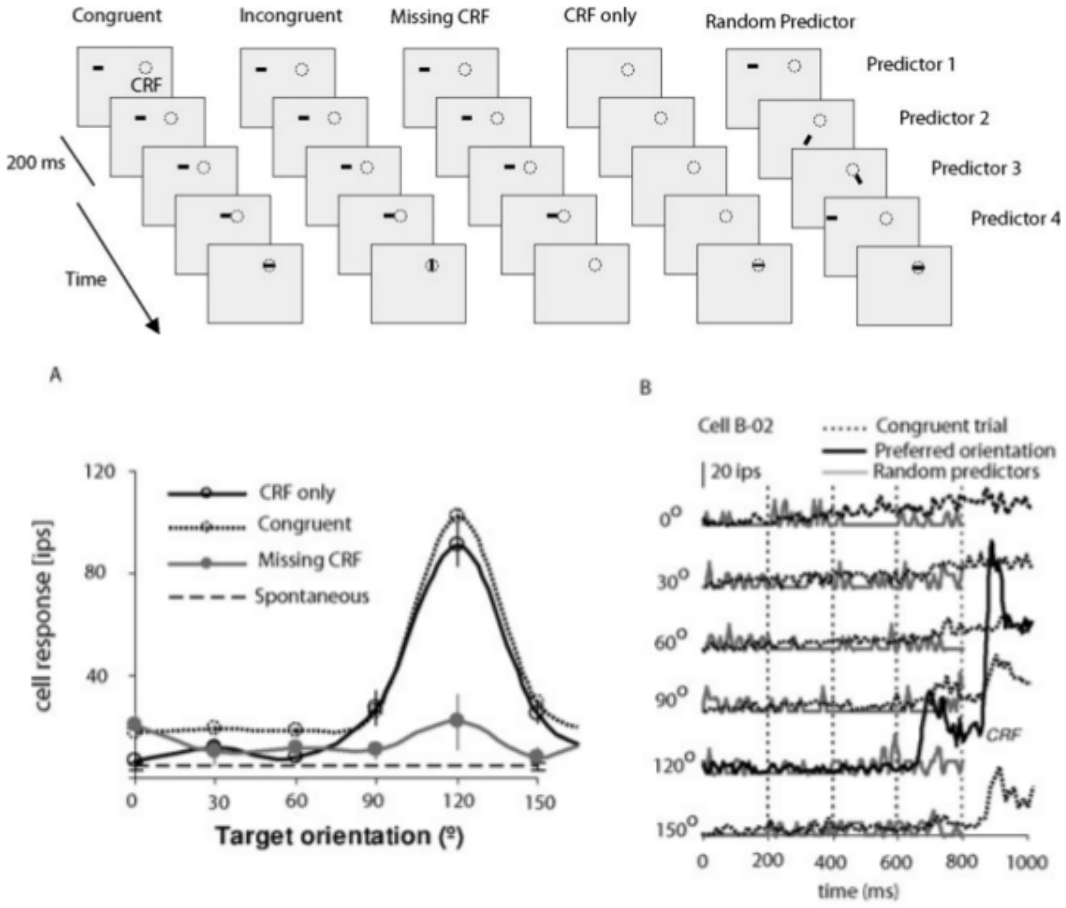

Figure 1.9: **Spatio-temporal prediction and inference by V1 neurons**, figure adopted from (Guo et al., 2007). **Top)** Different stimulation protocols 800 *ms* before arrival of stimulus to the CRF, **congruent:** Stimulus has the same orientation inside and outside of CRF, **incongruent:** orientation of stimulus is changed inside CRF, **missing CRF:** stimulus disappears inside CRF, **CRF only:** there is no trajectory before arrival of stimulus to CRF, **random predictor:** stimulus takes a random shuffled trajectory before arrival to CRF, **Bottom)** orientation preference of cell at different protocols and the response at each orientation. Early anticipatory response at preferred orientation and congruent trials is evident.

In some other studies, stimulation of the surrounding area has provided better understanding of motion processing by CRF. For example, modulatory effect of peripheral stimulations on the localization and temporal properties of neural response has been investigated (Alink et al., 2010). The most general question of all these works is the detailed correspondence between the spread of neuronal activity in the cortical area and the stream of motion in the visual field.

Coherent motion of a small stimulus excites many neural populations with overlapping CRFs and motion trajectory is mapped to a progressive code of position presentation in the cortex. As the selectivity of populations for the presented stimulus is not in a binary manner, it is more appropriate to look at neural code as a response composed of "votes" by all involved populations based on their preference. In a predictable motion, the hills of response are centered on the cells

coding the veridical position of stimulus and are shifting toward other populations which are connected to the next positions in the path.

Furthermore, some populations corresponding to more distant future in the trajectory, start to accumulate their belief on approaching arrival of stimulus to their CRF. Long range lateral connections in visual cortex are the most likely candidates for the propagation of information toward 'about to stimulate" populations. From this view, lateral connections alert about increasing probability of stimulation in the positions farther ahead in the trajectory.

In end-stopping cells, response to a preferred stimulus can be attenuated or even eliminated, if the surrounding area of CRF also gets stimulated with the same stimulus. Such a modulatory mechanism is called extra classical receptive field (eCRF) effect and has been reported to happen in V1, V2, V4 and MT in cortical areas (Harrison et al., 2007; Rao and Ballard, 1999). The eCRF effect has been suggested to be a way to implement predictive coding in the brain.

As mostly in natural images the intensity difference between neighboring areas is low, therefore response in CRF can be predicted from the neural code associated to eCRF. It has been discussed as a processing strategy to decrease computations related to redundant and predictable parts of the scene. Thus, eCRF probably acts as an error detector in processing of coherent motion, generating a residual signal composed of the difference between reference input signal and what has been predicted by an appropriate internal model. Then the residual signal is reported to higher order cortical areas and the regulatory signal is sent to striate cortex via feedback.

With similar questions Harrison et al. (2007) studied functional role of eCRF in striate cortex with fMRI. They have investigated how global and contextual information may affect local motion processing. By studying coherent versus incoherent motion they have found that increasing coherency of motion will result in more suppression in the response. This study also suggests that eCRF may enable the brain to implement predictive coding, by detection of regularities in the sensory input and reporting it to higher areas. Another fMRI study has reported that predictability of motion onset and motion direction would reduce the response in V1 (Alink et al., 2010).

As a recent psychophysical evidence for predictive coding of smooth motion, Roach, McGraw, and Johnston (2011) have conducted an experiment to study the modulatory effect of spatial regularity of stimulus on its detectability. In this experiment, an elongated grating stimulus with high contrast is illustrated as an inducer, and a small low contrast grating appears either at leading or trailing edge. They have explored dependency of stimulus detectability to the relative phase of inducer and target gratings. The results have been interpreted in the framework of constructive or deconstructive interference between stimulus and internal prediction. Therefore, small phase difference between induced and target stimuli would boost detectability of target as a constructive combination between both, and will induce a forward prediction of the spatial pattern.

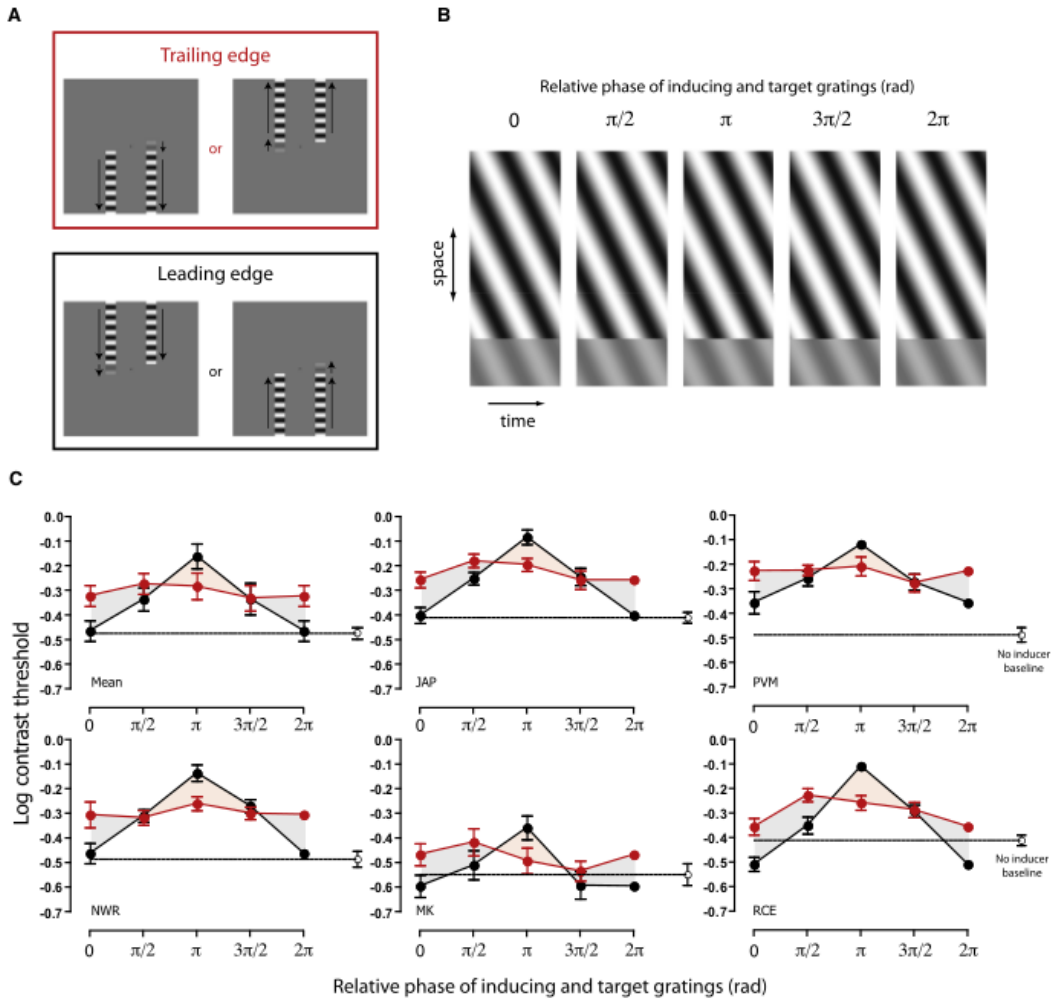

Figure 1.10: **Forward prediction of spatial pattern**, figure adopted from (Roach, McGraw, and Johnston, 2011). **A)** stimulus configuration, two elongated bars are presented to the subjects and a low contrast small stimulus would appear either at leading or trailing end of one of the bars. Elongated gratings act as inducers to study detectability of low contrast target. **B)** relative phase of two stimuli **C)** detectability of target stimuli at different phases, top left box illustrates mean response of five other subjects. Red and black traces respectively correspond to trials in which target appears at trailing and leading edges

Traveling neural waves in the visual cortex: Another line of research in the context of trajectory processing and long range spatio temporal motion integration is to study traveling waves in the visual cortex. There are experimental evidences that illustrate, for a localized stimulus, a systematic stream of activity diffuses to populations corresponding to the more ahead positions. This research area investigates large scale motion integration details by focusing on the measurement of diffused activities. The typical method to do that, is placing an electrode in a target population and study the development of response by stimulation of receptive field of a more distant population. Studies related to traveling waves have used VSD imaging (Benucci, Frazor, and Carandini, 2007; Bringuier et al., 1999; Grinvald et al., 1994; Slovin et al., 2002)

and LFP analysis (Busse, Wade, and Carandini, 2009; Kitano et al., 1994; Nauhaus et al., 2008; Ray and Maunsell, 2011). By measuring neural response in positions of the trajectory which are ahead of stimulation point, they have found that more distance from the excited receptive field and the electrode will elicit weaker and later response.

Two main candidate mechanisms have been proposed for traveling waves: first, the convey of delayed excitation via long range connections, and second feedback from higher areas. On the other hand the existence of traveling waves have been related to uncertainty in the stimulus. For example, when the contrast of stimulus is low, motion integration processes need to collect information over larger spatiotemporal windows before shaping the final neural code. Traveling waves provide propagation of such a distributed activity. They appear also in spontaneous activity of brain, but as soon as having a high contrast stimulus they do vanish.

1.8 Position coding in an interrupted trajectory

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 estimating 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 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 (Gollisch and Meister, 2010).

The problem of motion occlusion is a particular case of a more general problem in neuroscience: *extrapolation of interrupted trajectory*. 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 tends to follow smooth, straight trajectories. Such a 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 (Bennett and Barnes, 2003) and such a feat is only possible when observers have some knowledge on the path of motion (Graf, Warren, and Maloney, 2003). 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.

Followingly, we will review experimental studies and theories about extrapolation of an interrupted trajectory.

1.8.1 Extrapolation of interrupted motion trajectories: experimental evidence

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 at the time they judge the occluded target to have reached a particular point (Rosenbaum, 1975). 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 (Berry et al., 1999; Gollisch and Meister, 2010; Schwartz and Berry, 2008) or in higher cortical areas (Assad and Maunsell, 1995). Behaviorally, motion extrapolation was studied in the context of target catching (Nijhawan, 1994), apparent motion (Hogendoorn, Carlson, and Verstraten, 2008) and trajectory extrapolation for occluded or disappeared stimuli (Makin, Poliakoff, and El-Deredy, 2009), perceptual extrapolation of blurred visual target (Fu, Shen, and Dan, 2001), in audio visual targets (Wuerger et al., 2010), role of motion extrapolation in control of eye movements (Makin and Poliakoff, 2011), and blurred targets (Fu, Shen, and Dan, 2001).

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 (DeLucia, 2004). Here, we investigate visual, lateral motion extrapolation as a generic paradigm to challenge prediction algorithms and we will briefly review the relevant experimental evidences.

MT, MST and STS: The middle temporal (MT) and medial superior temporal (MST) areas in the superior temporal sulcus (STS) of primates process visual motion and oculomotor signals driving pursuit (see (Ilg, 1997) 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 (Allman, Kaas, and Lane, 1973; Dubner and Zeki, 1971). 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 (Maunsell and Van Essen, 1983). Pack and Born (2001b) 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 100 ms, 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; for instance during a transient image occlusion (Newsome, Wurtz, and Komatsu, 1988) or during tracking imaginary target covering the visual field in the outside of the receptive field which is currently recorded (Ilg and Thier, 2003).

Posterior parietal cortex: 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 (Assad and Maunsell, 1995).

Retina: In another study, Schwartz and Berry (2008) 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 (McAnany and Alexander, 2009).

Eye movements: The neural systems controlling smooth pursuit eye movements (SPEMs) are likely to be critically dependent upon motion extrapolation, in close synergy with saccades (de Xivry and Lefèvre, 2007). Several studies have shown that blanking a small moving target results in a very typical temporal profile of eye velocity. Eckmiller and Mackeben (1978) investigated smooth pursuit behavior in macaque monkey 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 (1985) 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 (1985) 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 (Barnes and Asselman, 1991).

There is an ongoing debate of whether the origin of motion extrapolation is within the oculomotor control system (Makin and Poliakoff, 2011) or rather occurs at the sensory level. Using event related potentials, Makin, Poliakoff, and El-Deredy (2009) have suggested on electrophysiological grounds that both systems may be contributing.

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. Becker and Fuchs (1985) 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 (2003) 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 Bogadhi et al. (2011) investigated 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 initiation. 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. What is the link between behavioral and neuronal signatures of motion extrapolation? Visual motion information is primarily used for gaze stabilization (Ilg, 1997; Kawano, 1999; Masson, Montagnini, and Ilg, 2010) and sensorimotor transformation underlying smooth pursuit eye movements (Lisberger, Morris, and Tychsen, 1987). 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 (Newsome, Wurtz, and Komatsu, 1988). 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. Also, as motion extrapolation of interrupted trajectories 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.

1.8.2 Motion inertia and extrapolation of interrupted trajectory

A tightly coupled phenomenon is motion inertia, which might be regarded as the perceptual equivalent of motion extrapolation for object identification. To put motion inertia in evidence, it has been shown with the following experiment: 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 the previous direction (Ramachandran and Anstis, 1983). 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 (Anstis and Ramachandran, 1987).

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 (Pavel, Cunningham, and Stone, 1992). 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.

1.8.3 Existing theories on extrapolation of interrupted trajectories

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 (Robinson, Gordon, and Gordon, 1986). Such models were refined to specifically address the problem of motion extrapolation (Churchland, Chou, and Lisberger, 2003) by including additional layers in a cascade model (Goldreich, Krauzlis, and Lisberger, 1992). These models may be subdivided into those where the predicted signal is based of some motor command (Bennett and Barnes, 2003) and those that specifically use the adaptation of an internal model (Madelain and Krauzlis, 2003). 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 (Krauzlis and Lisberger, 1989; Robinson, 1973; Robinson, Gordon, and Gordon, 1986).

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 is used to estimate the current state based on their reliability reflected from their variances. Studies investigating sensory-motor transformation already suggest 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 (Beers, Wolpert, and Haggard, 2002). Similarly, this may be expressed in as a Kalman filter, that is in a generic Bayesian framework with a clear hypothesis (Welch and Bishop, 1995).

Following the idea of Kalman filter and extending the work of Montagnini et al. (2007), Bogadhi, Montagnini, and Masson (2013) 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 (Weiss, Simoncelli, and Adelson, 2002). They are somewhat similar to some of the engineering models proposed earlier (Nowlan and Sejnowski, 1995), but allow for a more explicit formulation of the underlying hypothesis. Such a framework accommodates uncertainty in the motion information in the measurement likelihoods (Hedges, Stocker, and Simoncelli, 2011; Stocker and Simoncelli, 2006; Weiss, Simoncelli, and Adelson, 2002) and also expectation can be represented through the prior which can alter motion perception (Sotiropoulos, Seitz, and Seriès, 2011). 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. (2010) 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. 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 (Welch and Bishop, 1995), 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 (Burgi, Yuille, and Grzywacz, 2000). The approach from Bogadhi, Montagnini, and Masson (2013) 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.

Summary

In this chapter, we have introduced motion-based position coding as a theory which centralizes a family of neural data on visual motion processing. After a brief review on principals of

motion processing path, we have targeted the problem of neural delays and predictive position coding as a compensatory mechanism. Then we have reviewed various experimental evidences which demonstrate the influence of the motion signal on position coding. For instance, we have elaborated FLE as a well studied visual signature with various proposed theories. We also have reviewed the experimental studies associated with motion extrapolation in interrupted or occluded trajectories. At the end we have briefly introduced some of proposed models and theories for the coherent perception and tracking of interrupted trajectories.

In the next chapter we will focus on some computational models which are the most relevant with motion-based position coding theory, as a base for our model and methods described in chapter 3.

Chapter 2

Models of the predictive estimation of visual motion

Abstract

In this chapter, after a review on principal models of visual motion estimation, we will focus on the specific family of models based on a hypothesis on temporal coherency of motion. In the second section the review will continue by introducing models consistent with motion-based position coding theory.

2.1 Visual motion estimation: models

From the physics point of view, motion is defined as the change in position of an object with respect to reference points in time and space. It is usually associated with other concepts such as speed, direction and acceleration. More precisely, what matters for the visual system and can be detected by photoreceptors in the retina, is the relative displacement of a coherent volume against it's background, which leads to a systematic move in the balance of luminance in the visual scene (or optical flow). In the following sections we will introduce modeling studies on detecting and estimating of visual motion.

2.1.1 Spatiotemporal energy models for perception of motion (Adelson and Bergen, 1985):

Efforts to formulate the visual motion has lead to mapping it in terms of its spatiotemporal components in $x - y - t$ coordinates. An object moving in $x - y$ coordination would have an oriented trace in $x - y - t$ space and the tilt of this trace corresponds to the motion speed (See Fig 2.1).

Based on this idea, Adelson and Bergen (1985) proposed a motion energy model, to decode speed and direction by detection of their slanted spatiotemporal path. In Fig 2.2, the spatiotemporal oriented path has been shown for different speeds and directions.

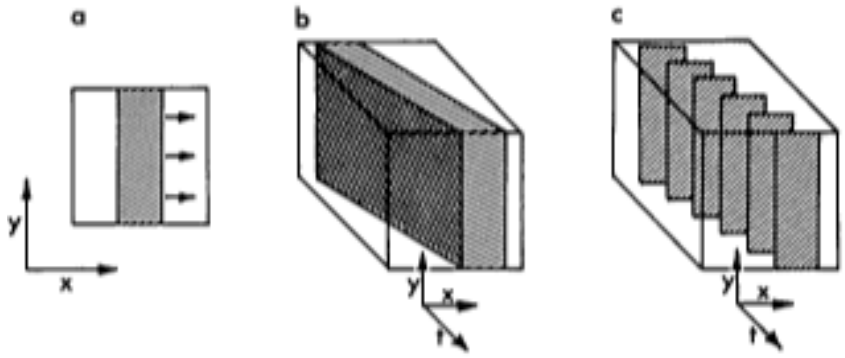

Figure 2.1: **Spatiotemporal components of motion in $x-y-t$ coordinates**, figure is adapted from (Adelson and Bergen, 1985): A bar moving to right is leaving an oriented trace in $x-y-t$ space. **b** and **c** show the slanted continuous and discrete paths

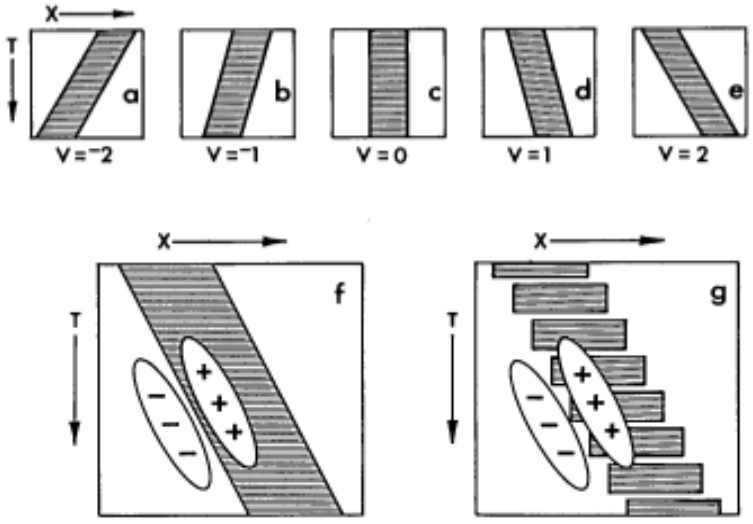

Figure 2.2: **Speed in motion energy model**, figure is adapted from (Adelson and Bergen, 1985): the slope of oriented path shown in Fig 2.1 represents the speed of motion (figures **a** to **e** are illustration of different speeds in $X-T$ space). It can be detected by a spatiotemporally oriented receptive field shown in **f** and **g**.

In the motion energy model, spatial and temporal response of motion detector sub-units are separable as shown in Fig 2.3. Then, the spatiotemporal impulse response of unit composed of positive and negative lobes would filter the motion path and rebuild the motion trace in x-y-t space (See Fig 2.4, b and c). These filters detect motion but are not direction selective. Direction selective Gabor filters can be built by different combination of spatial and temporal impulse responses as shown in Fig 2.3.

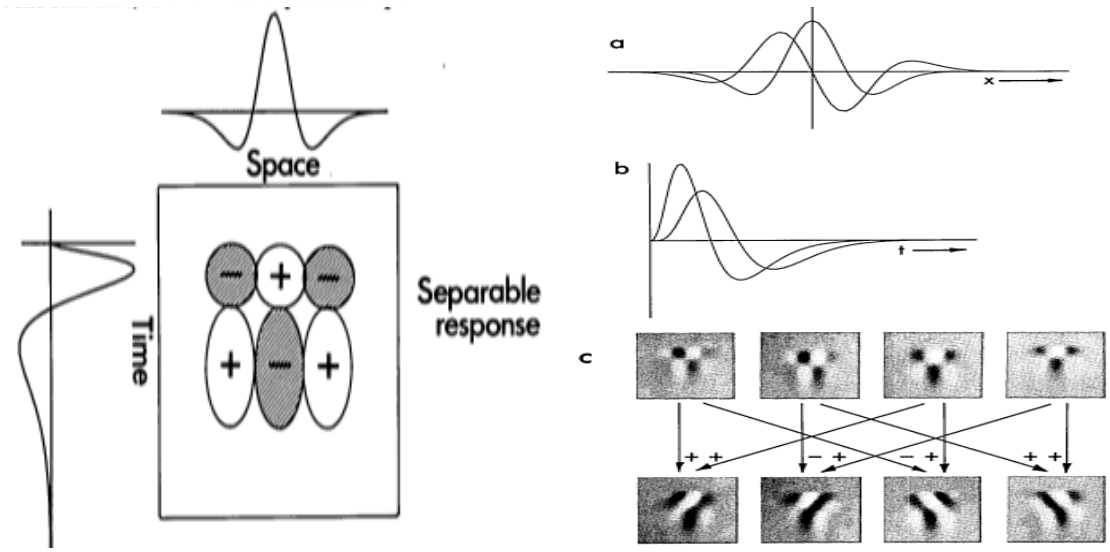

Figure 2.3: **Motion detector units in motion energy model**, figure is adapted from (Adelson and Bergen, 1985): **Left)** a motion detector unit can be built by illustrated spatial and temporal impulse responses. The resultant spatiotemporal transfer function of unit includes different positive and negative lobes which are separable. **Right)** two spatial (**a**) and two temporal (**b**) functions which are different from each there in phase and temporal shift are used to build spatiotemporal filters illustrated in (**c**). Different direction selective filters can be built by combination of these filters.

This motion energy model, by dedicating spatiotemporal transfer function to each unit, develops a network of specialized oriented detectors which are sensitive to the direction of motion, similar to simple cells in the early visual cortex. This model is matched to properties of the human visual system and has been successful to account for a vast body of physiological and psychophysical data and is used as a reference model visual motion detection.

Motion and correspondence problem: Physical displacement of objects in the visual scenes can be regarded as difference in the position of objects in consecutive discrete frames. Then detecting the motion will be a problem to solve: which features of a scene are most representative and useful to be tracked and matched between each two temporal steps? This problem is known as the ***Correspondence problem***. For example a movie is a sequence of discrete positions in discrete times which leads to ***Apparent motion*** and is perceived coherently. The correspondence problem in this example is to match some visual features like edges and corners constantly and infer a global velocity.

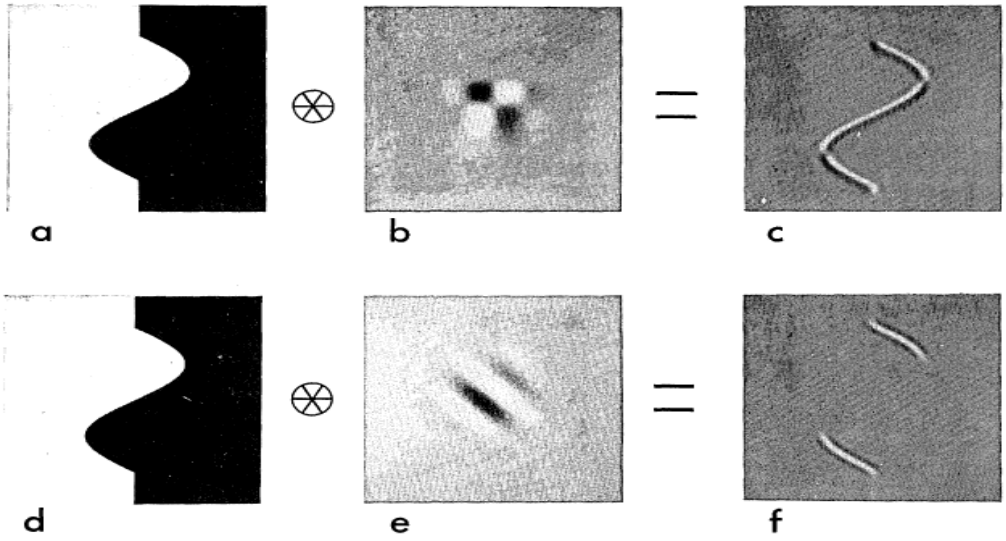

Figure 2.4: **Sinusoidal motion detection**, figure is adapted from (Adelson and Bergen, 1985): an example of sinusoidal motion (rightward-leftward-rightward) detected by **(top row)** spatiotemporal separable unit in Fig 2.3, **(bottom row)** spatiotemporal motion energy unit in Fig 2.3.

In various motion detection models different matching criteria have been proposed to solve the correspondence problem. ***Temporal coherency of motion*** or ***Motion coherence theory*** is one of them. In Fig 2.5 an example of correspondence problem is shown along with estimated velocity field of motion with motion coherency theory.

2.1.2 Temporal coherency of motion as a neural and computational constraint

Models of visual motion have been progressed in the direction of accounting for global motion perception, given local detected motion signals. One of the most important research lines to this aim is based on spatial integration of local motion with a large scale constraint known as ***Motion coherence theory***. This hypothesis assumes that moving objects in nature most probably travel through smooth trajectories and rarely make sharp turns, therefore motion is assumed to be temporally coherent.

Many important visual phenomena recorded by psychophysical experiments are explained with this theory. Yuille and Grzywacz (1989) proposed a mathematical analysis of this theory and modeled the optimal velocity field accounting for motion capture, aperture problem and motion cooperativity. In their mathematical framework, smoothness of motion is implemented by minimizing a cost function on estimated velocity field. They argued computational advantage of smooth velocity giving rise to perception of coherent motion.

Motion cooperativity is an example of estimated velocity field for short range motion and is explained by motion coherence theory. In motion cooperativity experiment, stimulus is set of several moving dots in random directions and at each time step the directions are chosen from

a uniform distribution with a slight bias on one specific direction (See Fig 2.6). According to psychophysical evidences, subjects are able to perceive a global motion in the direction of bias. Mathematical framework shown in equation set 2.1 estimates the correct velocity field for this experiment:

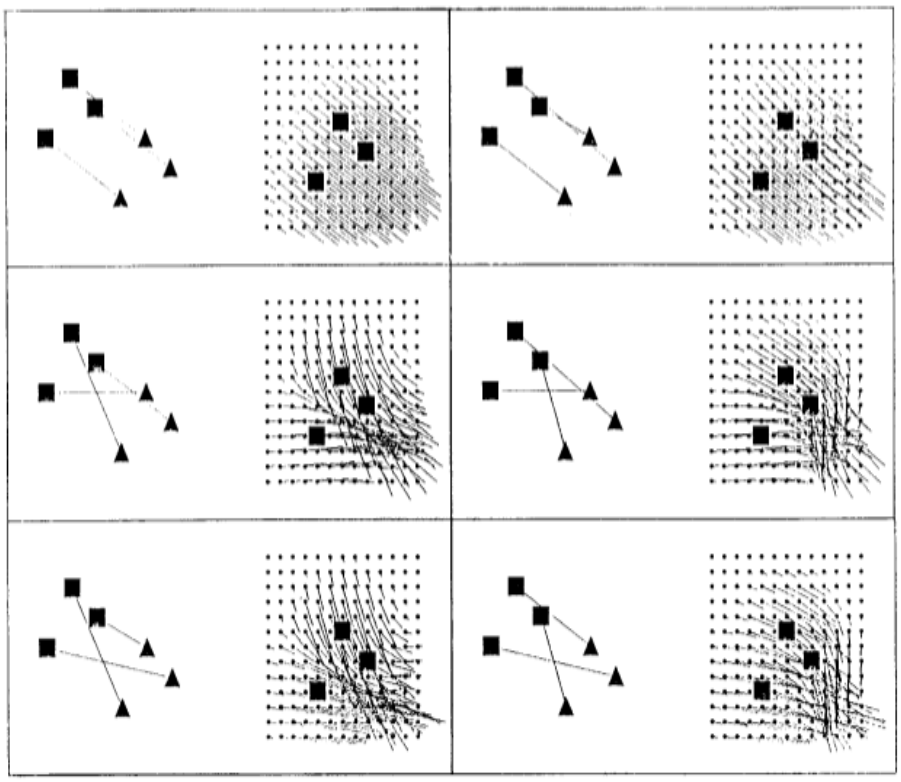

Figure 2.5: **Correspondence problem**, figure adopted from (Yuille and Grzywacz, 1989): correspondence problem solved by long range constraint of motion coherency. Squares and triangles match to 3 features of moving scene in frames number n and $n+1$. Out of all six possible combinations, the match proposed by motion coherency theory is consistent with the inference of the visual system.

$$E(\vec{V}(\vec{r}),\vec{U}_i) = \sum_i \ [\vec{V}(\vec{r_i}) - \vec{U}_i]^2 - \lambda \int \sum_0^\infty \ c_m D^m(\vec{V})^2$$

$$\vec{V}(\vec{r}) = \sum_i \ \frac{\beta_i}{2\pi\sigma^2} \ \exp\frac{-(\vec{r}-\vec{r_i})^2}{2\sigma^2} \tag{2.1}$$

$$(\lambda\delta_{ij} + G_{ij}) \ \ \beta_j = U_i \ \ , \ \ G_{ij} = \frac{1}{2\pi\sigma^2} \ \ \exp\frac{-(\vec{r_j}-\vec{r_i})^2}{2\sigma^2}$$

Where $\vec{V}(\vec{r_i})$ and $\vec{U}_i$ are respectively estimated velocity field and measured velocity of image at the position of the i_{th} dot. Then c, λ and σ are constant values.

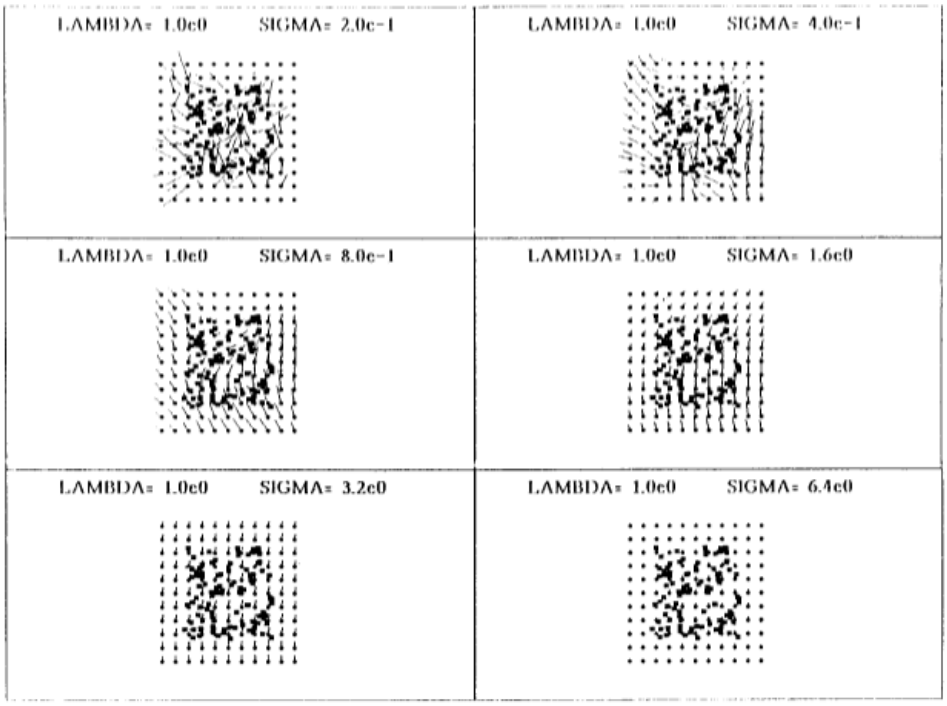

Figure 2.6: **Motion cooperativity phenomena**, figure adopted from (Yuille and Grzywacz, 1989): estimated velocity field for different values of σ and λ in equation set 2.1 . Hundred moving dots are placed in a square, their horizontal velocity is chosen from a uniform distribution between 1 and -1 and vertical velocities are chosen from the same uniform distributions which is slightly biased for 0.25. For certain values of σ and λ estimated velocity field shows a global upward velocity matched to the psychophysical results.

In another study, Grzywacz, Watamaniuk, and McKee (1995) have implemented a neural network based on the same theory which detects coherent trajectory of a signal dot among many noise dots. Relevant psychophysical experiments have been done by Watamaniuk, McKee, and Grzywacz (1995b) and the results illustrates ability of human subjects in detection of signal dot moving a pool of noise dots with brownian motion. The proposed model reports global response out of network which smoothly varies over time and over characteristics of neurons including preferred speed and direction, center and radius of their receptive field. This is reflected the equation 2.2:

$$\begin{aligned} E(t) \;=\; & \sum_{\vec{r},\vec{u},s,\lambda} R_l(t:\vec{r},\vec{u},s,\lambda) - R_c(t:\vec{r},\vec{u},s,\lambda) \\ & + \int_{\vec{r},\vec{u},s,\lambda} \psi_r (D_r R_c)^2 + \psi_s (D_s R_c)^2 + \psi_u (D_u R_c)^2 + \psi_\lambda (D_\lambda R_c)^2 \\ & + \int_{\vec{r},\vec{u},s,\lambda} \psi_t \int_{\acute{s}} W(\acute{s}) \Big(\frac{\partial R_c}{\partial t} + \acute{s} \bigtriangledown R_c \cdot \vec{u} \Big)^2 \end{aligned} \tag{2.2}$$

Where R_c and R_l are respectively estimated and measured response of network and $E(t)$ is the energy criteria to be minimized. $\vec{r}, \vec{u}, s, \lambda$ are respectively center of receptive field, preferred direction, preferred speed and size of receptive field. ψ_xs are corresponding constant to determine relative importance of each term and D_x is differential operator.

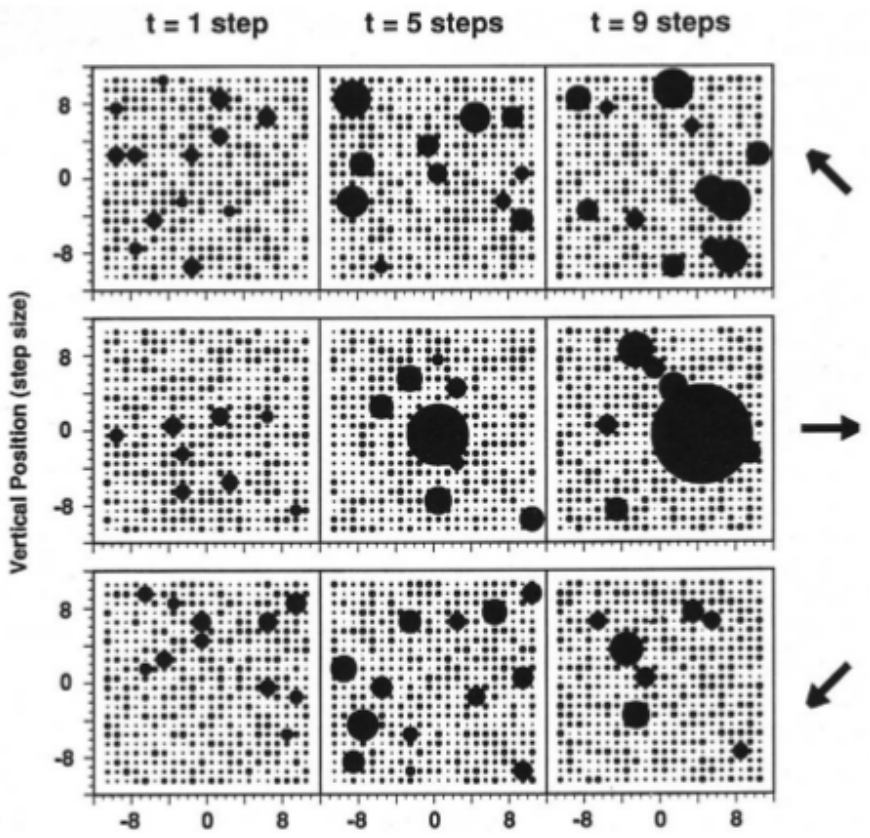

Figure 2.7: **Motion coherency in a neural network**, figure adopted from (Grzywacz, Watamaniuk, and McKee, 1995): neural network response to a horizontally moving dot in a pool of noise dots under brownian motion. Network includes 22×22 cells and response has been recorded from three population of cells with different preferred direction at three time steps. The diameter of dark blobs is proportional to the response.

In this network, local motion signals are made coherent in space and time and give rise to detection of coherent trajectories. Fig 2.7 illustrates gradual coherence detection of network for a signal dot moving rightward in a 22×22 size network with a dense background of randomly moving dots.

2.1.3 Temporal coherency of motion and probabilistic motion estimation

Extending previous studies on temporal coherency of motion, Burgi, Yuille, and Grzywacz (2000) proposed a probabilistic framework for motion estimation. In this model, motion is defined as a time varying signal including position and velocity components and is represented probabilistically. Taking advantage of predictability of smooth motion trajectory, their probabilistic motion

estimation theory utilizes Bayesian generalization of standard Kalman filter in order to integrate visual motion.

In earlier works from the field of computer vision, it is common to demonstrate temporal coherence of successive states of image luminance with a Markov chain and define motion estimation problem in a Kalman recursive linear state estimator. Standard Kalman theory ensures optimal combination of **Prior** knowledge on behavior of system and instantaneous observed data known as **Likelihoods**. However, the Kalman theory makes the hypothesis of Gaussian distributions for modeling variations in observed data and transition of successive states. Thus, it can not fully account for psychophysical evidences on motion occlusion and motion outliers or more complex scenes including multiple moving objects.

The Generalized Kalman filter proposed by Burgi, Yuille, and Grzywacz (2000) is a solution for this limitations by applying non Gaussian probability distributions for nonlinear state estimations. By emphasizing the temporal coherency of motion and smoothness of trajectories, the theory also recruits **Prediction** as a powerful mechanism in temporal grouping of visual data and to interpret the moving scenes. The estimated motion finally is composed of predicted state updated by observed data.

In a most simplified Bayesian form, temporal coherence of trajectory would play the role of **Prior** information in initial state to be corrected by observed **Likelihoods** in the next step. Equations 2.3-2.4 respectively illustrate estimation and prediction steps of the algorithm, where $\vec{x_k}$ and $\vec{z_k}$ are state vector and observation vectors and $Z_k = \vec{z_0}, \vec{z_1}, ..., \vec{z_k}$ includes all observed history of the trajectory. State estimation for future step employes Z_k for predicting and then $\vec{z}_{k+1}$ to correct the estimation.

$$P(\vec{x}_{k+1}|Z_{k+1}) = \frac{P(\vec{z}_{k+1}|\vec{x}_{k+1})}{P(\vec{z_{k+1}}|Z_k)} \;\; P(\vec{x}_{k+1}|Z_k) \tag{2.3}$$

$$P(\vec{x}_{k+1}|Z_k) = \int P(\vec{x}_{k+1}|\vec{x}_k) \;\; P(\vec{x}_k|Z_k) \; d_{\vec{x}_k} \tag{2.4}$$

This Bayesian formulation can be rewritten by definition of motion as an update in the flow field on image during time step δ shown by $\vec{v}(\vec{x}, t)$. Then $\phi(\vec{x}, t) = \{\phi_1(\vec{x}, t), \phi_2(\vec{x}, t), ..., \phi_M(\vec{x}, t)\}$ illustrates local observation all over image at time t and equations 2.3-2.4 turn to the following equations:

$$\textit{Estimation:} \quad P((\vec{v}(\vec{x}, t)|\Phi(t)) = \frac{P[\vec{\phi}(\vec{x}, t)|\vec{v}(\vec{x}, t)]}{P[\vec{\phi}(\vec{x}, t)|\Phi(t - \delta)]} \;\; P[\vec{v}(\vec{x}, t)|\Phi(t - \delta)] \tag{2.5}$$

$$\textit{Prediction:} \quad P[(\vec{v}(\vec{x},t)|\Phi(t-\delta)] = \int P(\vec{v}(\vec{x},t)|(\vec{v}(\vec{x},t-\delta))\ P(\vec{v}(\vec{x},t-\delta)|\Phi(t-\delta))\ \ d_{\vec{v}(\vec{x},t-\delta)} \tag{2.6}$$

The main computational advantage of the theory shown in equations 2.5-2.6 is the spatial factorability of probability distributions. It means that **Prior** and **Observation** distributions are regarded as independent coherent state transitions and local observations allover the image as illustrated by equations 2.7-2.8 .

$$P_p(\vec{v}(\vec{x},t)|\vec{v}(\vec{x'},t-\delta)) = \prod_x p_p(\vec{v}(\vec{x},t)|\vec{v}(\vec{x'},t-\delta)) \tag{2.7}$$

$$P_l(\phi(\vec{x},t)|\vec{v}(\vec{x},t)) = \prod_x p_l(\phi(\vec{x},t)|\vec{v}(\vec{x},t)) \tag{2.8}$$

Factorability assumption in formulation of motion estimation implies parallel computation of motion and is derived from the independence of these measurements in each spatial point of image for the cost of a reduction of spatial coherency. Burgi, Yuille, and Grzywacz (2000) also have implemented this theory in a neural network with two layers for priors and observations of motion. At each image pixel, the network includes banks of receptive fields tuned for all velocities. Therefore at each spatial position two neural populations are coding for likelihoods and priors having special tuning curve to respect temporal coherency demonstrated by equations 2.3-2.8. The network implements state transitions in position and velocity based on prior on temporal coherency of motion with a multiplicative modulation of likelihood functions by observation layer (See Fig 2.8).

Fig 2.9 depicts network response to two well known psychophysical experiments: motion outlier and motion occlusion. In outlier experiment distractor dots are not moving coherently enough to gain confidence in motion estimation.

The probabilistic motion estimation network of Burgi, Yuille, and Grzywacz (2000) is an efficient implementation of temporal coherency theory. Bayesian essence of network using motion prediction and motion observations gives rise to emergence of global motion direction based on stochastic local computations.

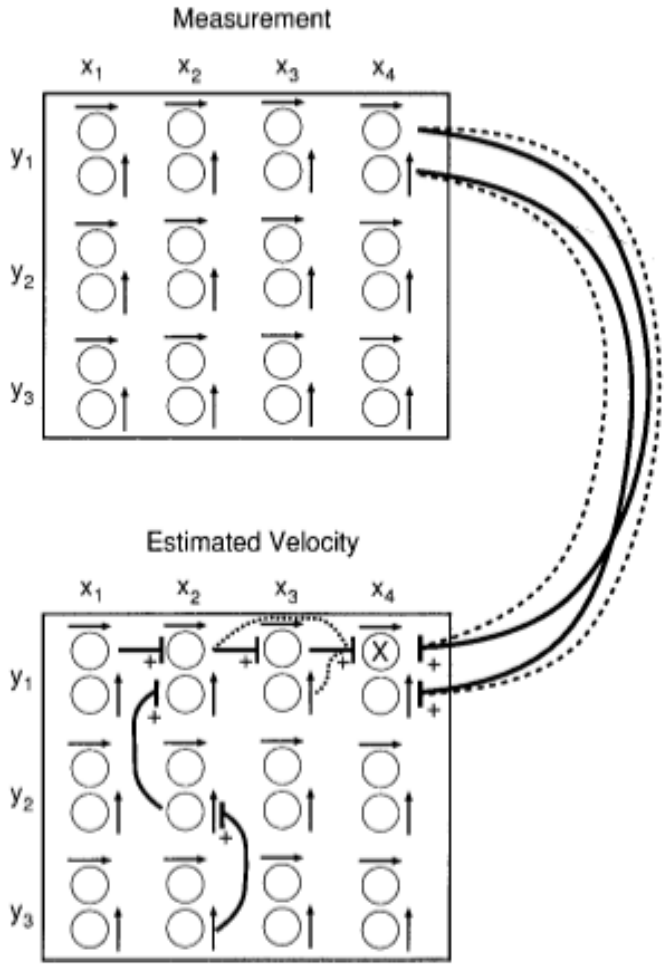

Figure 2.8: **Structure of a neural network based on temporal coherency of motion**, figure adopted from (Burgi, Yuille, and Grzywacz, 2000): neural network implemented by Burgi, Yuille, and Grzywacz (2000) for motion estimation. Network includes two distinct layers for measurement (observations) and estimation. 12 spatial positions have been shown in the $x - y$ plane and each position contains cells tuned for various velocities (only horizontal and vertical velocities are indicated in the figure) measurement layer affects estimation by strong multiplicative excitatory connections between cells with similar velocity preference and weaker ones between less similar cells. Importantly, trajectory prediction is implemented in estimation layer, where cells with similar velocity preference excite each other in row.

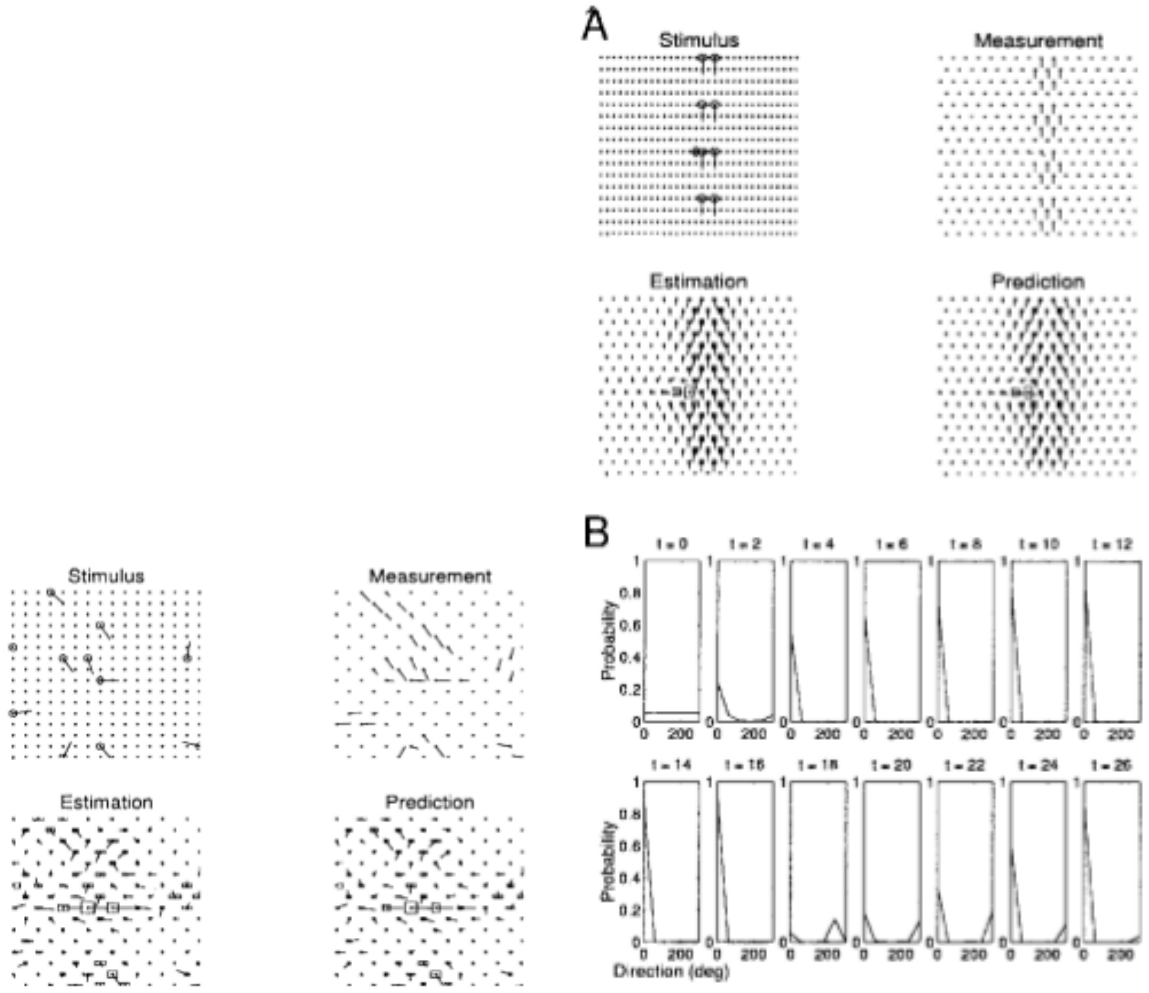

Figure 2.9: **Motion outlier and motion occlusion experiments in network model** of Burgi, Yuille, and Grzywacz (2000). **Left)** network response to motion outlier, a signal dot moving rightward in a pool of distractor dots under brownian motion. Unlike distractor dots, trajectory of signal dot has enough coherency to be detected by network. **Right) A:** stimulus and simulated response to motion occluder. Stimulus is shown for the time step just before entering to the occlusion area ($t = 18$) and occluder illustrated by downward moving dots. As shown in **B**, during occlusion time the strong peak of estimated direction at 0° starts to fade out and another peak is developed around 270° (motion direction of occluder dots)

2.2 Motion-based position coding: models

As reviewed in previous section, the systematic relation of position coding with the history of ongoing motion is supported by various experimental evidences. A very typical experimentation which reveals the difference in position coding of stationary flashed and moving stimuli is flash lag illusion (FLE) that has been repetitively studied from many behavioral and physiological aspects. Reproducing FLE in different experimental efforts usually has been accompanied with new insights into the delay compensation mechanisms of the brain.

Motion based coding of position is one of candidate mechanisms for FLE and is known as motion extrapolation after proposition by Nijhawan (1994). The competency of this hypothesis in explaining a vast body of behavioral and experimental data keeps it under continuous verification leading to design of new models and experiments to find its limitations.

In chapter 1 we reviewed main experimental studies in the literature and here we will go through proposed models based on this hypothesis. Motion extrapolation is observed as early as the retina by recording firing rate of ganglion cells (Berry et al., 1999), and then it is supported to be implemented in striate (Jancke et al., 2004) and extra striate cortex (Maus, Fischer, and Whitney, 2013; Maus et al., 2013) and also at behavioral levels. As all neural populations covering the visual path from retina to higher areas are of various cell types and network properties, there are few proposed models, accounting for specific data of some studies and here we will review a brief summary of them.

2.2.1 Internal models of visual perception

In a network model of inhibitory and excitatory cell populations, Erlhagen (2003) studied position code of moving stimuli versus a brief localized flashing target. The model, implements a mean field representation of network excitation and inhibition and proposes an efficient internal model for representation of moving objects. This self stabilized internal model would accelerate processing of moving objects by accumulation of sub-threshold excitatory activities and pre-shaping the neural response in the positions which are not yet touched by the stimulus. The notion of internal model is consistent with bottom-up and top-down contributions to motion processing.

As such, the coded position for moving object at each time step would be a combination of bottom-up sensory input and top-down internally represented input. Any mismatch between these two input sources would be sent back to higher areas to correct the internal model. Functional advantages of having an internal model for motion perception is explored specially for a motion occlusion and FLE like experimental setups (See Fig 2.10, Fig 2.11 and Fig 2.12).

Position field simulations suggest coherent motion presentation whenever there is an interruption or occlusion in the sensory input. In this context, as demonstrated by FLE, neural response to moving objects can be regarded as spatiotemporal overlapping wave patterns; which prime activation of next positions in the trajectory. While a brief localized flash stimulus lacks such a smooth state transfer.

This model, being consistent with the approach of Nijhawan (1994), emphasizes the competence of the internal model in accounting for motion induced spatial extrapolation of moving objects, and also for anticipatory neural responses to overcome neural transfer delays. In addition, this framework introduces a common candidate mechanism for motion processing at early and middle trajectory parts, by simulating the Fröhlich effect and FLE. (See Fig 2.13).

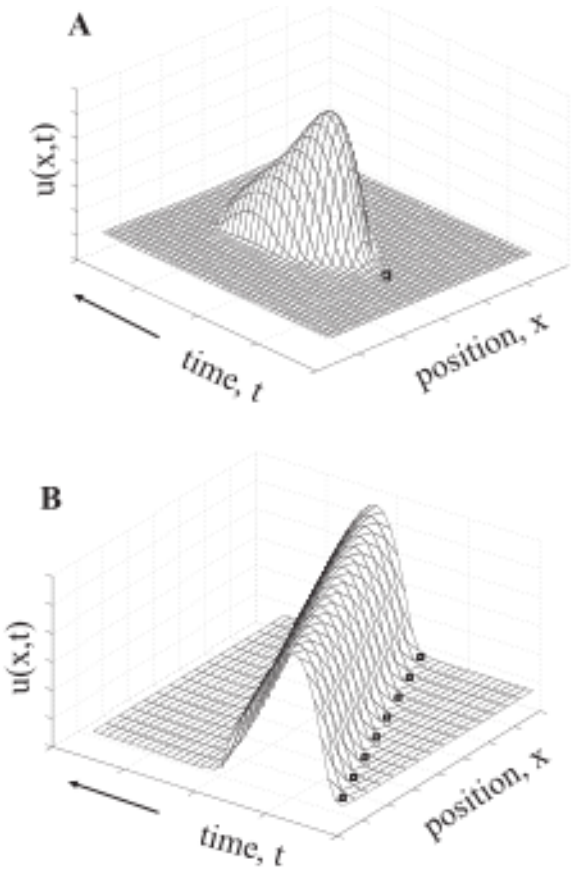

Figure 2.10: **Internal model of visual motion**, figure adopted from (Erlhagen, 2003). Spatiotemporal response properties in the field model **A)** $u(x,t)$ is excitatory field response for a localized flash response, the black square shows center of stimulus. The response increases to its maximum and then fades away. **B)** $u(x,t)$ is excitatory field response for a moving stimulus. Response lags behind center of stimulus shown by black square

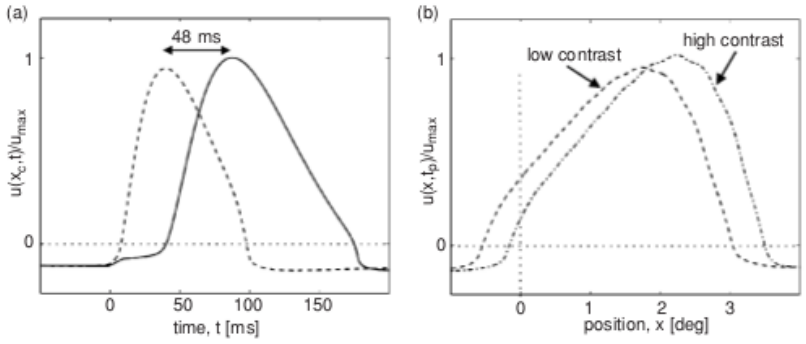

Figure 2.11: **model simulation of FLE in the field model of** Erlhagen (2003). **A)** Temporal evolution of excitatory field response in the position of flash. Response to the flash (solid line) lags behind the response of the moving object (dashed line). **B)** Dependency of FLE to contrast simulated in the model.

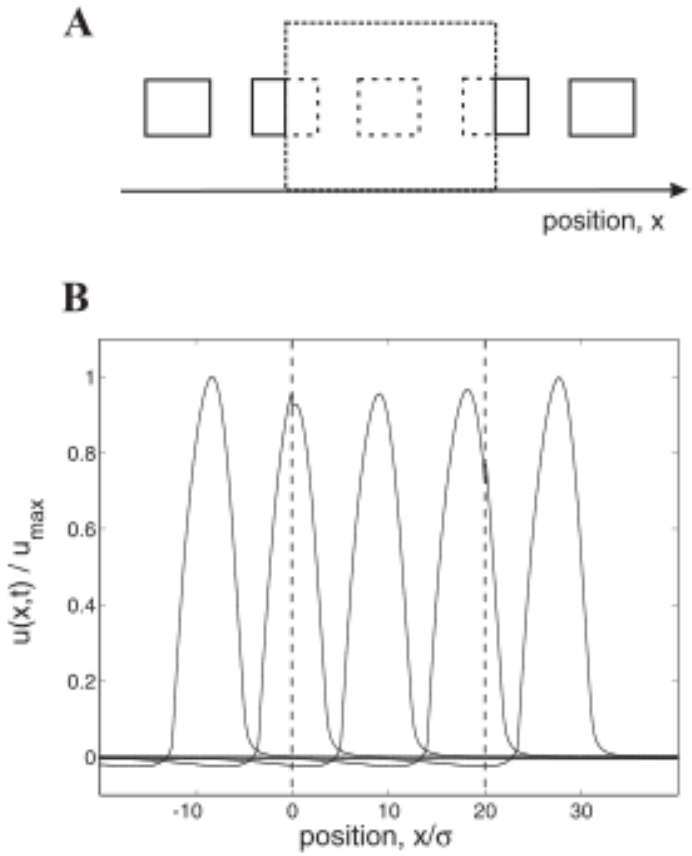

Figure 2.12: **Motion occlusion in the field model**, figure adopted from (Erlhagen, 2003). **A)** Schematic demonstration of motion occlusion. **B)** Temporal continuity of excitatory field response to occluded stimulus.

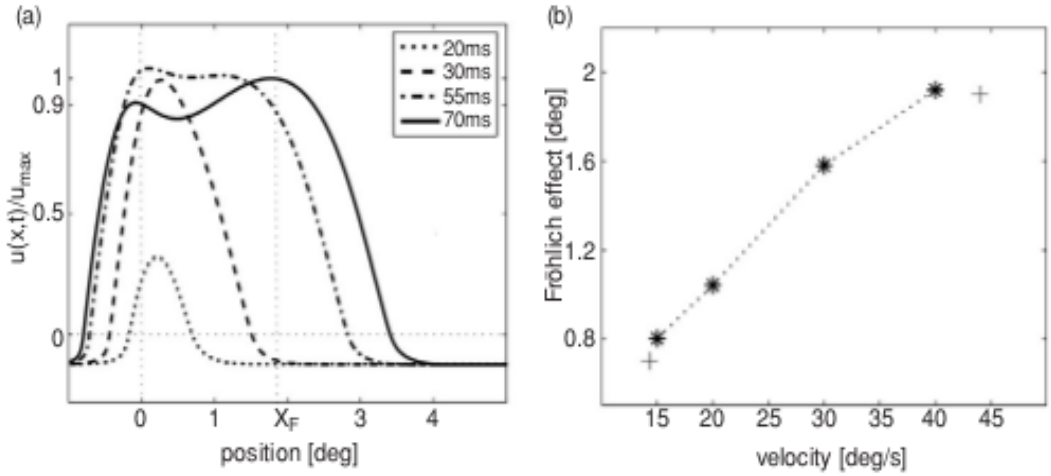

Figure 2.13: **Model simulation of the Fröhlich effect**, figure adopted from (Jancke and Erlhagen, 2010): **A)** . Four successive snapshots from development of excitatory field at the begining of trajectory. **B)** Dependence of Fröhlich effect to speed. Model results and experimental data are respectively shown by pluses and asterisks.

2.2.2 Model of motion anticipation in retina

Berry et al. (1999) proposed a model for quantitative description of light response of ganglion cells, along with their experimental evidences on anticipatory response in retina. To study the sensitivity of anticipation to contrast of stimulus, they recorded the spatial profile of anticipatory response by ganglion cell populations (See Fig 2.14).

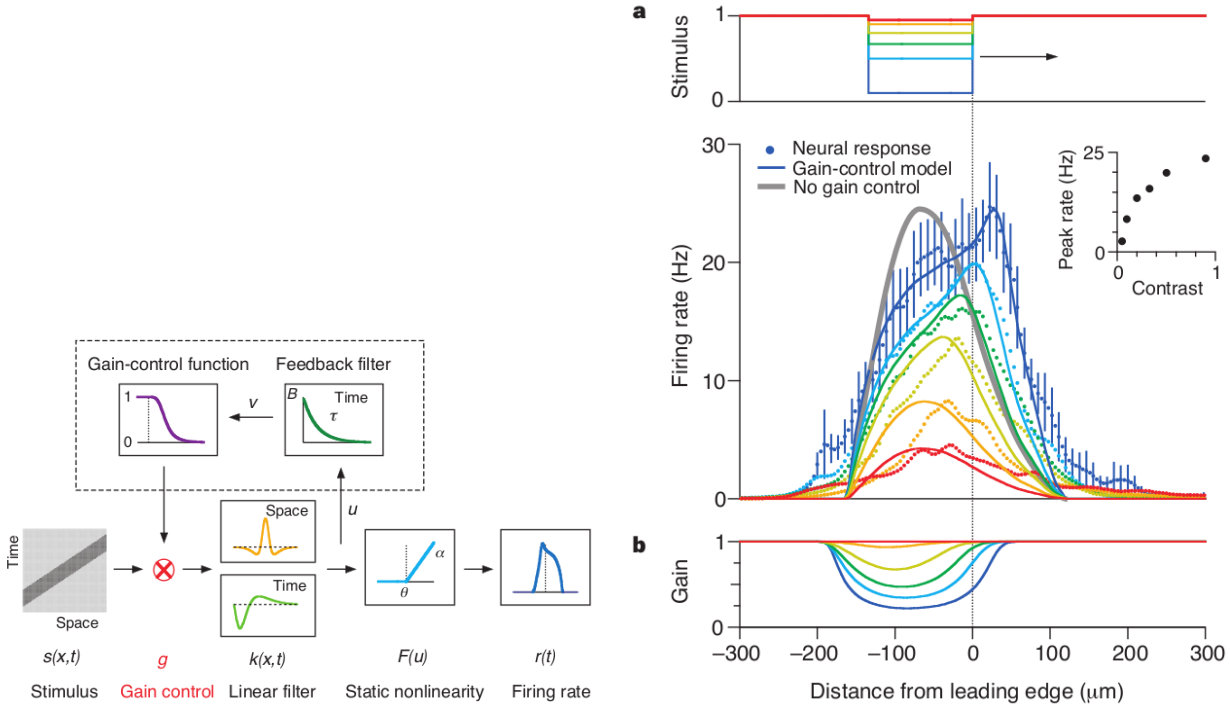

Figure 2.14: **Gain control model for motion extrapolation in the retina**, figure adopted from (Berry et al., 1999). **Left)** gain control model for light response of ganglion cells in the retina. The model is described in equations 2.9 - 2.10. The output of a spatiotemporal filter will be applied to an exponential and gain control filter, then firing rate will be generated after getting rectified with a nonlinear function. **Right)** contrast response of ganglion cells along with model response, with and without gain control. Contrast range is shown in **a** and traces in **b** correspond to gain control at each contrast condition.

Their data illustrates that at high contrasts anticipatory signature is more evident, causing the peak of spatial profile to be ahead of leading edge of stimulus. This signature is described with a contrast gain control model: receptive field of ganglion cell is represented with a weighting function $k(x,t)$ which integrates the input light in space and time and determines the firing rate. The model stated by following equations includes a negative feedback. That is activated in case of having a strong stimulation in long time, and reduces the gain and controls the response for the continuation of stimulus stream.

$$u(t) = g(v) \int_{-\infty}^{\infty} dx \int_{-\infty}^{t} d\acute{t}\ s(x, \acute{t})\ k(x, t - \acute{t}) \tag{2.9}$$

$$v(t) = \int_{-\infty}^{t} d\acute{t}\ u(\acute{t})\ B\ exp\ \left(-\frac{t - \acute{t}}{\tau}\right) \tag{2.10}$$

$$g(v) = \begin{cases} 1 & v < 0 \\ \dfrac{1}{1 + v^4} & v > 0 \end{cases}$$

$$F(u) = \begin{cases} 0 & u < \theta \\ \alpha\ (u - \theta) & v > \theta \end{cases}$$

2.2.3 Diagonal neural pathway model for delay compensation

In a recent work on the role of sensory prediction in delay compensation, Nijhawan and Wu (2009) updated the model of Berry et al. (1999) with more emphasis on the modulatory effect of starburst amacrine cells on ganglion cell populations. The model (so called diagonal model) in the level of mechanism unifies two well know accounts for delay compensation and FLE: motion extrapolation approach and the account based on different processing delays for moving and stationary objects. The theory proposes diagonal neural pathways between retinotopic positions in retina and cortical positions corresponding to the slightly ahead positions in the trajectory of moving object (See Fig 2.15). Thus, cortical neurons in the path of motion (which are predicted to be stimulated soon) are primed and the response latency is reduced. The slope of the diagonal pathway would match to the distance that stimulus is predicted to be traveling during delay latency. Theoretically, the neural populations corresponding to the actual position of stimulus would respond with a significantly reduced delay.

The model differs from the model of Berry et al. (1999) in its more detailed physiological description by the association of gain control function with the activity of amacrine startburst cells. Indeed, the very particular structure of starburst amacrine cells with a small soma and dense dendritic tree emits an asymmetric inhibitory current field toward ganglion cells. For a rightward stimulus, asymmetric inhibition depresses more ganglion cells that are in left hand side of actual position of stimulus than the ones in the right hand side. Therefore, gain control function dedicates more gain to the cells situated slightly ahead of stimulus and shifts the spatial profile of response in the direction of motion (See Fig 2.16). This physiological mechanism is consistent with the diagonal activation of future cortical positions to encode the actual position of stimulus and to compensate for transfer delays.

The diagonal model is a comprehensive explanation of the motion extrapolation mechanisms started in retina. It is consistent with physiological signatures and anatomical properties of retina as well as being a more generic computational mechanism which is likely to be implemented in other cortical areas for the aim of delay compensation and sensory prediction.

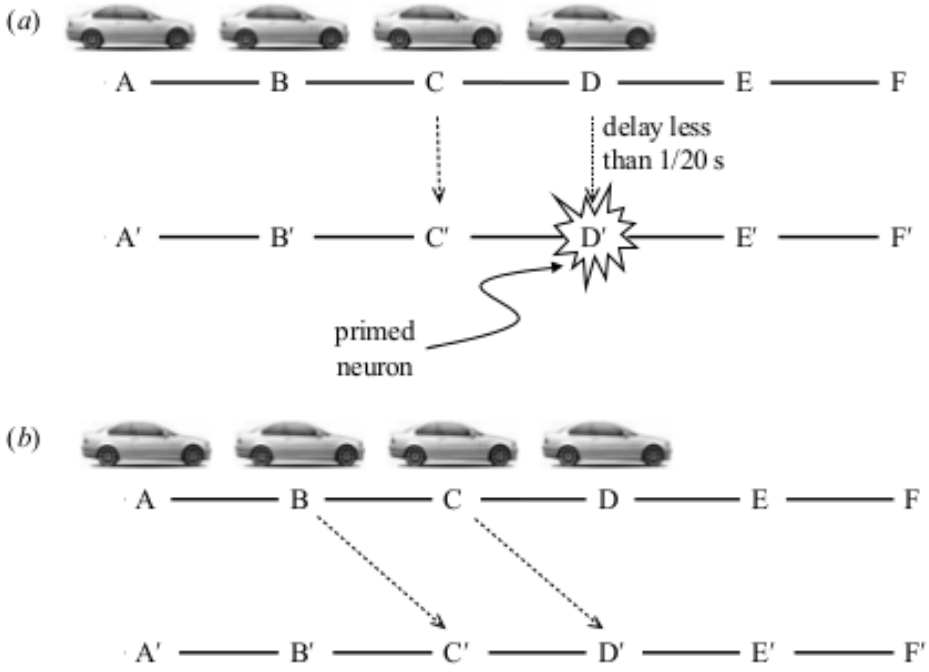

Figure 2.15: **Diagonal model of motion extrapolation**, figure adopted from (Nijhawan and Wu, 2009): Comparison of DL (different latencies for moving and stationary objects) and diagonal motion extrapolation model. A, B, C,..., and $\acute{A}$, $\acute{B}$, $\acute{C}$, ... respectively represent retinal and cortical positions of stimulus. **a)** DL model illustrates the reduction of transfer delay by priming the actual position of stimulus via lateral connections in the cortical area. **b)** diagonal model illustrates activation of actual position of stimulus by spatial shifting of activity in the retinal layer.

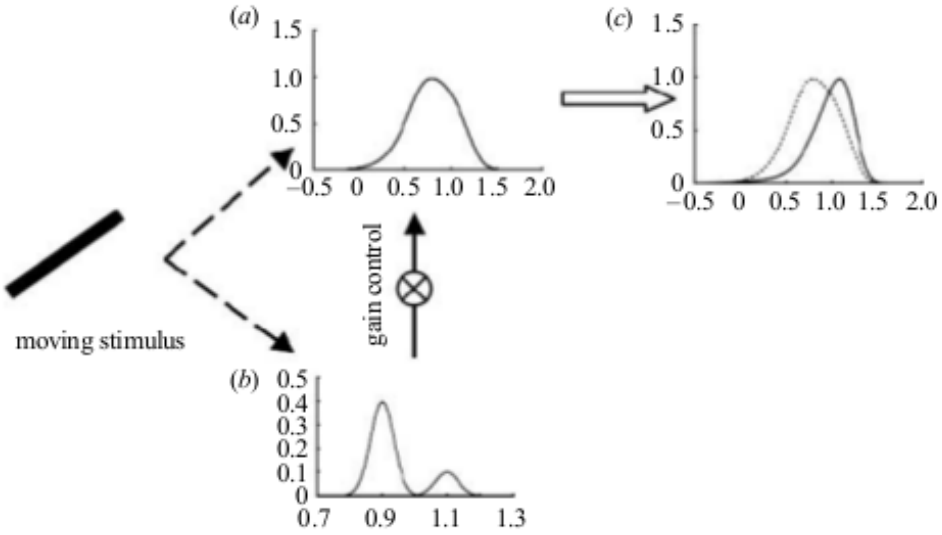

Figure 2.16: **The implementation scheme of diagonal model in the retina**, figure adopted from (Nijhawan and Wu, 2009). **a)** illustrates the symmetric spatial profile of response in ganglion cell to a horizontally moving stimulus, before modulatory effect of starburst amacrine cells. **b)** illustrates the asymmetric inhibition imposed by starburst amacrine cells to ganglion cells. By gain control and inhibition the symmetric response turns into the asymmetric one shown in **c**.

2.2.4 Neural network models of FLE

In a simple feedforward network of leaky integrate and fire neurons Baldo and Caticha (2005) have simulated FLE and various psychophysical attributes of that. The model does not explicitly implement a specific theory and FLE arises from spatial advancement of activity in a network of lateral connections by excitatory and inhibitory neurons. This model emphasizes top down facilitation of neurons and facilitations generated by spatial interactions along a pathway, being consistent with motion extrapolation account. This study has not investigated the influence of velocity coding in spatial facilitations.

Other neural network models have addresses the question of neural delays and motion extrapolation at single neuron level (Lim and Choe, 2006; Lim and Choe, 2008). In these models, neurons are sensitive to the rate of change in the input and via extrapolative activations they estimate the state of the external world at $time = t + \Delta t$ instead of $time = t$. Facilitatory activity is derived from the present and past activity of network and synaptic efficacy implements kind of smoothness constraint and coherency in the spiking activity. Therefore spiking activity is extrapolated in the direction of change via STDP learning rule along with facilitated synapses.

Summary

In this chapter, we have reviewed the basic models of visual motion estimation. In the first section we have focused on the classic motion energy model of Adelson and Bergen (1985), then we have introduced the modeling studies based on temporal coherency of motion and their application to well known problems in visual processing: correspondence problem, motion cooperatively phenomena, motion outliers and motion occlusion.

In the second section we have focused on the the models consistent with motion-based prediction theory, including an internal model of visual perception with a demonstration of FLE, A model for motion anticipation in the retina, diagonal model for delay compensation and some relevant neural network models.

Results of this thesis provides a generic computational framework to study the influence of motion-based information on position coding of moving object. In the next chapter, we will describe the basic methods of our study as well as computational details of the Motion-based prediction (MBP) model.

Part II

Model and methods

Chapter 3

Motion-based prediction

Abstract

In this chapter, we will first briefly review Bayesian motion estimation and Kalman filtering as the most common frameworks in studying visual motion prediction. Then we will introduced the CONDENSATION algorithm (Isard and Blake, 1998) as a generalization of Kalman filter and a powerful particle filtering method in studying visual motion. Then we will elaborate our motion-based prediction model from Bayesian master equations to the implementation details. Finally we will discuss the novelty of the current implementation for the objectives of this study.

3.1 Predictive coding in motion detection

From the point of view of evolution, there are two most fundamental abilities for all species to survive: first they need to precisely locate preys to hunt, and second they must have efficient reactions against predators. These basic survival conditions can be stated in a motion estimation framework, as successful movements of the animal toward targets or away from enemies is rooted in precise information about their instantaneous location.

Performance of neural systems in fulfilling these tasks is in conflict with the existence of various sources of noise or delay in the delivery of information from physical world to processing areas. This issue has attracted great research efforts and has led to development of a research line as motion prediction. Motion prediction theory assumes that visual system takes advantage of motion history to be prepared for processing of the most expected locations of moving objects. Functional benefits of such a predictive system are validated by a wide range of neuronal and behavioral evidences (See chapter 1), giving rise to development of relevant solid theoretical frameworks. Bayesian motion estimation and Kalman filtering approaches are the most dominant theories in motion prediction research.

In this section, we will briefly review the theoretical principles of predictive coding in motion detection, as a base for our model and methods described in Section 3.2.

3.1.1 Bayesian inference and Markov chain

Theories based on considering the brain as a Bayesian machine have grown fast during the last decades, mainly because of the intuitive explanation that they provide, about the neural structure, sensory processing and decision making. Central simplification in these models comes by assuming that regularities of the external world, known by physical rules, are reflected in the structure and processing patterns of the brain. This regularities are called ***prior*** informations and this assumption leads to studying neural functions in a reduced and well defined state space, composed of the most expected ones. Then the ultimate inferred ***posterior*** is created by combination of priors and the most recent measurements from the world. Equation 3.1 summarizes the Bayesian inference for an unknown parameter, x:

$$posterior(x) = prior(x) * likelihood(x) \tag{3.1}$$

Another flexibility of Bayesian formulation of neural signatures is a probabilistic description of state variables. According to neural recordings, different trials of the same stimulation will never produce exactly the same responses, but they are rather distributed in a range definable with statistical parameters.

Bayesian theory has been widely used to describe various behavioral and physiological data sets in motion prediction. In some cases, Bayesian inference is applied to reproduce a specific data set in physiology or behavior (Bogadhi et al., 2011; Hürlimann, Kiper, and Carandini, 2002; Montagnini et al., 2007). These studies, for instance, are useful to examine the effect of variously defined priors or to provide a better understanding from the way that the brain infers motion.

In another category of models, like the the network model of Burgi, Yuille, and Grzywacz (2000), Bayesian inference is implemented in a purely theoretical framework, based on temporal coherency of motion. Another elegant Bayesian model of motion estimation describes the posterior velocity of image motion based on the prior information on luminance conservancy of the image, which is associated with temporal coherency of motion (Weiss and Fleet, 2001).

In the current study, we also have temporal coherency of motion as the hypothesis but implemented in a different framework. As we reviewed in chapter 1, this hypothesis is suitable to explore more generic and large-scale properties of visual motion processing, by simulating well known visual experiments. It also allows various experimental assessments, which may be hard to test in real experimental setups.

Bayesian motion estimators are implemented as ***Markov chain***, where every state includes the instantaneous position and velocity of the moving object. In the Markov chain, motion is dynamically estimated, based on likelihoods and most expected motion states. This expectancy is determined by prior information about smoothness of motions in the nature. The achieved posterior distribution represents the estimated motion at a specific moment. This dynamical

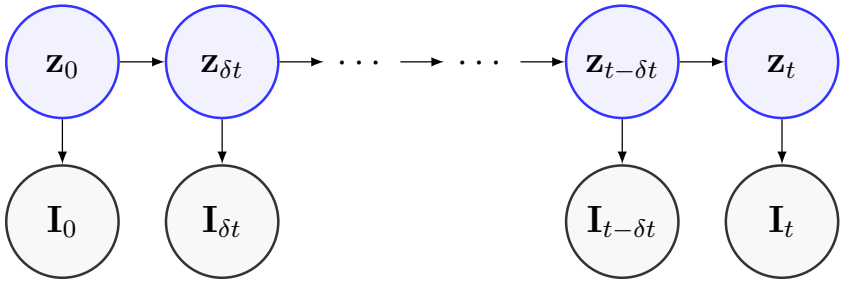

Figure 3.1: **Classic Markov chain for state transition of system**: in the current study, the estimated state vector is $z_t = \{x_t, y_t, u_t, v_t\}$ composed of position and velocity of moving stimulus and instantaneous measurements from the image are represented by I_t

motion estimation has been illustrated in Fig 3.1 and is summarized in following equations:

$$\textit{estimation:} \quad p(z_t|I_{0:t}) \propto p(I_{t-\delta t:t}|z_t) \cdot p(z_t|I_{0:t-\delta t}) \tag{3.2}$$

$$\textit{prediction:} \quad p(z_t|I_{0:t-\delta t}) = \int dz_{t-\delta t}\ p(z_t|z_{t-\delta t}) \cdot p(z_{t-\delta t}|I_{0:t-\delta t}) \tag{3.3}$$

Where $z_t = \{x_t, y_t, u_t, v_t\}$ is the state vector and motion signal, measured from a moving object or estimated for a target motion, is represented in a probabilistic fashion. Meaning that, instead of allocating a deterministic scalar to motion, one may describe it in terms of value of belief about existence of motion with a specific velocity in a particular location. This description of motion is also appropriate to depict the uncertain nature of activity propagation in the neural system. In the visual areas with retinotopic representation of space, the tuning property of neurons to a certain motion signal (including the position and velocity) is not deterministic and rather it is considered like an uncertain estimation.

Equations 3.2-3.3 are master equations of our framework which instantaneously update the estimated motion vector, based on measurements and expectations of the internal model. Equation 3.2 illustrates Bayesian combinations of latest measurements of motion with the most predicted state vector of the internal model. State transitions, as illustrated in Fig 3.1, are determined by conditional probabilities $p(I_{t-\delta t:t}|z_t)$ and $p(z_t|I_{0:t-\delta t})$, where the former is measurement pdf and latter refers to the the predicted state of motion.

3.1.2 Kalman filter as a motion predictor

The Kalman filter (Kalman, Rudolph, and Emil, 1960) is an optimal Bayesian estimator, in which state estimation is fulfilled by means of the prior knowledge about operational principles of the system, in combination with indirect and uncertain observations from its output. In Kalman formulations, these two sources of information are stated as ***Process model*** and ***Measurment model***, respectively indicating to the predictive and corrective steps of the filter. The reliability of each model is defined by additive Gaussian noises and a crucial point in Kalman filtering is the relative contributions of these uncertainty sources in final state estimation. This is reflected

in ***Kalman gain***, used to adjust the internal precision of the filter. The gain is recursively calculated to minimize the prediction error at each step and as such, the Kalman filter is the optimal estimator for this formulation.

Kalman filtering approaches have been widely used to model the internal expected motion (Bogadhi, Montagnini, and Masson, 2013; Deneve, Duhamel, and Pouget, 2007; Xivry et al., 2013) in the neural, behavioral and eye movement levels.

Despite successful and various applications of the Kalman filter, it has some limitations. First, the process and measurement models need to be linear, and second, the additive noises in both models are required to be Gaussian. It is not always easy to write system equations in a Kalman format and use the classical algorithm to find the Kalman gain. Models based on generalization of Kalman filter have been developed to deal with more complex systems, where standard Kalman is not sufficient to model system nonlinearities.

3.1.3 Particle filtering and CONDENSATION algorithm

Particle filtering (PF) methods are a family of mathematical approximations which are used to extend a range of Kalman-like applications. They may be addressed also as Sequential Importance Sampling (SIS), Markov Chain Monte Carlo (MCMC) and Sequential Monte Carlo (SMC) methods. The shared property of all these methods is that probability distribution functions (pdfs) in the system and measurement models can be represented by a set of weighted samples, so called particles. In these algorithms, measurements and estimation are done in discrete time steps and the posterior is represented by an approximated sum of weighted particles. The motion estimation algorithm that we have used in this work is a type of particle filtering implemented by the CONDENSATION algorithm.

The CONDENSATION (Conditional Density Propagation) algorithm was first proposed by Isard and Blake (1998) to model the tracking of curves in visual clutter. This algorithm is based on a probabilistic representation of motion information and conditional propagation of them in time. The main advantage of the CONDENSATION algorithm is that, unlike the Kalman filter, there is no assumption on linearity of process and measurement models, and of normality of perturbations, or uni-modality of input distributions. In CONDENSATION algorithm, as illustrated in Fig 3.2, probabilistic distributions are represented by a set of samples. The generality of the algorithm along with its simpler implementation than Kalman, makes it a powerful candidate for a general state estimation model of predictive coding.

The CONDENSATION algorithm may be adapted to study motion-based prediction. Tracking of moving targets has been modeled in different efforts in machine vision and neuroscience domains. A large family of works, as we have reviewed them in chapter 1, are built based on the hypothesis of smoothness of motion trajectories. In this context, Kalman filtering approach as an optimal recursive state estimator has been widely used. The CONDENSATION algorithm provides a probabilistic representation of motion state space in a simpler framework and with lower computational cost. Being stated in Markov chain guarantees being dependent only on the

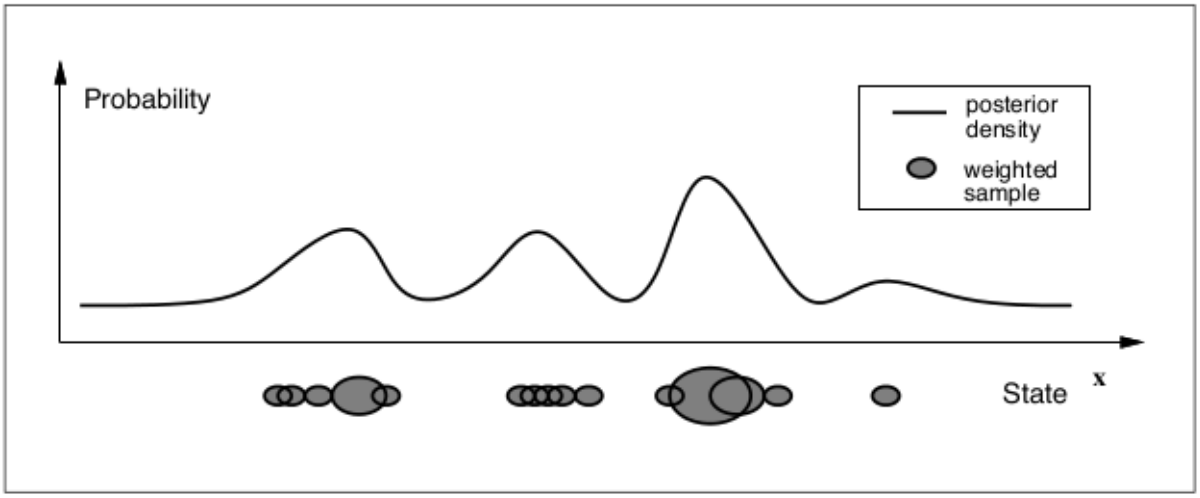

Figure 3.2: **Weighted sampling in CONDENSATION algorithm**, figure adapted from Isard and Blake (1998): continuous posterior distribution has been represented by a set of weighted circles.

preceding state and nearly real time motion estimation performance. In general, to study visual signatures, particle filtering gives us flexibility to model motion integration at a more precise level.

Motion-based predictive coding, as used as a theory in the context of this thesis, highlights the importance of motion signal as a potential modulator of visual processing. Physical motion of a stimulus, defined as a signal composed of position, speed and direction, after translation to a sensory signal is accurately delivered to specialized neural populations in different areas of the visual system. Our global motion perception originates from distributed neural map of the motion, as a measurable physical signal.

In the next section we will describe our abstract motion estimation framework and allover this chapter motion-based predictive coding will be referred to as MBP. This model bridges between the coherency of motion trajectory and neural stimulation sequences in time and position. This provides a simple and efficient abstract tool to investigate interaction between predictability of motion trajectory and the coherency of the corresponding neural code.

3.2 Methods: MBP model implemented by CONDENSATION algorithm

The basic framework used in this thesis was first developed to model the motion integration process in the aperture problem (Perrinet and Masson, 2012). Master equations of the model are equations 3.2−3.3 which represent probability distributions of the estimated motion and the predicted motion, in terms of position and velocity components. The model uses elementary motion detectors at its input layer similar to the ones proposed by Adelson and Bergen (1985), and the generative model of Weiss, Simoncelli, and Adelson (2002) for the conversion of luminance

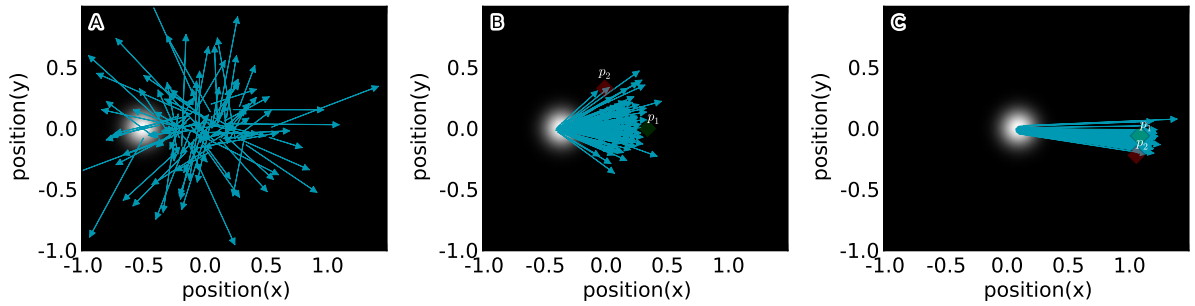

Figure 3.3: **Estimated motions in three different frames of the trajectory**, illustrated as a distribution of weighted particles. For simplicity we have plotted only one tenth of all particles. Stimulus is a horizontally moving Gaussian dot which is displaced one spatial unit in one temporal period (in this example it starts to move from $x = -0.5$ to $x = 0.5$, at speed $= 1$). The relative magnitude of particles represents the estimated speed **A)** estimated motion in the first frame is completely random (initialization of the algorithm) **B)** estimated motion in 10th frame is relatively concentrated on the position of stimulus with a reasonable dispersion around physical motion **C)** estimated motion in 40th frame is well placed in the center of dot with a low variance around its direction. In **B** and **C** two sample particles have been highlighted with green and red markers $(p1, p2)$, representing respectively for high weighted and low weighted local estimations.

to motion measurements. The temporal coherency theory of Burgi, Yuille, and Grzywacz (2000) is implemented as prior knowledge or motion state transfer model.

In this section we will first describe approximation of estimated motion by the CONDENSATION algorithm and steps of particle filtering. Then we will explain internal model of motion and finally we will elaborate the luminance-based motion detection procedure.

3.2.1 Weighted samples of estimated motion

We have used particle filtering to approximate probabilistic representation of motion. This decreases the computational cost of the model by removing unnecessary calculations as well as giving biologically plausible interpretations to the model.

Indeed, probabilistically estimated motion $p(z|I_{0:t})$ (calculated in equation 3.2) can be approximated by $\hat{p}(z|I_{0:t})$ as follows:

$$\hat{p}(z|I_{0:t}) = \hat{p}(\vec{x}, \vec{V}|I_{0:t}) \approx \sum_{i \in 1:N} w_i \cdot (\delta(\vec{x} - (x_i, y_i)), \delta(\vec{V} - (u_i, v_i))) \tag{3.4}$$

In this framework, we represent estimated motion as a set of N particles defined as $\pi_i = \{x_i, y_i, u_i, v_i\}$ for $i \in \{1, 2, ..., N\}$, with corresponding normalized positive weights $w = \{w_i\}_{i \in 1:N}$, $\sum_{i \in 1:N} w_i = 1$.

Fig 3.3 illustrates an example of estimated motion represented by particle sets.

Model initialization: Stimulus of the model is a movie composed of certain number of

frames presenting a horizontally moving target. In most of experiments of this thesis, the stimulus is an image of a Gaussian dot. At very first frame of the stimulus, before having any measurements available, the model is randomly initiated. In other words, it is assumed that at the frame number 1, position measurements are random probability distributions over the possible range, and velocity measurements are composed of random samples from a certain range of slow speeds and directions between 0 and 2π. (See Fig 3.3-A).

3.2.2 Internal model: Motion coherency

After initialization of particles at frame number 1, the distributed particles allover the luminance profile go under smooth trajectory prediction. The predictive step is done by the internal representation of motion that favors smooth trajectories. Having $z_t = \{x_t, y_t, u_t, v_t\}$ as the state vector then $p(z_t|z_{t-\delta t})$ in the equation 3.3 can be calculated based on temporal coherency of motion at each step, implemented by equations 3.5-3.6:

$$\begin{aligned} x_t &= x_{(t-\delta t)} + u_{(t-\delta t)} \cdot (\delta t) + \nu_x \\ y_t &= y_{(t-\delta t)} + v_{(t-\delta t)} \cdot (\delta t) + \nu_y \end{aligned} \tag{3.5}$$

$$\begin{aligned} u_t &= \gamma \cdot u_{(t-\delta t)} + \nu_u \\ v_t &= \gamma \cdot v_{(t-\delta t)} + \nu_v \end{aligned} \tag{3.6}$$

$$\nu_x, \nu_y \propto \mathcal{N}(x, y; 0, D_X \cdot \delta t) \tag{3.7}$$

$$\nu_u, \nu_v \propto \mathcal{N}(u, v; 0, (\sigma_p^{-2} + D_V^{-1})^{-1} \cdot \delta t) \tag{3.8}$$

Where ν_x, ν_y are Gaussian distributions of position blurring and $D_X \cdot \delta t$ is blur value sampled at each time step. Blurring of velocity is done with a sample from ν_u and ν_v, Gaussian distributions with standard deviation of $(\sigma_p^{-2} + D_V^{-1})^{-1} \cdot \delta t$. This standard deviation is combination of two other standard deviations (σ_p and D_V) respectively matched to the prior information on slowness and smoothness of motion.

Equations 3.5-3.6 compose a generative model to implement smoothness of the trajectory at each step. Particles at each frame are transported by these equations and then are weighted based on motion measurements.

3.2.3 Luminance-based detection of motion

The motion detection task in the model is done based on the luminance profile of successive frames. After initialization at frame number 1 particles are transported by the generative model. Then at the frame number 2 predicted particles are weighted based on available measurements. In this step, mismatch between predicted position of target (suggested by each particle) and real position of it in the current frame is measured. The particles (the quality of estimations) are weighted based on their mismatch, which is stated in terms of luminance energy.

At each frame, to compute luminance mismatch energy associated with each particle we translate the reference image with estimated velocity of each particle. A local estimator particle is represented as $\pi_i = \{(x_i, y_i, u_i, v_i), w_i\}\ for\ i \in \{1, 2, ..., N\}$. To be able to asses the contribution of estimated velocity at the exact estimated position of the particle, we need to calculate $p(I_t|\pi_i)$ (I_t is luminance information). To this aim we approximate this probability by defining a Gaussian mask, M_{π_i}, around π_i in 2-D visual space:

$$M_{\pi_i} = \exp - \frac{(x - x_i)^2 + (y - y_i)^2}{2 * \sigma_{RF}^2} \tag{3.9}$$

Where σ_{RF} is the standard deviation of the mask which may be interpreted as the radius of the positional receptive field centered on the particle (the radius of red and green areas in Fig 3.5). The value of σ_{RF} will affect the precision of measurements and eventually the estimated motion (See Fig 3.6). Having I_t as the reference image, the luminance mismatch and energy of each particle is given by:

$$E_{\pi_i} = ||(I_t - \hat{I}_{\pi_{i(t-\delta t)}}) * M_{\pi_i}||^2 \tag{3.10}$$

then particles are weighted based on their energy:

$$\acute{w}_i = \exp(\frac{-E_{\pi_i} * C^2}{2}) + \epsilon \tag{3.11}$$

where C is an measurement contrast factor which controls the number of high weighted particles and ($\epsilon \ll 1$) ensures that none of weights will be zero. Final weights are normalized over all particles, as stated in equation 3.12.

$$p(I_t|\pi_i) \ \simeq \hat{p}(I_t|\pi_i) \ = \ w_i = \frac{\acute{w}_i}{\sum_{i \in 1:N} \acute{w}_i} \tag{3.12}$$

Fig 3.4 and Fig 3.5 illustrate the evaluation procedure of particles. In Fig 3.4 two frames have been chosen (frame numbers 11 and 41) and for each frame two sample particles are shown from estimated motion at previous frames (frames number 10 and 40 respectively). Particles (p_1 and p_2) are indicated with green and red markers and the gray circle matches to the actual position of stimulus, which is the reference of evaluation. Green and red circles indicate the stimulus position as predicted by p_1 and p_2.

Fig 3.5 illustrates the luminance energy mismatch for p_1 and p_2, previously indicated in Fig 3.4-A. The position of each particle is marked with a Gaussian mask of the same color. Mismatch at each case is a crescent resultant from subtraction of two circles (actual position and predicted position). Estimation by p_1 is very close to the reference, resulting in a very thin mismatched area. While, mismatch area associated to p_2 is bigger (its direction is far enough from horizontal). Then according to equations 3.10 and 3.12, p_1 and p_2 respectively get high and low weights.

Fig 3.8 illustrates luminance distribution of a Gaussian dot as stimulus, versus noisy mea-

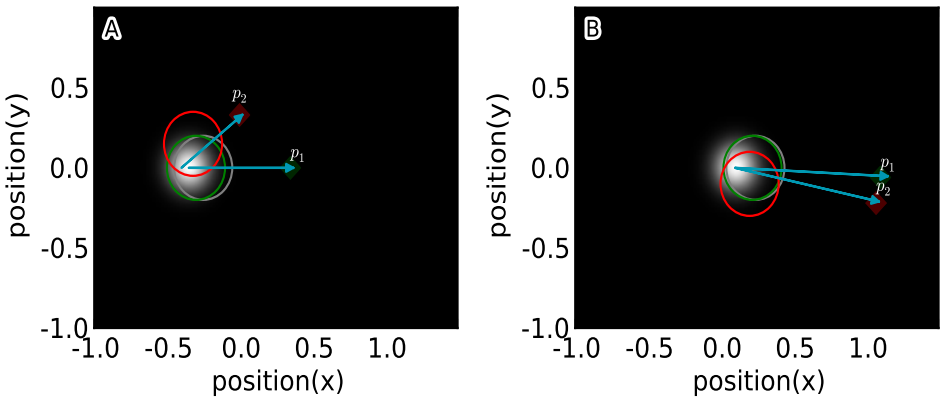

Figure 3.4: **Motion measurements to evaluate four sample particles illustrated in Fig 3.3**: in each figure p_1, p_2 respectively symbolize high weighted and low weighted local estimations. Green and red circles are matched to the positions of stimulus predicted by p_1 and p_2. Gray circles represent the actual location of stimulus at next frame, which is the reference to evaluate particles.

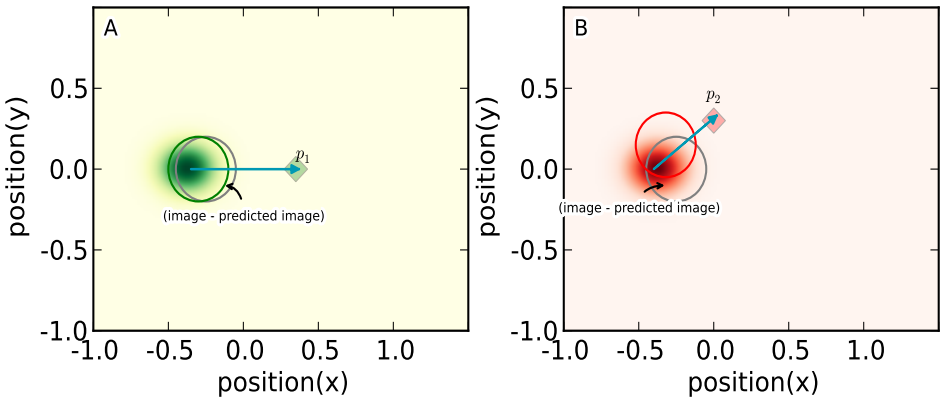

Figure 3.5: **Evaluation and weighting of local motion estimations based on their likelihood**: the actual position of stimulus in the next frame has been shown with gray circle (reference image), while green and red circles are predicted position of stimulus by each particles. The likelihood of each particle is determined by subtraction of reference image and the circle of corresponding color. Overlap between subtracted area and the Gaussian mask centered by the particles defines energy of each particle. **A)** The resultant crescent by the subtraction *(reference image - green circle)* is very thin and located on the tail of green mask. This particle has little luminance energy. **B)** The area *(reference image - red circle)* is a crescent with a significant surface and overlap with the red Gaussian mask. This particle contains high luminance energy.

surements of it due to the mask described in equation 3.9. In Fig 3.7 estimated position of stimulus have been shown along with instantaneous location of luminance.

3.2.4 Particle resampling

After measurement, evaluation and dedicating relative weight to local motion estimators, weights will go under a resampling algorithm. The resampling step is done to avoid a usual numerical problem in particle filtering, known as particle impoverishment. The problem is associated with presence of particles with small weights in the estimated motion.

We have used a classical resampling routine similar to the one used in the CONDENSATION algorithm (Isard and Blake, 1998). Allover this study we had resampling factor equal to 100%, meaning that the total number of particles remains the same from the first frame to the end. By resampling, particles based on their weights are either eliminated or duplicated. In fact similar to the histogram equalization algorithms, it gradually reduces the difference between weights (quality of local estimations). This has been illustrated in Fig 3.9, where weight distributions gets unified by advancing in frame numbers.

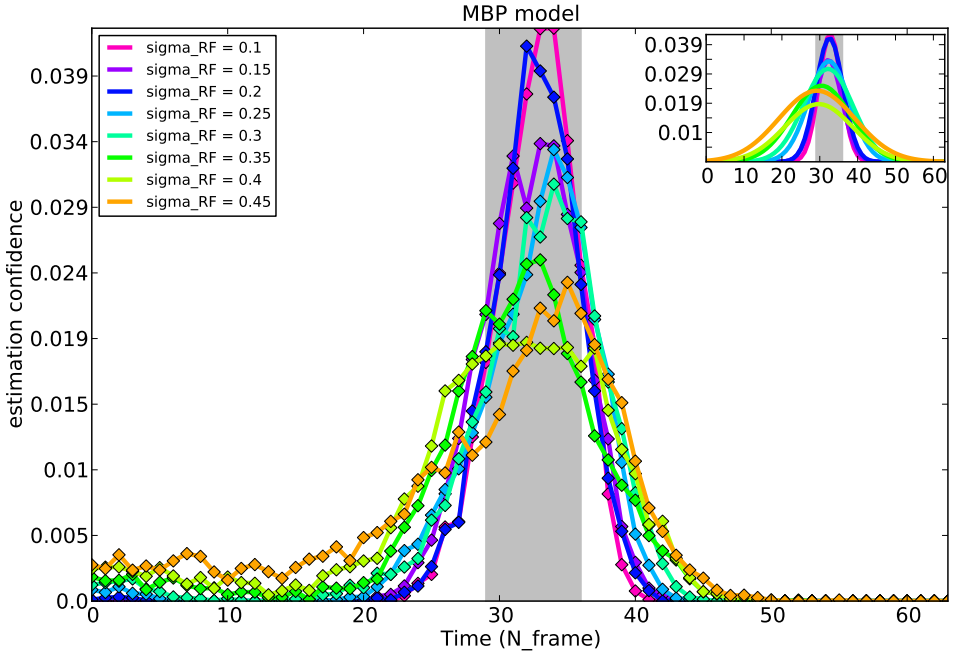

Figure 3.6: **The effect of σ_{RF} in estimated motion**: the figure illustrates the estimated motion in a specific point in the trajectory of stimulus, while σ_{RF} is changing between 0.1 and 0.45 of the screen (the visual area covering whole path of the stimulus). The inset includes the Gaussian fits of each condition. The shaded area indicates the temporal period in which that specific point is covered by the luminance of stimulus. The peak of confidence is small because this example shows the confidence of the estimated position in a small positional bin. In the tracking states the instantaneous integral of confidences over positional bins matched to the surface of the stimulus would be 1.

Considering that weights represent relative quality of estimations, resampling algorithm moves toward creating more particles around high weighted ones and decreasing the number of particles with lower weights. To this aim Cumulative Weight Function (CWF) is calculated (See Fig 3.10). With low resolution, CWF seems to grow linearly between 0 and 1 and over all particle. Evidently, very localized assessment of CWF will reveal non equal contribution of each particle. As shown in Fig 3.11, CWF is sampled by a unified rate equal to $r = \frac{1}{N}$ (where N is number of particles), indicated by blue dotted lines. Any particle with considerably bigger weight than the rate is likely to be sampled more than once, and on the other hand, particles which cause a small jump in CWF are likely not to be sampled at all. Two example particles in the index range of $[0, 10]$ have been marked by green and red diamonds, where green is sampled twice and red is not sampled. Sampling a certain particle index twice means replicating its estimated coordinates and labeling it with a index coming from one of removed particles.

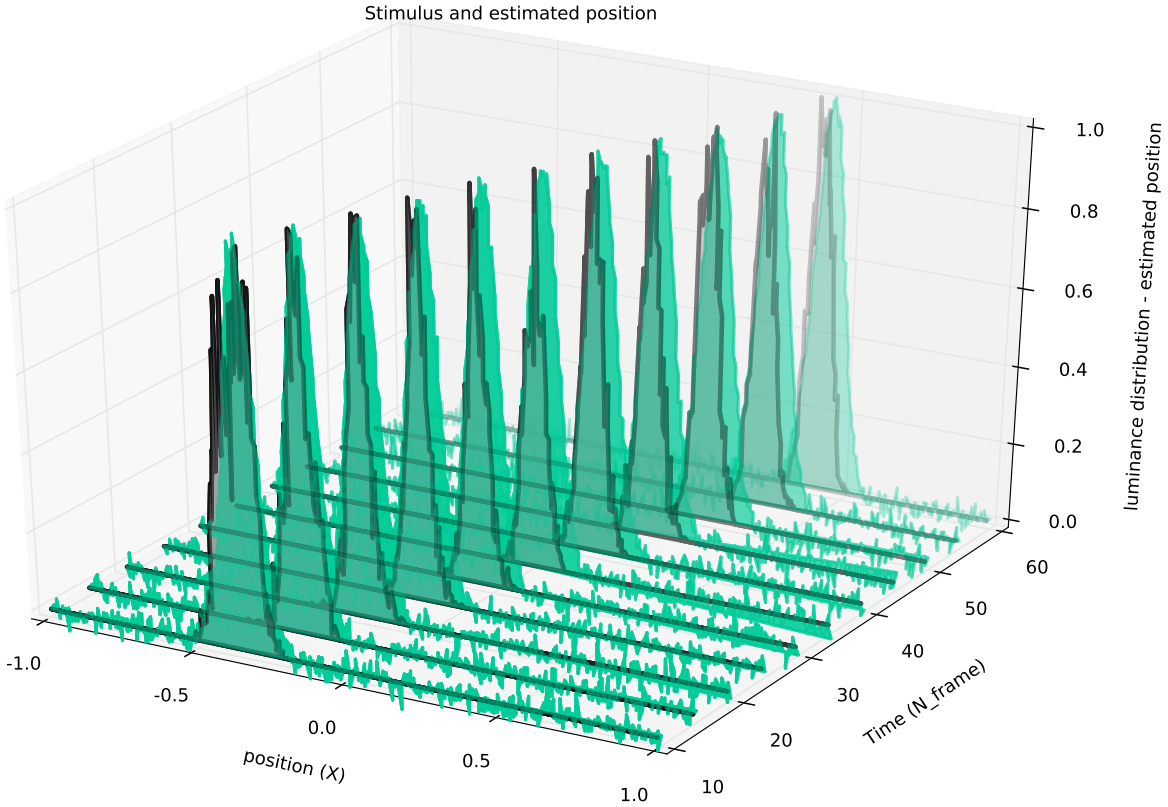

Figure 3.7: **Estimated position of stimulus versus luminance distribution**, shown for sampled frames from the trajectory: stimulus is moving from -0.5 to 0.5, green shadowed areas refer to luminance while estimated position is plotted by black solid lines. For visibility we have multiplied the value of estimated peaks with a factor around number of particles.

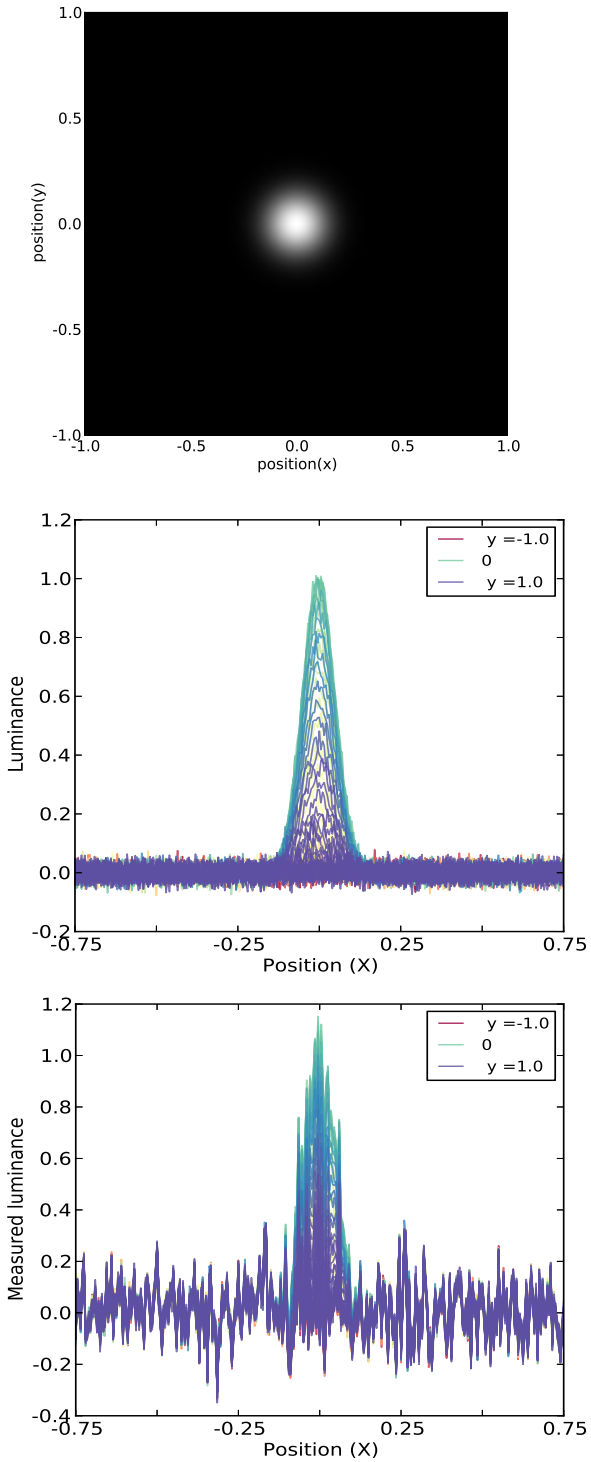

Figure 3.8: **Luminance of the stimulus and luminance measurements**: **Top)** Luminance profile of a Gaussian dot as stimulus, a frame matching to a position in the middle of trajectory **Middle)** Luminance distribution of Gaussian dot over position(X), where each colored trace from red to purple corresponds to luminance at position(y) between -1 and 1 **Bottom)** Luminance measurements from stimulus with the Gaussian mask of each local estimator, as described in equation 3.9

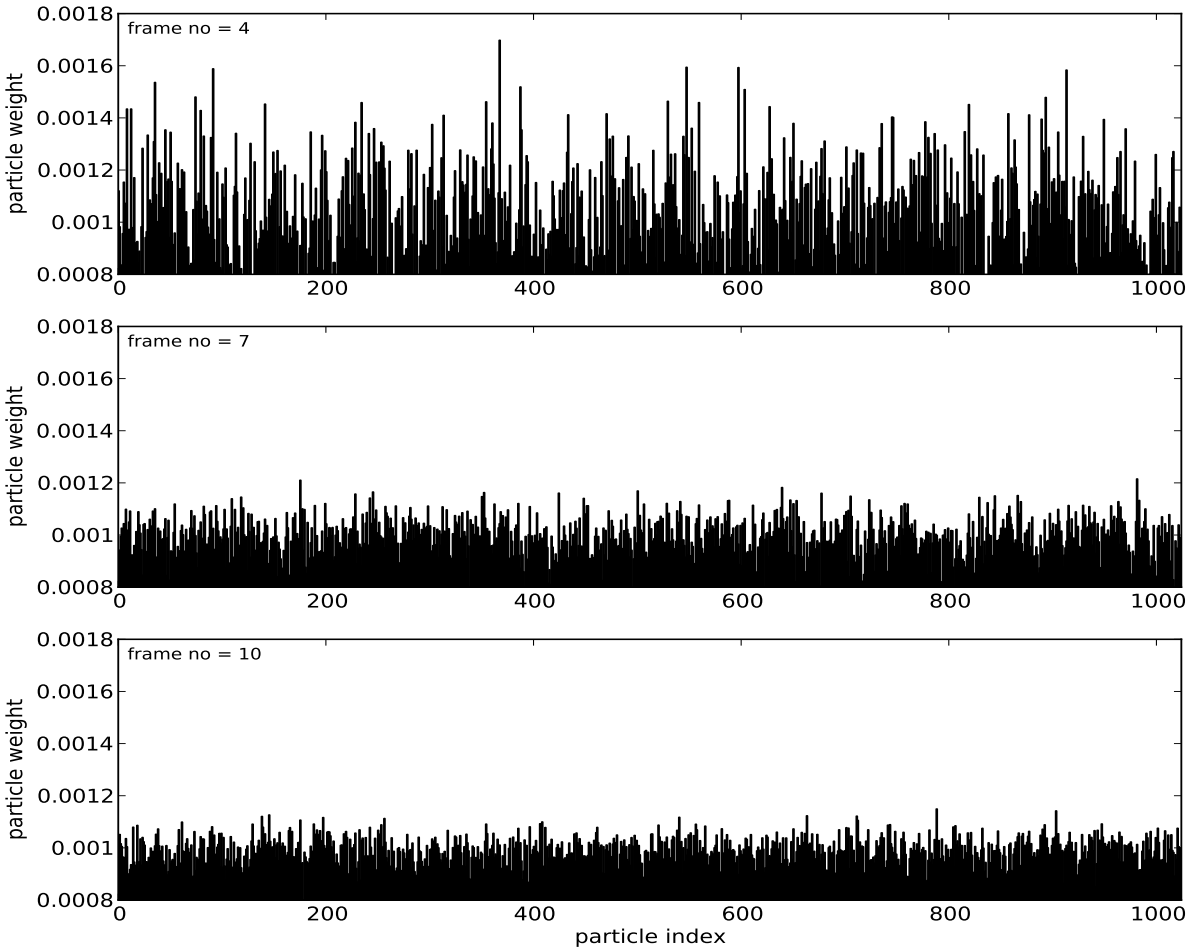

Figure 3.9: **Weight distribution of particles in three sample frames of stimulus**: due to resampling weights are gradually moving toward a unified distribution

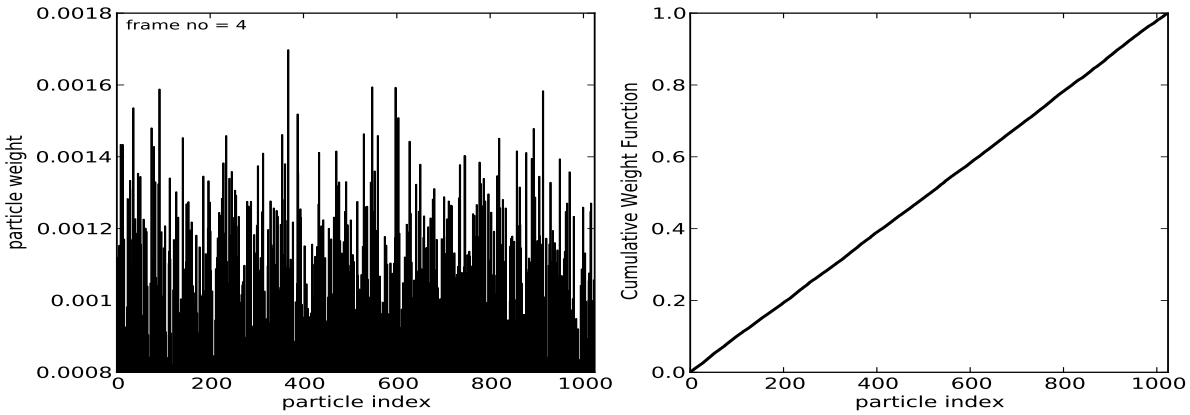

Figure 3.10: **Weight distribution of particles in frame No 4 and corresponding Cumulative Weight Function**, as it is clear from weights, increments in CWF are not necessarily equal, even though it looks linear between 0 and 1.

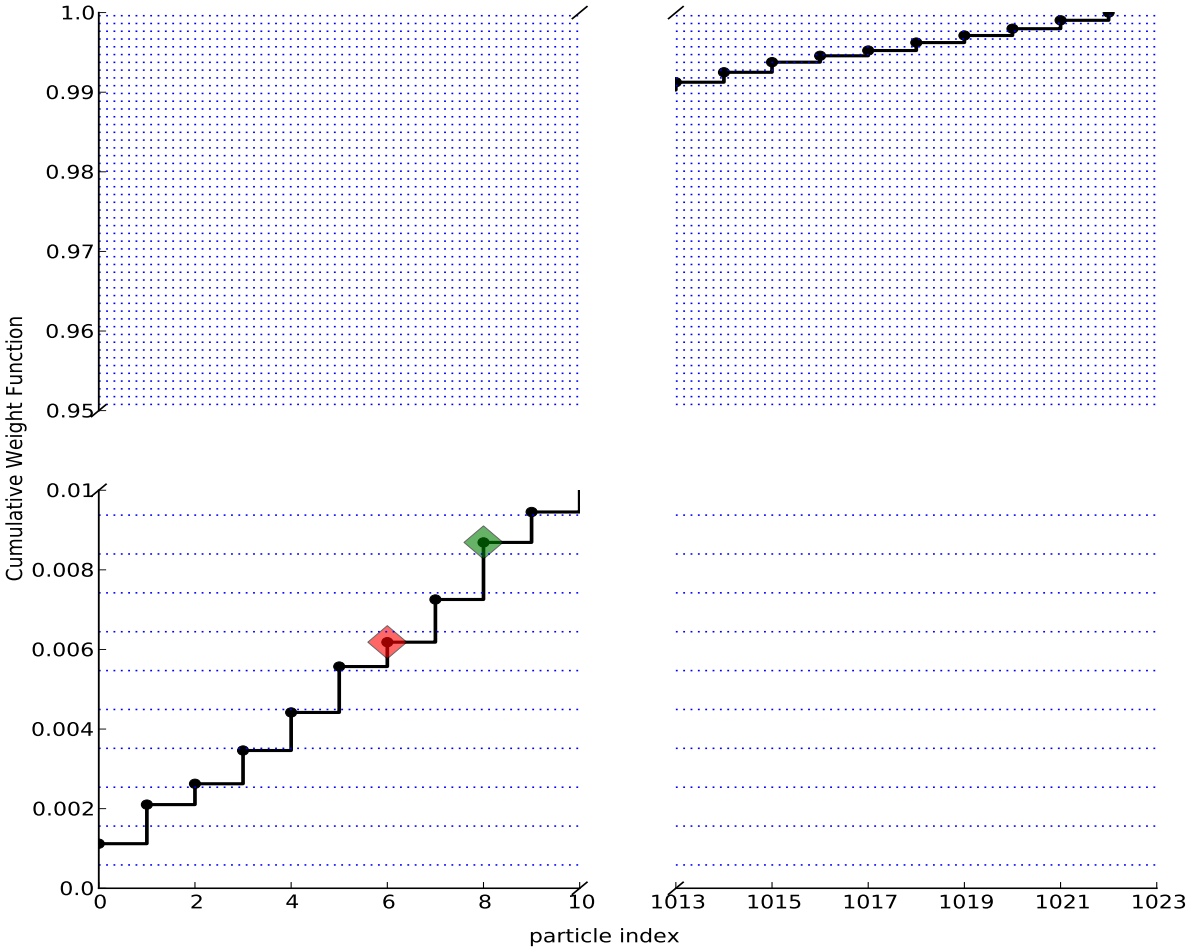

Figure 3.11: **Particle resampling**: CWF illustrated in Fig 3.10 has been shown for two localized parts, including 10 first and 10 last particles (note broken axes). As the resampling factor is 100%, the total number of particles is kept at each frame (1024 allover this study). The weight range $(0, 1)$ is divided to 1024 segments (dotted blue lines) and the intersection of each segment with CWF determines the resampled particles. High weighted particles (like the one marked with green diamond) will have two intersections with segments, versus some other particles with no intersection (compare with the particle marked with red diamond). This resampling routine at each frame will eliminate less confident particles and replicate the particles with high confidence.

Simulation details: All simulations were conducted using a cluster of Linux nodes and python environment (version 2.7), with Numpy (Oliphant, 2007) and Scipy modules (respectively versions 1.6.2 and 0.10.1). Visualizations were performed in Matplotlib (Hunter, 2007). All scripts are available upon request.

3.3 Computational advantages and biological implications

In this section we will summarize the advantages of our modeling framework for the purpose of this study.

Position of motion as a random variable: There is one main difference of this implementation with previous studies which results in a more complete representation of motion. Indeed, previous models as Burgi, Yuille, and Grzywacz (2000) and Weiss, Simoncelli, and Adelson (2002) and some other feedforward models proposed for studying motion detection are lacking an important biological constraint in detection and estimation of visual motion. This constraint is about the subtle way of mapping motion information from physical world via specialized receptive fields to the visual processing areas.

In the motion detection models developed for computer vision purposes, it is usually neglected that visual motion is composed of successive stimulation of neighboring positions on the trajectory of moving objects. In the previous feed forward models, motion is estimated as a velocity distribution over the whole visual space and position of motion does not play any role in the estimations. Therefore, if one shuffle all local velocity estimations, the global readout of motion would remain still the same.

In the current implementation, position of motion is an important piece of information to be considered, meaning that all local measurement and estimations have precise biological meanings and indeed there is a distance based dependency between local information. Based on the temporal coherency hypothesis, for the visual system motion is a distributed stimulation along a coherent path and spatial arrangement of local motion in the visual field matters.

In MBP model, position of motion is a random variable which along with velocity defines the motion. Thus, in the probabilistic representation of motion instead of $P(\vec{V}(\vec{x})|I)$ we use $P(\vec{x}, \vec{v}|I)$ (where I is the image luminance information). This representation of motion is consistent with retinotopic representation of motion and can base a more realistic, functional model to study visual computations. In addition, motion estimation problem moves from technical optimization issue to an abstract predictor of neural activation in specific parts of the retinotopic map.

Generalized motion estimator: As we mentioned earlier in this chapter, the biggest advantage that comes with Generalizations of Kalman filters is their ability to deal with non-Gaussian and non unimodal inputs and non Gaussian noise as system and measurement perturbations. In the CONDENSATION algorithm (Isard and Blake, 1998) and the particle filter used in the current study, multi modality adds flexibility and precision to the model and makes it a more competent candidate to study visual motion. Particularly, studying motion

integration with a particle filter, provides new theoretical insights on the non isotropic and asymmetric propagation of local motion estimators toward a global decision-like motion. As the base model of the current study, Perrinet and Masson (2012) have elaborated various aspects of motion integration by modeling the aperture problem. For instance, as they proposed, in tracking stimuli with more complex shape like a tilted bar, particle filtering of motion results in emergence of two line endings. This interesting feature arises solely from the dynamics of the model, instead of an ad hoc constraint.

In addition, as having multimodal posterior distribution is a feature of particle filters, the current model is able to deal with tracking multiple objects in one scene. Theoretically, this type of experimental setups in model are useful for provide predictions about the relative tracking quality of multiple object based on the difference in their contrast, shape and trajectory.

Discrete time state estimation and computational cost: Particle filtering algorithm uses the power of Markov chain to model motion dynamics, therefore provides discrete time propagation of state densities. On the other hand, as an efficient and precise enough approximation, motion states are represented by weighted samples of posterior probabilities. The efficient approximation of estimations is a big advantage of the current framework, as it decreases the computational cost dramatically. In addition, with a certain complexity and fixed number of particles the model is able to deal with a range of stimuli from simple dots to real-world videos (See Fig 3.12).

Particles as abstract receptive fields: Abstract motion estimator used in this study is structurally compatible with motion processing in the visual system. In the model, estimated motion as a posterior probability includes a set of weighted particles, probabilistically distributed over certain positions and with certain velocities. The two dimensional posterior may be considered as a predictor for specifications of neural activity in retinotopic areas of the visual system. While tracking a smoothly moving stimulus, the position and velocity of particles are continuously updated, based on smoothness constraint and instantaneous sensory input. This simple framework depicts a crude image from relative stimulation of receptive fields lying over the trajectory of the stimulus. Indeed, the position of each particle may be considered as the center of the activated receptive field, where the associated weight quantifies the relative contribution of that receptive field in the population response.

Simple map of physical motion into neural activity: The current implementation provides a particular framework to highlight the encoding of physical motion into probabilistic neural activity. In fine grained maps for estimated position and velocity of stimulus, one can study the contribution of different components of the motion signal in modulation of corresponding neural response. In chapters 3 and 4 we have elaborated these three configurations of the model as MBP, PX (position-based prediction) and PV(velocity-based prediction) models. Indeed,

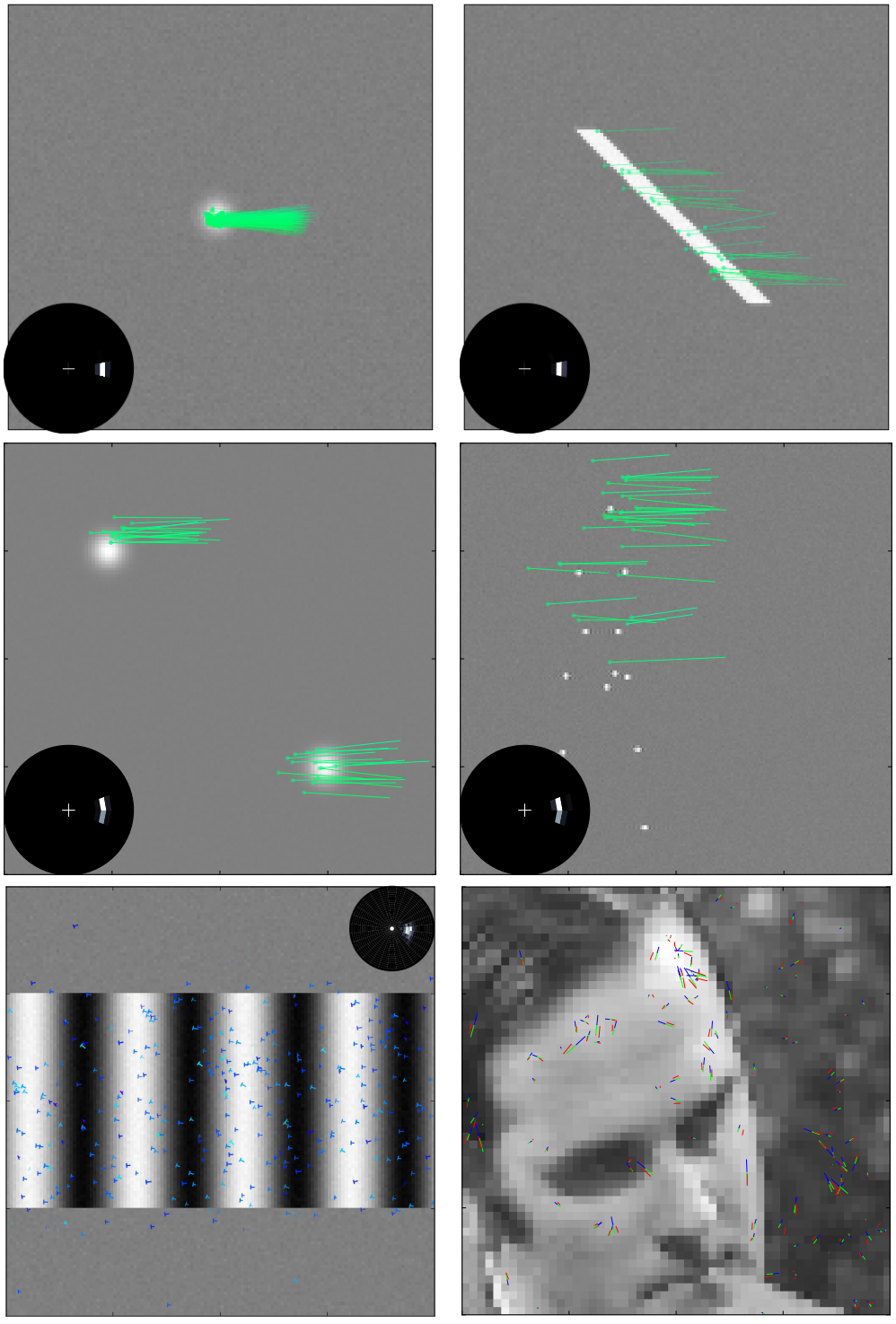

Figure 3.12: **Snapshots of different moving stimuli tested on model**: from top left to bottom right: dot, tilted bar, multiball, walking cartoon, grating and a natural video. The inset illustrates the distribution of estimated directions for motion.

PV model would be equivalent to the models like Weiss, Simoncelli, and Adelson (2002) and PX model in principal is similar to the model of Tlapale, Masson, and Kornprobst (2010). In various cases we have highlighted the difference of MBP model with these two models as control. Furthermore, the spatiotemporal readouts of the model are particularly informative in studying the effect of delays in position coding. We have explored these type of questions in chapter 5.

Prediction as a large scale constraint: Prediction of motion trajectory as a large scale constraint may modulate the temporal and spatial distribution of neural activity, independent from detailed and neural connections specific to each visual areas. The current framework, implements the internal model of smooth trajectories, with a particular emphasis on the coherency of neural activity being caused by prediction. In this context, it is important to study the role of strong and weak internal models on the precision and quality of positional code.

Indeed, simple transport of position and velocity information in predictable smooth trajectories explains a large range of visual computations. Implementation with particle filter brings the advantage of studying mentioned transports in probabilistic distribution of local motion estimators. In the current framework, by a set of local motion estimators we are able to highlight the importance of local positions of a global motion in the efficient motion tracking.

Generic sensory computations: The motion-based prediction hypothesis implemented with particle filtering is precise enough to provide insights to visual computations and also generic enough to be a base toward asking similar questions on sensory integration of other modalities. First, as an abstract model for visual motion, it explains some shared neural mechanisms among different motion processing stages of the visual system. For instance, diagonal motion extrapolation that we have proposed in chapter 5 can partly be implemented from retina to the higher level processing areas like MT, independent from neural specifications and connection properties of each area. Likewise, visual motion integration elaborated by this model has a counterpart in motion integration of somatosensory and tactile senses.

Summary

In this chapter, we described the implementation details of the motion-based prediction model. We elaborated motion detection, motion measurements, motion prediction and motion estimation steps. In the next chapter we will use the MBP model and two control configurations of it (PV and PX models) to study robust motion estimation in noisy or fragmented trajectories.

Part III

Results

Chapter 4

Motion-based prediction and robust tracking of moving objects

Abstract

As we discussed in previous chapters, prior knowledge on the smoothness of motion can be implemented in a robust motion estimation framework. In other words, gaining enough knowledge from trajectory of moving object can lead to emergence of an internal representation for the trajectory, which keeps predicting the motion by less reliance on the instantaneous sensory input. In this chapter, we have surveyed the functional advantages of such a velocity-dependent internal representation. We have challenged the MBP model (motion-based prediction), during temporal absence of target or at the presence of other sources of uncertainty. To this aim, we have explored the sufficiency and competency of two other incomplete configurations of the model in robust tracking. These control models are configured by excluding one ingredient of motion : PX (position-based prediction) model and PV (velocity-based prediction) model. Furthermore, by assuming prior information on statistical slowness of motion, we have tested how each configuration behaves, with weak and strong preferences for slow motions.

According to our results, even for a simple stimulus such as the one that we have used (a gaussian dot), PX model provides a non efficient internal representation and a weak tracking. Between PV and MBP configurations, MBP model has a more robust tracking than PV model over a wider contrast range.

Also, in simulations of a blanked trajectory, we assessed how internal representation may play as a mechanism to fill the gap between trajectory fragments, and in the efficient estimation of motion at reappearance of stimulus. In this case, MBP model provides an efficient internal representation and quicker catch up at trajectory restart point. We also have explored the interaction between prior information on the slowness of motion and contrast of stimuli. Having a very strong prior preference on slow speeds does not let tracking to develop, even in the MBP model.

Our results in this chapter suggest that, the velocity dependent internal representation of motion, as implemented by MBP model, may serve as motion extrapolation mechanism to keep sensory coherency in the case of fragmented trajectories, as well as providing an efficient catch up of the stimulus after reappearance.

4.1 Problem statement: Internal representation of motion and tracking

Efficient and robust tracking of moving stimului is reflection of the precise sensory information about the position and velocity of target. Under realistic situations, motion trajectories are likely to be fragmented or temporally occluded or to be embedded in a low contrast context. These are situations in which smooth sensory flow is perturbed but visual system is still able to keep coherency and efficiency of motion perception. (See Fig 4.1)

This suggests existence of an internal representation for motion which may emerge after accumulation of enough knowledge from the trajectory, to make a robust neural code against transient sensory perturbations. Such a representation is also consistent with theories on temporal coherency of motion. Statistically, most of motions in the physical world are smooth and predictable. This fact has been the base for a theory on temporal coherency of motion (See chapter 2) and consequently some theoretical works have studied efficiency of neural codes of moving object from this point of view.

As we reviewed in chapter 1, there are experimental studies on assessing neural activity or behavioral efficiency in situations in which the visual motion trajectory is transiently disappeared or occluded. Here, we mention two most relevant experimental studies with similar questions:

As an example of behavioral studies, in a recent work Bogadhi, Montagnini, and Masson (2013) have recorded eye movements in tracking blanked motion trajectories. They report that for short absence of stimulus, velocity of eyes drops down to a certain level (between 40%-60% of target velocity), and catches the moving stimulus efficiently after reappearance (See Fig 4.2). Assad and Maunsell (1995) recorded the activity of neurons in parietal posterior cortex of monkey to study the inferred motion during temporal disappearance of stimulus (See Fig 4.3 and Fig 4.4). They have found some neurons which stay active during absence period of stimulus. Results of this study suggest that, even when stimulus is not present, there may be a neural mechanism to monitor the displacement or keep an abstract representation of the stimulus.

In this chapter, we have studied robustness of estimated motion in presence of uncertainty sources like in transient disappearance of stimulus and low contrast conditions. To this aim, in an abstract Bayesian framework, we have disentangled the relative importance of position and velocity informations in motion tracking performance and coherency of internal representation. We also have assessed the role of prior information on slowness of motion in robust tracking in low contrast conditions.

This chapter is organized as follows: First we will have a brief introduction to the relation of motion-based prediction theory and motion extrapolation. Then in Section 4.3 we will review the principals of the MBP model and other configurations (PX and PV) used to survey tracking and internal representations of the trajectory. Section 4.4 includes the results of the study and finally in Section 4.5 we will conclude our finding in this chapter.

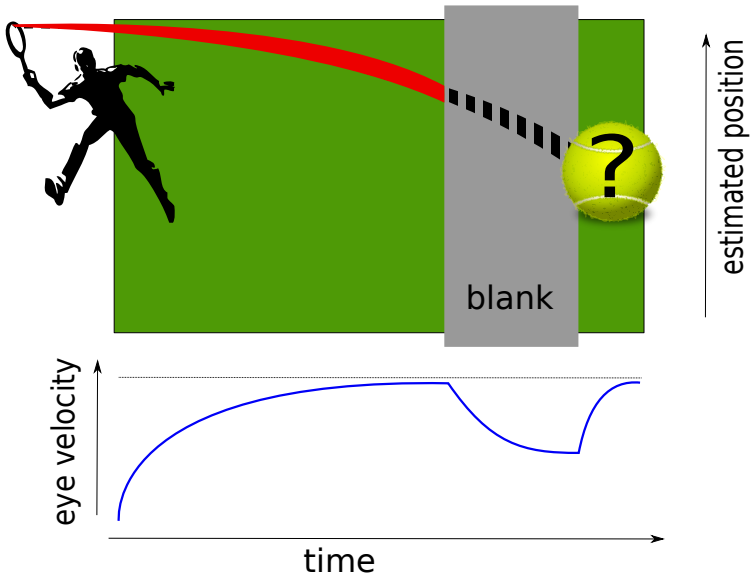

Figure 4.1: **The problem of motion extrapolation in the fragmented trajectories**: as an object moves in visual space (as represented here for commodity by the red trajectory of a tennis ball in a space-time diagram with an one-dimensional space on the vertical axis), the sensory flux may be interrupted 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? 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, as it is suggested by Becker and Fuchs (1985).

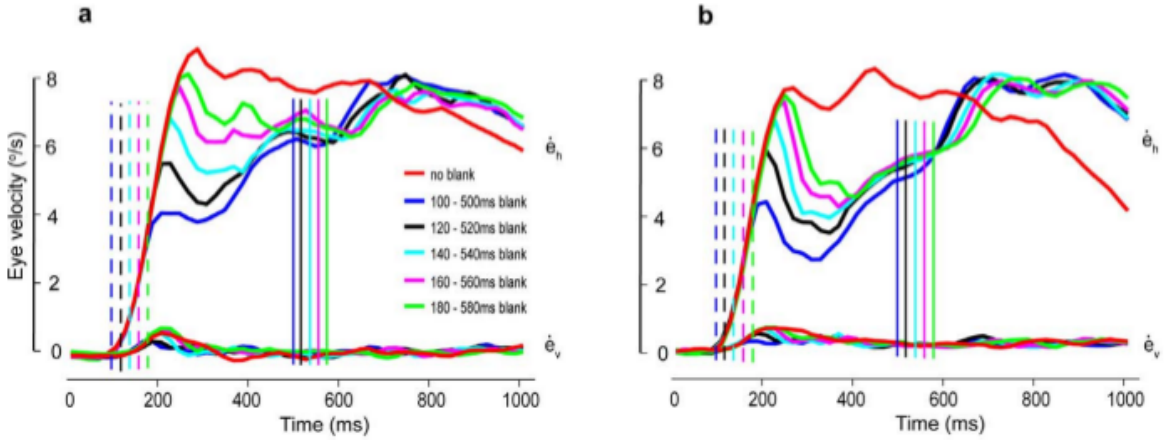

Figure 4.2: **Eye movement recordings from a recent study**, figure adopted from Bogadhi, Montagnini, and Masson (2013). The stimulus is a horizontally moving tilted bar which is blanked for 400 ms during its early trajectory. Traces with different colors correspond to different start time of blank. Start and end of blanks have been shown with dashed and solid lines, respectively. **a)** horizontal and vertical eye velocities, recorded from a naive subject, **b)** horizontal and vertical eye velocities recorded from a non-naive subject

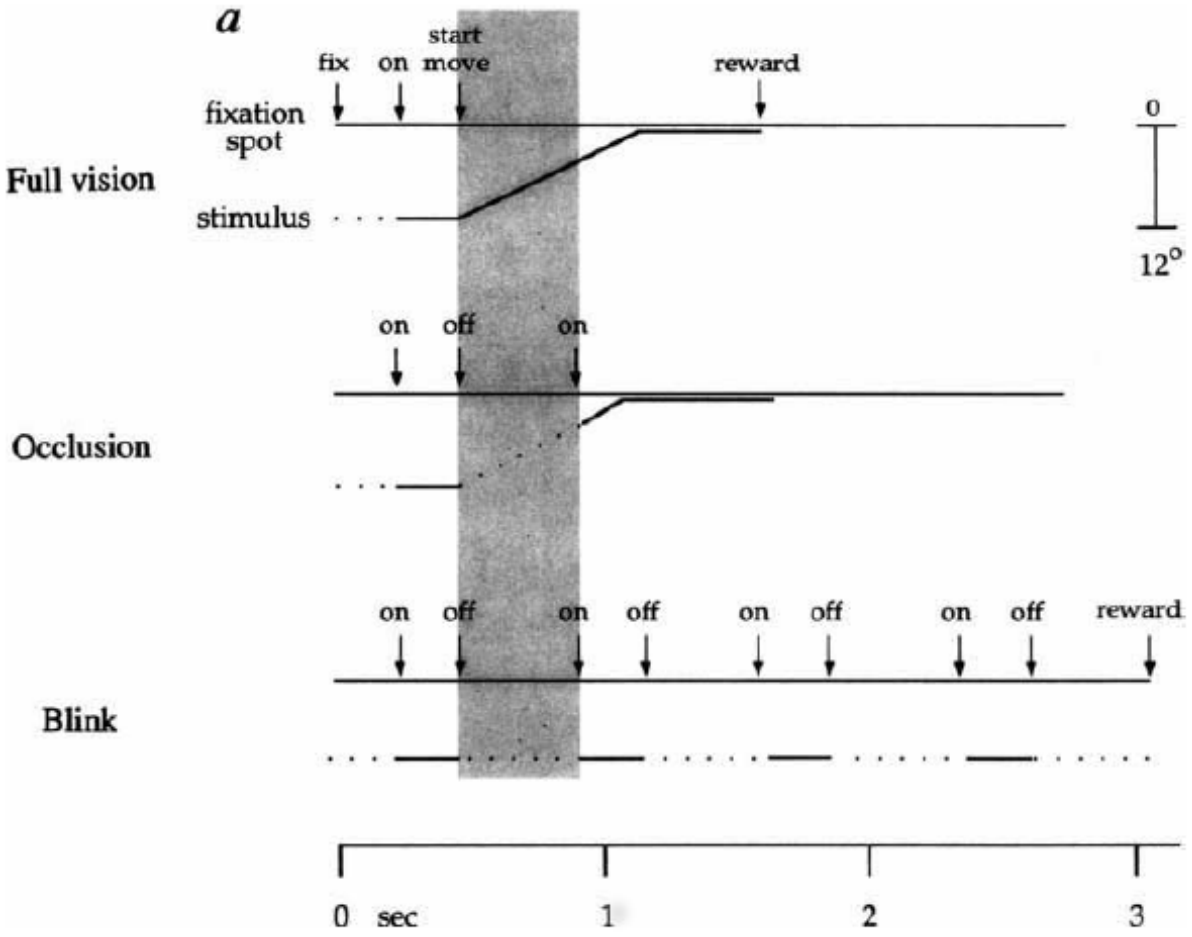

Figure 4.3: **Experimental protocol of Assad and Maunsell (1995) to study inferred motion during transient absence of stimulus**: neuronal firing rates of primate parietal posterior cortex have been recorded for three stimulus types including full vision, occlusion and blink. Time is shown in the horizontal axes and the horizontal line represents the fixation point. In full vision trials, after initial fixation stimulus appears in a distance from fixation point. Then it starts to move toward it, meanwhile the spiking response is recorded for the shaded duration. In occlusion trials initially stimulus appears at the same position of full vision trial but then it disappears and reappears close to fixation point, as if it was moving behind an occluder for the absence period. The presence and absence of stimulus is marked with on and off notations and spiking activity is recorded for the absence period which is shaded. Each blink trial includes an arbitrary sequence of four stationary blinks of stimulus, meaning that after absence period, stimulus reappears at the disappearance position and spiking activity is monitored during off periods shown by dashed area.

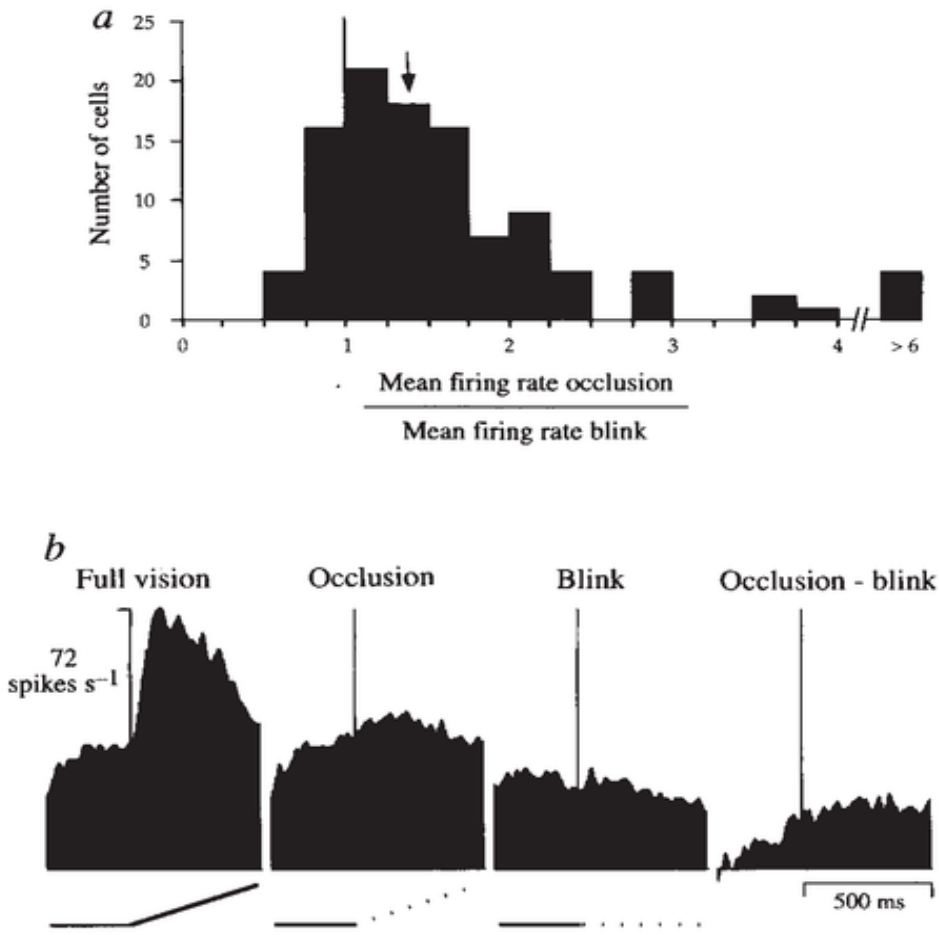

Figure 4.4: **Finding of Assad and Maunsell (1995) in experimental protocol illustrated in Fig 4.3**: **a)** Histogram of relative response of recorded cells in parietal posterior cortex to occlusion and blink stimulus types. **b)** Population PSTH of 46 recorded units for each stimulus type and also residual response (occlusion-blink). Start of trial is marked by vertical line and transient absence of stimulus is marked with dotted lines.

4.2 Motion-based prediction and motion extrapolation

As we reviewed in chapter 2, the relation between smoothness of motion (temporal coherency) and efficiency of neural presentations of motion has been extensively studied. Yuille and Grzywacz (1989) have shown that the efficiency of motion integration is highly dependent on the smoothness of the trajectory of the stimulus. Also, according to behavioral data, 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 (outlier detection).

Considering temporal coherency of motion during occlusion of stimulus (the case without sensory measurements), velocity estimation is degraded and probabilities are diffused in space and time. However, a model based on temporal coherency may still have enough momentum or motion inertia to propagate estimations about the position of the target. This process will break

down if the occluder gets too long but the motion inertia effect of target motion on distractors is visible (Watamaniuk, McKee, and Grzywacz, 1995a). Inertia of estimated motion will lead to extrapolation of motion trajectory, to fill transient gaps of sensory flow.

Such a prior on the temporal coherency of motion can be defined in a probabilistic framework (See chapter 3). As a consequence, *motion-based prediction* comes into play, that is, including both the position and velocity from the trajectory of motion to predict the future.

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 and 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 in interrupted and noisy trajectories.

4.3 Model: motion-based prediction

In this chapter, to study neural and behavioral signatures of motion extrapolation in interrupted trajectories, we emphasize on the ingredients of motion-based prediction: position and velocity of motion. In simulation of relevant experimental data on eye tracking, we also have studied PV configuration of the model. PX configuration also has been used to asses sufficiency of previous position information in precise estimation of the trajectory. In addition, we have explored the role of prior information on slowness of motion in the estimated velocity of tracking and particularly tracking during the absence of stimulus. Further more, we have assessed robustness of tracking to noise in all configurations: PX, PV and MBP model. All these studies have been repeated for strong and weak priors on the slowness of motion.

Master equations in the model as described in chapter 3 are:

$$p(z_t|I_{0:t-dt}) = \int p(z_t|z_{t-dt}) \cdot p(z_{t-dt}|I_{0:t-dt}) \cdot dz_{t-dt} \tag{4.1}$$

$$p(z_t|I_{0:t}) = p(I_{t-dt:t}|z_t) \cdot p(z_t|I_{0:t-dt}) / p(I_{t-dt:t}|I_{0:t-dt}) \tag{4.2}$$

Generative model of motion coherency defines dependence between successive state vectors of motion z = (x, y, u, v) and $p(z_t|z_{t-dt})$ is computed based on that:

$$\begin{aligned} x_t &= x_{t-dt} + u_{t-dt} \cdot dt + \nu_x \\ y_t &= y_{t-dt} + v_{t-dt} \cdot dt + \nu_y \end{aligned} \tag{4.3}$$

$$\begin{aligned} u_t &= \gamma \cdot u_{t-dt} + \nu_u \\ v_t &= \gamma \cdot v_{t-dt} + \nu_v \end{aligned} \tag{4.4}$$

with

$$\nu_x, \nu_y \propto \mathcal{N}(x, y; 0, D_X \cdot dt) \tag{4.5}$$

$$\nu_u, \nu_v \propto \mathcal{N}(u, v; 0, (\sigma_p^{-2} + D_V^{-1})^{-1} \cdot dt) \tag{4.6}$$

ν_x, ν_y are Gaussian distributions of position blurring where $D_X \cdot \delta t$ is blur value sampled at each time step. Blurring of velocity is done with a sample from ν_u and ν_v Gaussian distributions with standard deviation of $(\sigma_p^{-2} + D_V^{-1})^{-1} \cdot dt$.

Similar to prior information on smoothness of motion, to be more realistic, one can define a constraint on statistically significant slowness of natural motions. As defined in (Weiss, Simoncelli, and Adelson, 2002) prior information in slowness and smoothness of motion can be formulated in velocity prediction, by standard deviation as $(\sigma_p^{-2} + D_V^{-1})^{-1} \cdot dt$ (See equation 4.6). Setting ${\sigma_p}^{-1} = 0$ is equivalent to having no prior on slow speeds.

This constraint can be reflected as γ in equation 4.4 and is defined as $\gamma = (1 + \frac{D_V^2}{\sigma_p^2})^{-1}$. Clearly this factor plays as a gain for tracking and varies depending on the configuration of the model. The update rule (see (Perrinet and Masson, 2007) for a derivation) assumes independence of the prior on slow speeds with respect to predictive prior on smooth trajectories.

Equation 4.2 estimates motion at each time step by Bayesian combination of measured evidence ($p(I_{t-dt:t}|z_t)$ in the window between $t - dt$ and t) with what is predicted by internal model of trajectory expressed in equation 4.1. Based on this, PX, PV and MBP models can be explained as in following:

Position-based prediction (PX configuration): In this configuration of model, only position estimation benefits from internal model and velocity estimation is mainly shaped by measurements. The value of D_V needs to be high, to largely diffuse velocity distribution at each step, therefore motion estimation is independent from velocity information.

Velocity-based prediction (PV configuration): In this configuration, only velocity estimation uses internal model and position estimation is not benefited from that. Value of D_X is high and according to equations 4.3 and 4.4, internal model of trajectory diffuses position information.

Motion-based prediction (MBP configuration): This model utilizes both components of motion signal, position and velocity, for trajectory estimation. ($D_X.dt$) and ($D_V.dt$) are set respectively to 1 and 1.5.

4.4 Results

This section includes results of our investigations on the role of prediction in extrapolation of trajectory, 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 the model upon prediction versus current sensory input. Results are presented in three categories: motion estimation of trajectory (without blank) in presence of increasing levels of background noise, motion estimation

of blanked trajectory with PX, PV and MBP models, and motion estimation of blanked trajectory with background noise. Stimulus is a horizontally moving dot with constant speed and we have used traces of estimated velocity as output of MBP and PV models, as well as histograms of estimated velocity and position in all three configurations of the model.

4.4.1 Predictive tracking in the presence of varying noise levels

Robustness of tracking: We tested the robustness of the 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 (Watamaniuk, McKee, and Grzywacz, 1995a). 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.

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 4.5). 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 configuration to minimally rely on position predictions. To do so, we chose a D_X value high enough so that the model did not favor any estimation in particular. We then repeated the same experiment but with the 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. Also, we found a range of contrast 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 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 4.5 are similar to the outlier detection experiment observed in psychophysics by Watamaniuk, McKee, and Grzywacz (1995a), 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 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 the 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 (Naka and Rushton, 1966) that can adequately describe the different aspects of motion integration (See (Perrinet and Masson, 2007)) (see figure 4.6). 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 of tracking (as defined in figure 4.5). 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 MBP model exhibited a significant change 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 MBP model and complements the analysis done by (Perrinet and Masson, 2012).

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 (Rust et al., 2006; Simoncelli and Heeger, 1998). 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.

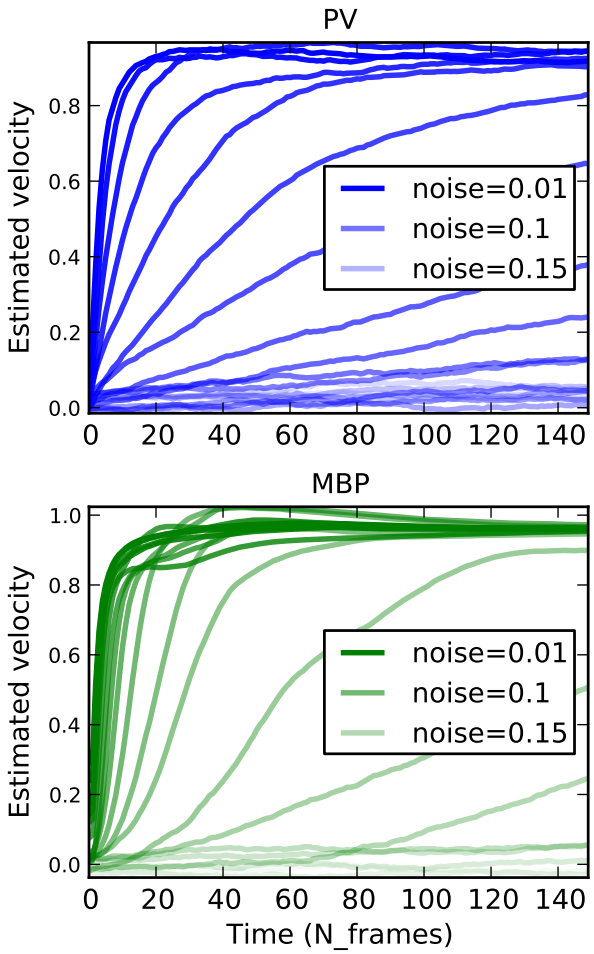

Figure 4.5: **Estimated velocity of PV and MBP configurations averaged over 20 trials**: the stimulus is a horizontally moving dot with $u = 1$ which includes different noise values at background and no blank in trajectory. Colors from dark to light correspond to the response to the stimulus with noise 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.

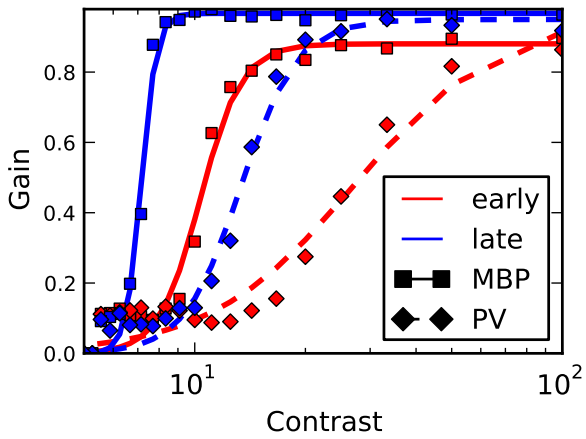

Figure 4.6: **Response gain functions are plotted with best-fit Naka-Rushton functions for both PV and MBP models**: increasing contrast produces a sigmoidal increase in response gain whose shape changes with both time and model configurations. Similar to the psychophysical reports by Watamaniuk, McKee, and Grzywacz (1995a), 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 MBP configuration, indicating a transition from no-tracking to tracking states.

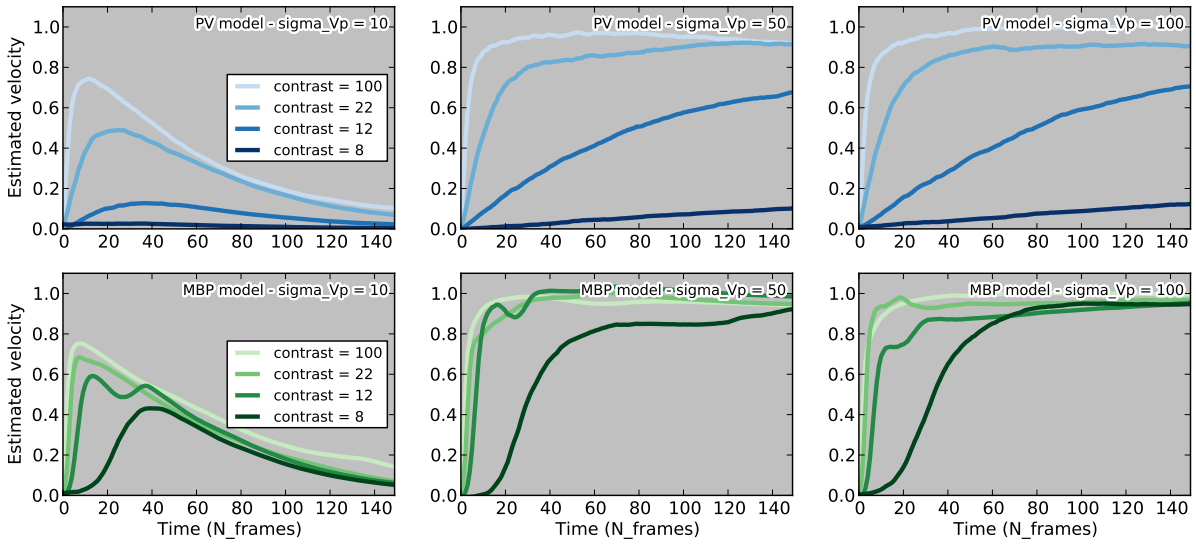

Figure 4.7: **Estimated velocity at different σ_p and contrasts**: **Top row)** PV model **Bottom row)** MBP model

Robustness of tracking and prior information favoring slow speeds: After surveying contrast sensitivity of motion-based prediction, we have studied how preference of internal model for slow speeds may affect tracking in low contrast conditions. To this aim, we have calculated estimated velocities by PV and MPB models for four contrast levels and three σ_p (variance

of speed distribution centered by zero, stated in equation 4.6). Also, histograms of estimated positions by PX, PV and MBP models were calculated, for mentioned contrasts at low and high σ_p values.

Fig 4.7 includes estimated velocities. According to this figure, the MBP model in high enough σ_p is robuster against noise. Figs 4.9- 4.10 illustrate histograms of estimated position by PX, PV and MBP models at strong σ_p, where following Figs 4.12- 4.13 are the same results by imposing weak σ_p.

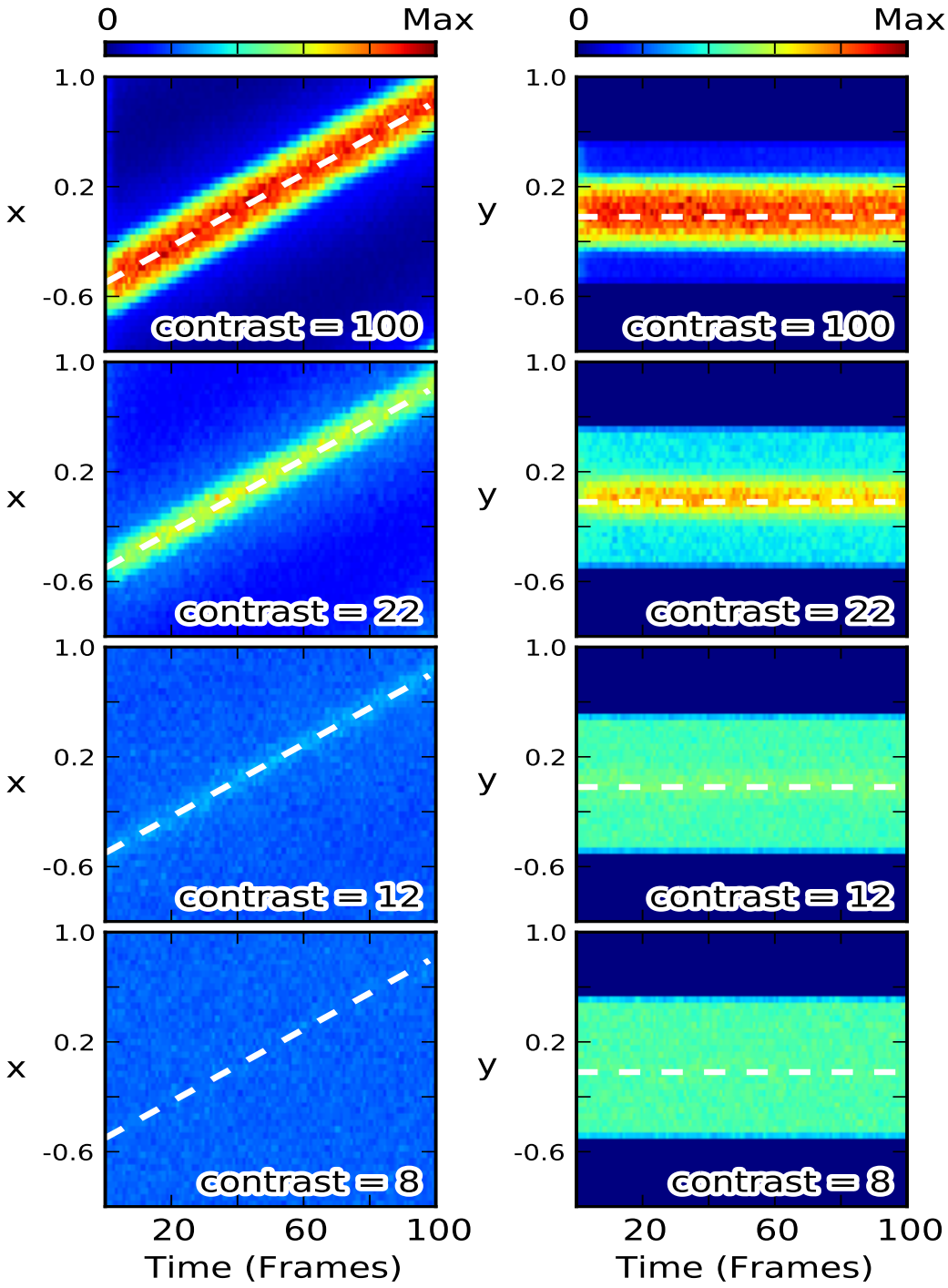

Figure 4.8: **Histograms of estimated position (x, y) by PX model at eight contrast levels and** $\sigma_p = 10$: stimulus is a dot moving horizontally from $(-0.6, 0)$ to $(1, 0)$, dashed white line indicates the center of stimulus

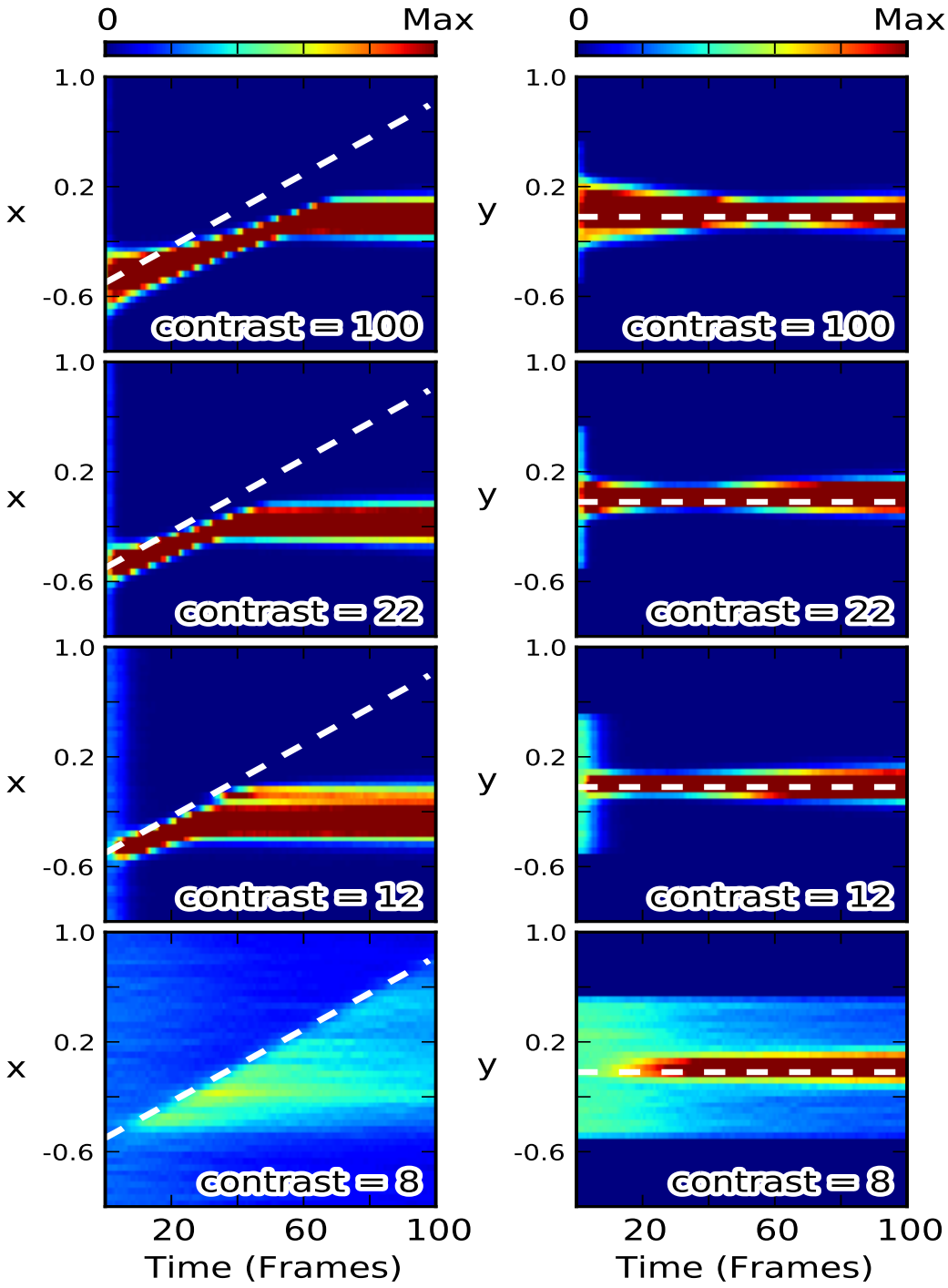

Figure 4.9: **Histograms of estimated position (x, y) by PV model at eight contrast levels and** $\sigma_p = 10$: stimulus is a dot moving horizontally from $(-0.6, 0)$ to $(1, 0)$, dashed white line indicates the center of stimulus

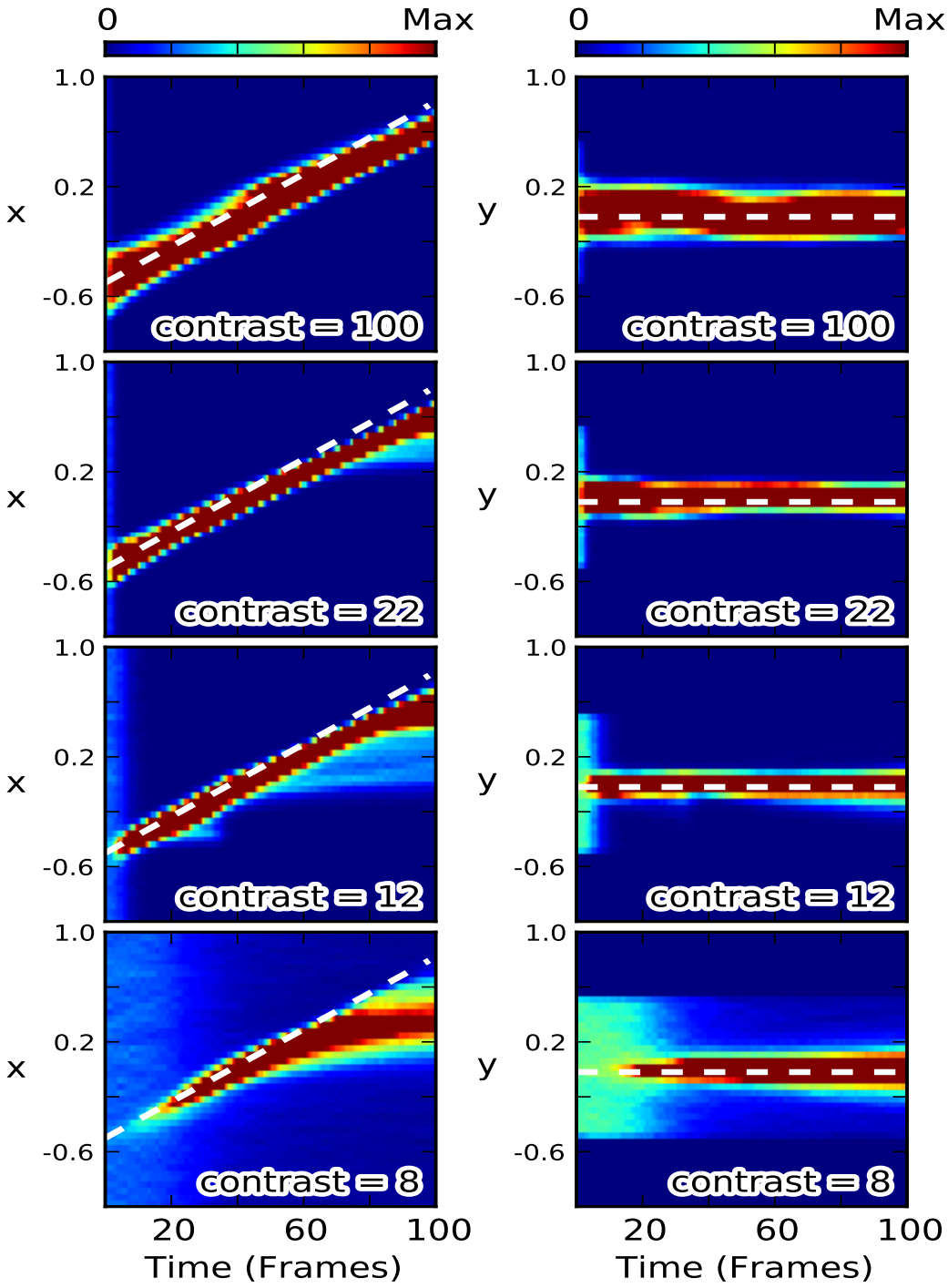

Figure 4.10: **Histograms of estimated position (x, y) by MBP model at eight contrast levels and** $\sigma_p = 10$: stimulus is a dot moving horizontally from $(-0.6, 0)$ to $(1, 0)$, dashed white line indicates the center of stimulus

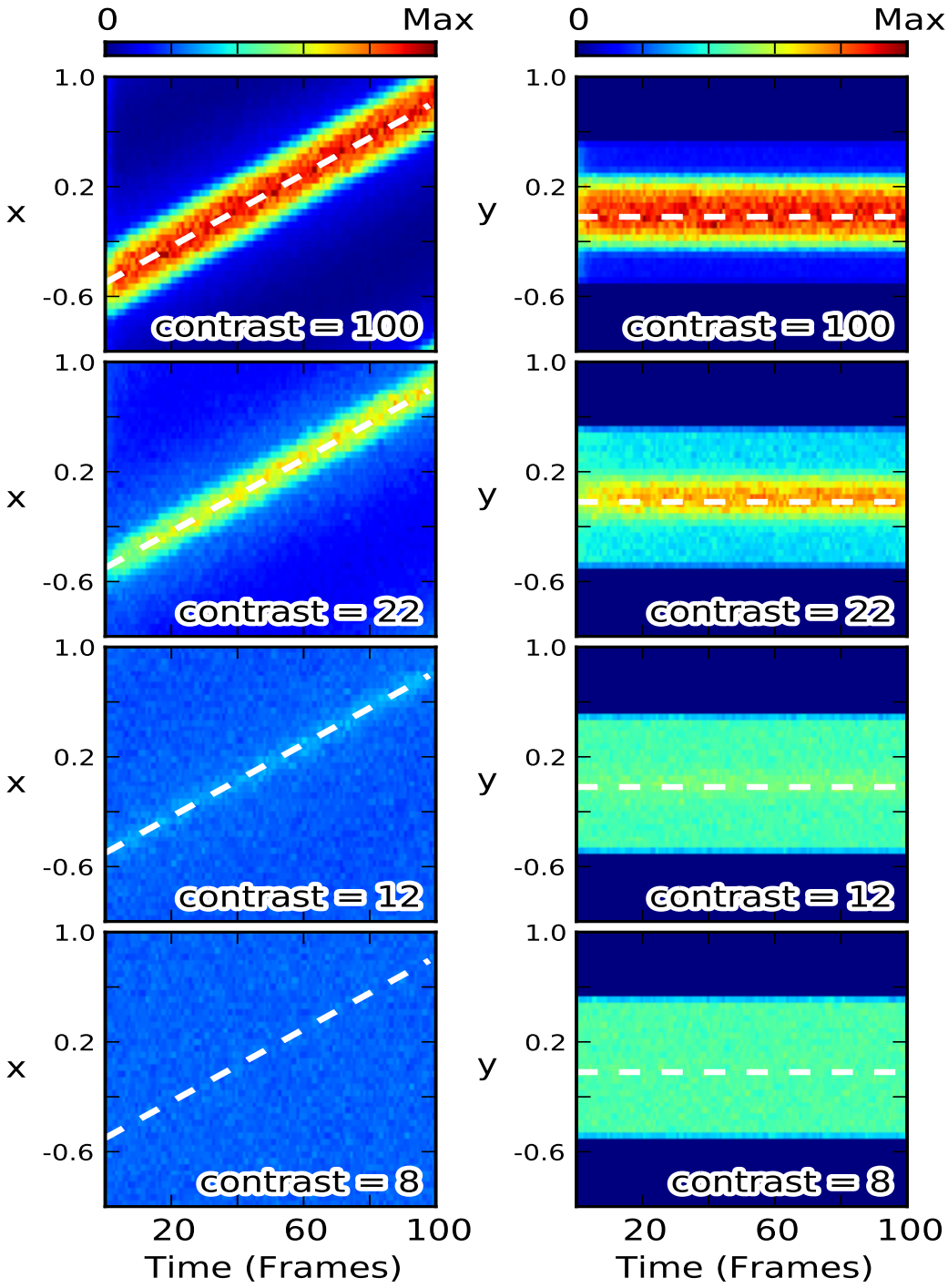

Figure 4.11: **Histograms of estimated position (x, y) by PX model at eight contrast levels and** $\sigma_p = 100$: stimulus is a dot moving horizontally from $(-0.6, 0)$ to $(1, 0)$, dashed white line indicates the center of stimulus

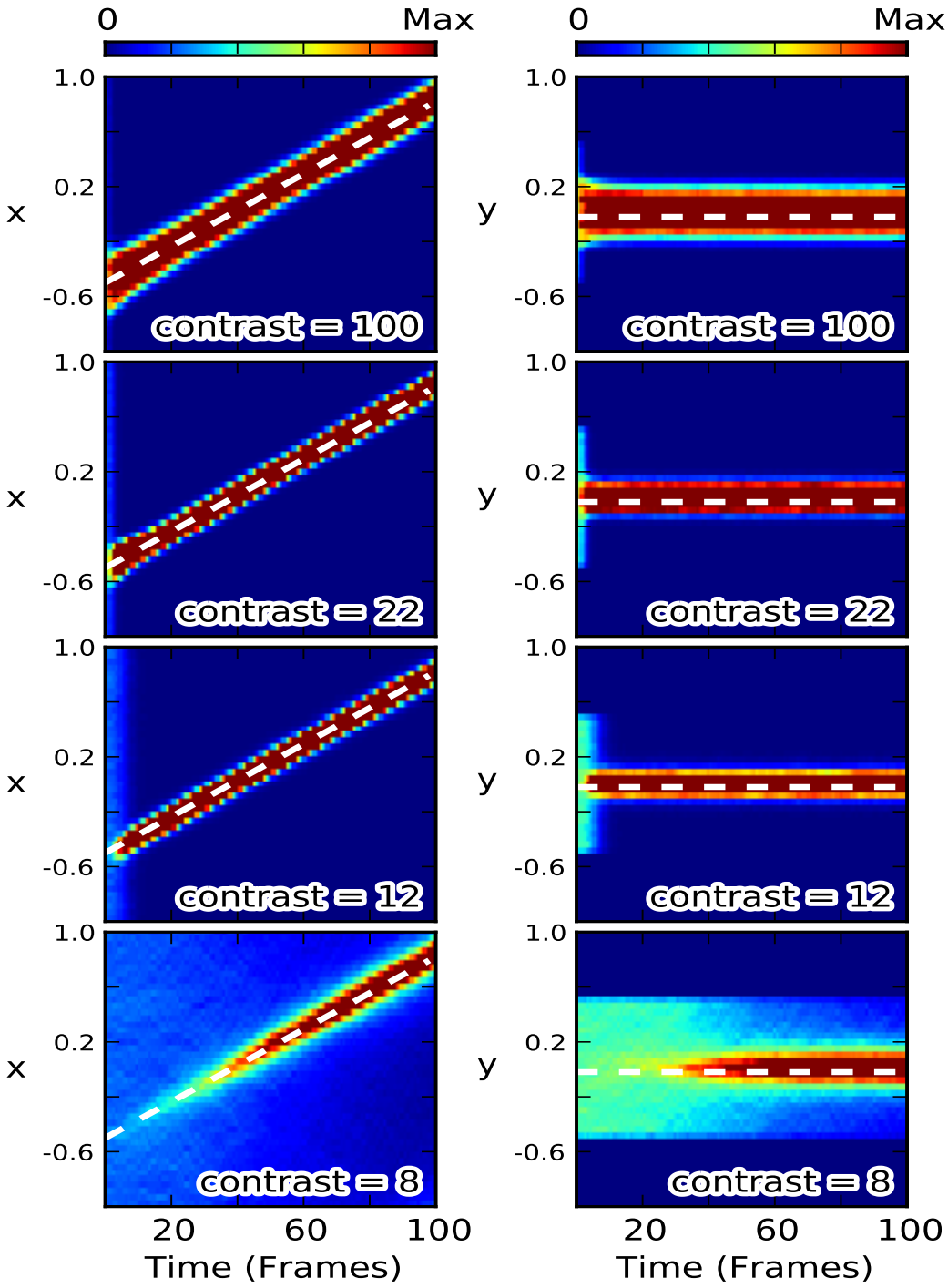

Figure 4.12: **Histograms of estimated position (x, y) by PV model at eight contrast levels and** $\sigma_p = 100$: stimulus is a dot moving horizontally from $(-0.6, 0)$ to $(1, 0)$, dashed white line indicates the center of stimulus

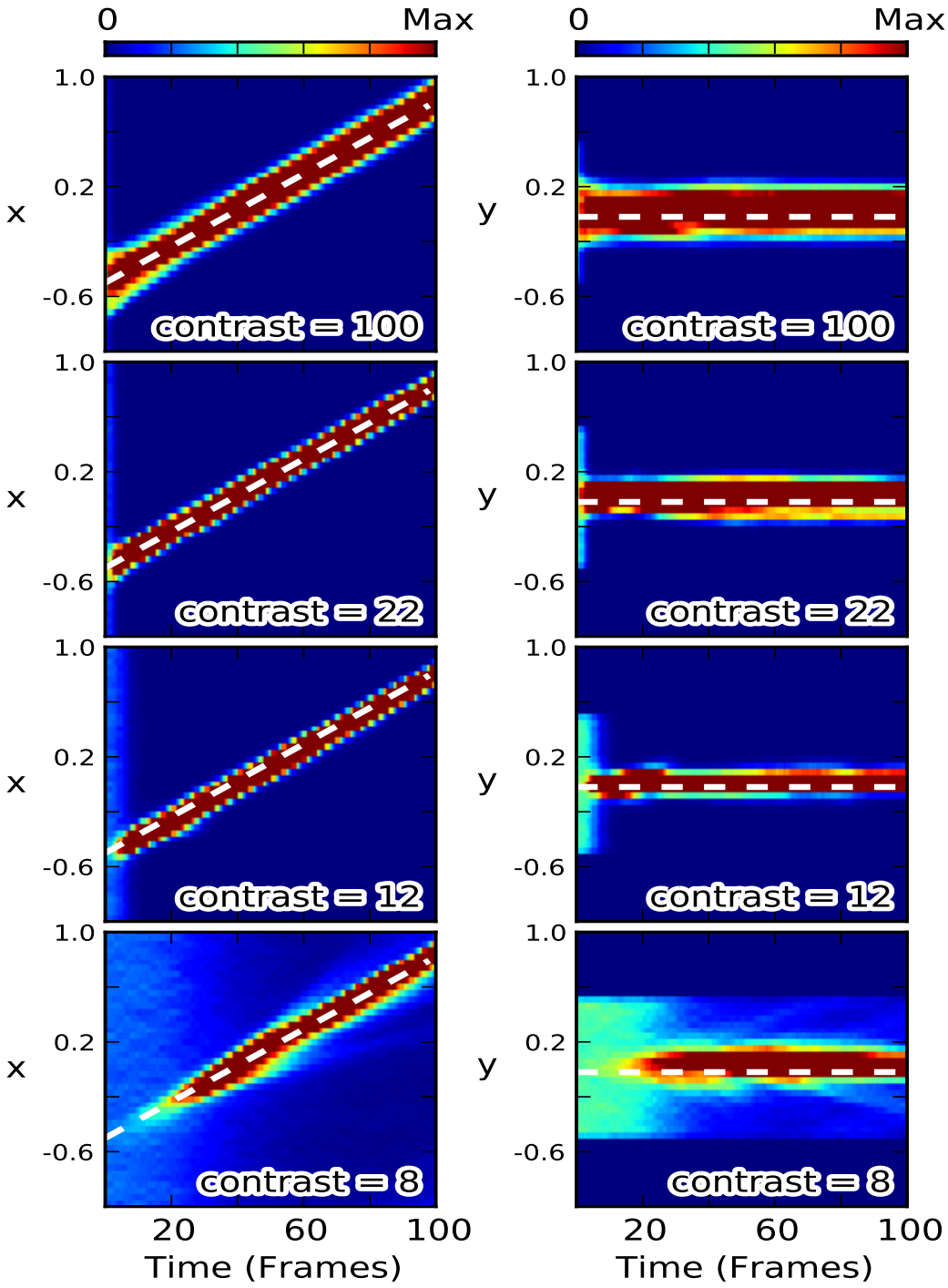

Figure 4.13: **Histograms of estimated position (x, y) by MBP model at eight contrast levels and** $\sigma_p = 100$: stimulus is a dot moving horizontally from $(-0.6, 0)$ to $(1, 0)$, dashed white line indicates the center of stimulus

4.4.2 Extrapolation of motion information in a blanked trajectory

Motion estimation in the temporal absence of stimulus: In this set of model experiments we used as an input the movie of a 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 the previous physiological (Assad and Maunsell, 1995), behavioral (Bogadhi, Montagnini, and Masson, 2013) and theoretical studies (Burgi, Yuille, and Grzywacz, 2000). 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 4.14, we plot the histogram of the estimated positions and velocities obtained with the three different models: PX, PV and MBP. Remember that PX and PV were obtained simply by choosing high values of D_V and D_X, respectively.

In Figure 4.14, for the earliest frames, velocity histograms first spread over a larger area but progressively 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 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 4.15. 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 Fig 4.15, because of the very large variability of estimated velocity observed across trials (See Figure 4.14).

Moreover, the rather small difference observed between PV and MBP models is due to the simplistic dot motions used in the present study. As explained earlier in chapter 3, the sensory

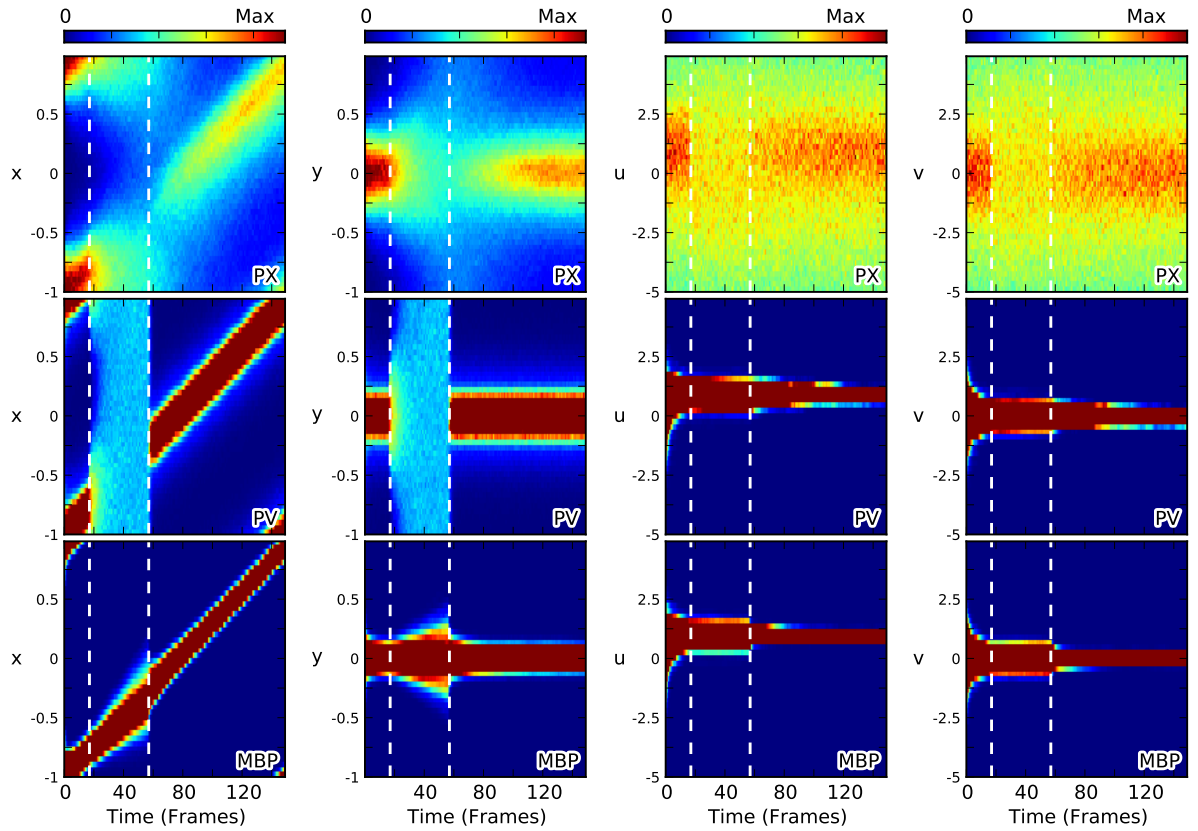

Figure 4.14: **Histograms of estimated motion for a horizontally moving dot with a blanked trajectory under three predictive configurations of the model**: blanking period is indicated with dashed white lines and each row represents full motion estimation under the configuration denoted by the inner title. Each plot illustrates the probability distribution function of a relevant variable (vertical axis) with respect to time (horizontal axis) as in Figure 4.1. 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).

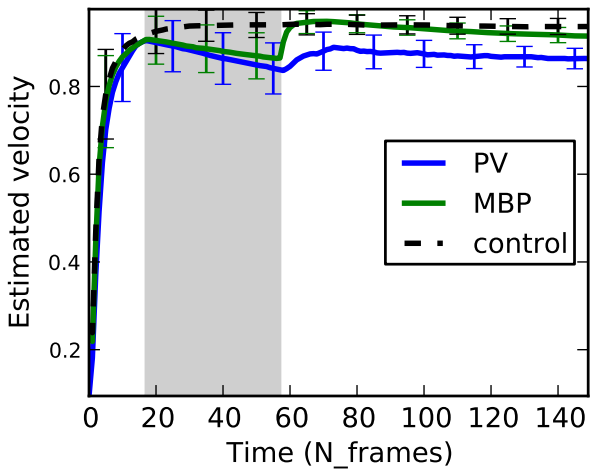

Figure 4.15: **Estimated velocity in PV and MBP configurations of Figure (4.14)**: estimation is measured as the response of the model to a horizontally moving dot with $u = 1$ and a short blank in the trajectory. Blanking period has been shown with shaded area. The trace in black dashed lines represent 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.

layer of both models is made of a bank of motion energy filters which are highly efficient in local detection of straight dot translations. Choosing 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 (Perrinet and Masson, 2012).

Motion estimation during blank, with prior information favoring slow speeds: To study how trajectory estimation during blank may be modulated by prior information on slowness of the trajectory, we have simulated estimated velocities by MBP and PV models for a range of σ_p, from very strong to very week. The results have been summarized in Fig 4.16. In the top left panel, the scheme of σ_p has been plotted, from strongest (lightest hue) to weakest (darkest hue). The top right figure, includes the estimated velocity by MBP model in a control condition (no blank in the trajectory) and for a range of σ_p. Bottom left and bottom right figures respectively correspond to estimated velocities by PV and MBP model, in a blanked trajectory and for shown range of σ_p. Duration of blank has been marked with dashed lines. Two figures shown in the bottom are extension of Fig 4.15 for different levels of preference for slow speeds.

Results of control model highlights necessity of imposing a moderate enough σ_p in the MBP model. Indeed, even in the case of highly detectable stimulus (high contrast dot), strong preference of internal model for slow speeds may cease emergence of tracking (note traces with

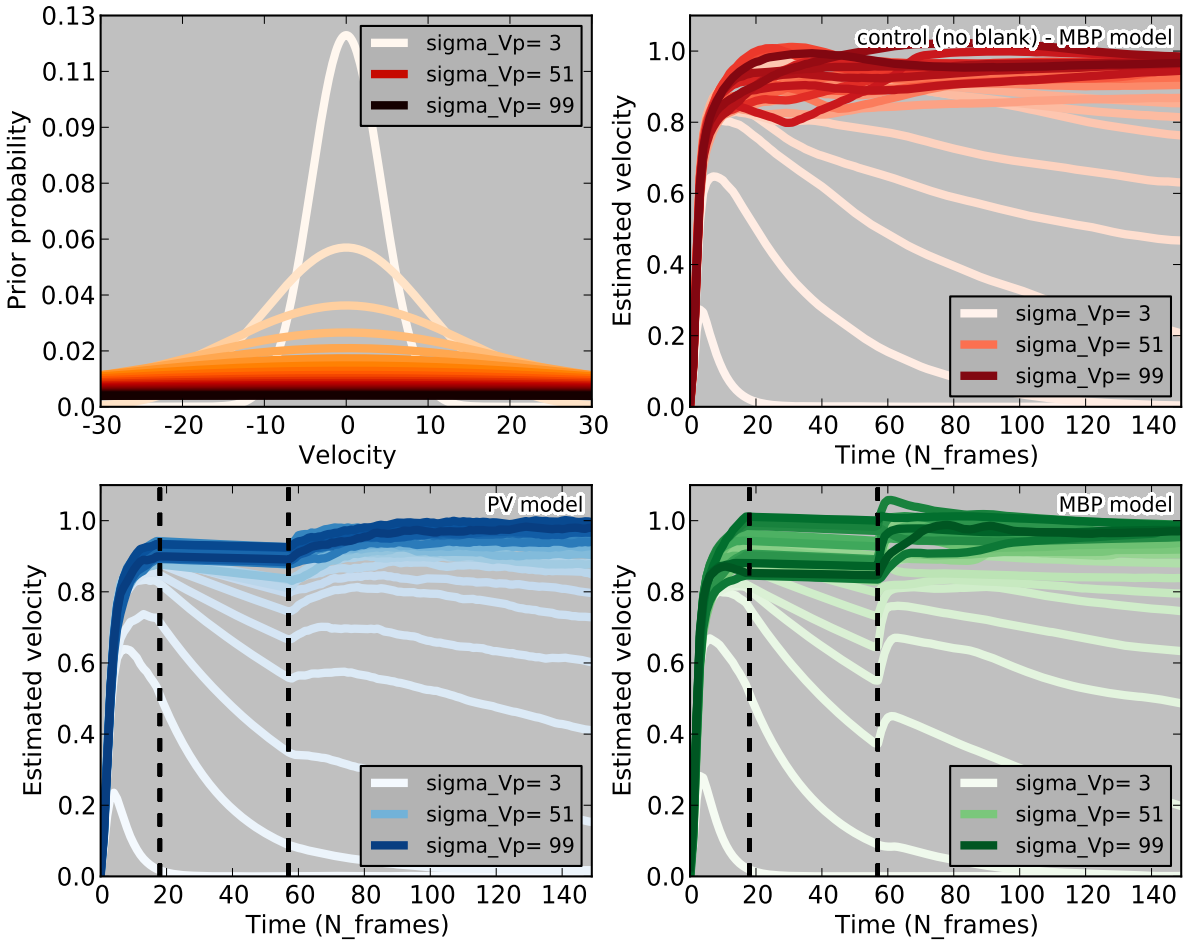

Figure 4.16: **Velocity estimation by PV and MBP models for range of σ_p (prior information on slowness of velocity)**: in all four panels, traces with lightest and darkest colors correspond to the strongest and weakest σ_p, respectively. **Top left):** strength of prior information on velocity, ech trace is a gaussian with standard deviation of σ_p and mean = 0 **Top right):** estimated velocity by MBP model in control condition (no blank in the trajectory), **Bottom left):** estimated velocity by PV model in a blanked trajectory, **Bottom right):** estimated velocity by MBP model in a blanked trajectory

lightest color). Meanwhile, the results of PV and MBP models on blanked trajectories suggest importance of σ_p in velocity drop during absence of stimulus. The velocity traces have been produced with a very simplistic ocular motor plant, integrating and averaging estimated velocity allover positions. If we interpret our results in the context of experimental data shown in Fig 4.2, two signatures of estimated velocity needs to be discussed: First, the fair slope of velocity decrease during the blank and second, quick catch-up of stimulus at the end of blank. In PV and MBP models, the velocity drop, similar to the experimental data, happens for strong enough σ_p, where this desired value would result in poorer tracking and catch-up after blank. For precise motion estimation after reappearance of stimulus, MBP model shows better performance than PV model. (compare velocity catch-up at the end of blank for both models).

4.4.3 Motion extrapolation in noisy blanked trajectory

In the last series of experiments, we combined two different sources of uncertainty by simulating a noisy dot moving along a partially blanked trajectory. We have shown that motion extrapolation requires enough accumulation of information from the observed trajectory parts, to allow the emergence of the tracking state. Moreover, we found that there is a 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 (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 4.17). We extend the results shown in figure 4.15 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 4.15 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 contrasts lower than contrast thresholds, no such velocity catch-up was observed and the models in fact remained in the no tracking state (see figure 4.17). 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 model. 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 4.5 and figure 4.17.

The 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 4.14). 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.

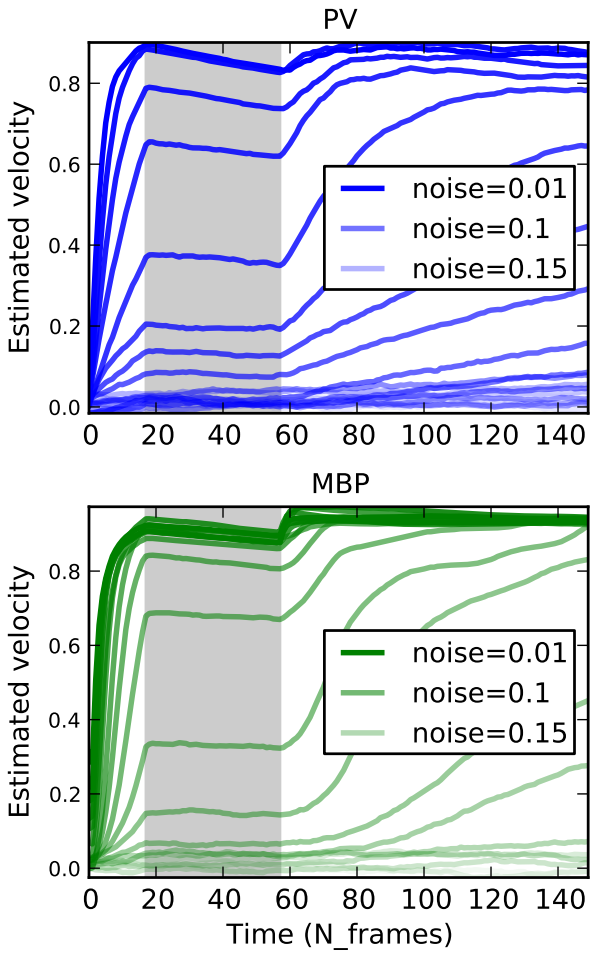

Figure 4.17: **Motion extrapolation with sensory noise**: stimulus is a horizontally moving dot with $u = 1$ which includes a blank as shown with shaded area. In addition there is a sensory noise and colors from dark to light correspond to noise levels increasing linearly between 0.01 to 0.2. **Top)** 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 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.

4.5 Conclusion and Discussion

In this chapter we have explored the competence of motion-based prediction (MBP model) in robust motion estimation. After studying the effect of contrast and σ_p (preference for slow speeds) in robust tracking, particularly we have focused on the problem of fragmented motion

trajectories. In this section, we conclude implication of our results and the accordance with some experimental data.

4.5.1 Motion-based prediction and robustness of tracking

To explore the model dynamics, we tested the robustness of MBP model 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. (Spering et al., 2005)) showing a strong nonlinear relationship between pursuit gain and contrast (see (Masson and Perrinet, 2012) 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 (Rust et al., 2006; Simoncelli and Heeger, 1998)— emerges here as a property of the dynamical system. The theoretical link between Bayesian inference and divisive normalization has been already suggested by several studies (e.g. (Barthélemy et al., 2008; Hürlimann, Kiper, and Carandini, 2002)).

Current results emphasize that dynamical inference as implemented here can also reproduce the temporal dynamics of normalization mechanisms through lateral interactions (Reynaud, Masson, and Chavane, 2012). 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.

Another issue that we explored is the dependence of robust tracking to σ_p (preference for slow speeds). To do so, we have studied contrast dependency of tracking in strong and weak σ_p values. Estimated velocity results highlight that for the same value of σ_p (weak prior on the slowness of motion), tracking by MBP model tolerates higher noise levels. Results of estimated positions by PX, PV and MBP models in strong and weak σ_p values may be concluded as following:

First, in the cases with very strong prior on the slowness of motion, PV model can track the stimulus only at very high contrast, estimated position by PX model lags behind the instantaneous position of stimulus and in lower contrasts this lag happens earlier in the trajectory. The MBP model has better tracking performance than PX, but for low contrast stimului it stops tracking at the end of trajectory. For weak σ_p, tracking of PX and MBP models is accurate but PV model still is able to track only at highest contrasts. Our results on tracking in low contrast are compatible with results of Weiss, Simoncelli, and Adelson (2002) about slow tracking in noisy conditions.

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 (2000) 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.

4.5.2 Motion-based position coding in fragmented trajectories

We investigated the role of motion-based prediction in motion extrapolation during target blanking. This is 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. (Assad and Maunsell, 1995; Becker and Fuchs, 1985; Bogadhi, Montagnini, and Masson, 2013)).

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 the 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 MBP model is more consistent with both physiological (e.g. Assad and Maunsell (1995) and Newsome, Wurtz, and Komatsu (1988)) and behavioral (e.g. Becker and Fuchs (1985), Bennett and Barnes (2003), and Bogadhi, Montagnini, and Masson (2013)) observations. Interestingly, the comparison between PV, PX and the 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 (Burgi, Yuille, and Grzywacz, 2000; de Xivry and Lefèvre, 2007; Perrinet and Masson, 2012).

In addition, we have explored the role of σ_p (preference for slow speeds) in motion estimation during blank. According to experimental results (Becker and Fuchs, 1985; Bennett and Barnes, 2003; Bogadhi, Montagnini, and Masson, 2013), velocity of eye starts to drop at the beginning of blank and then stabilized in a level around half of target velocity and then after reappearance of stimulus starts to grow quickly. In our results, we have reproduced this velocity slope during blank by PV and MBP models. Even though that stronger σ_p would result in quicker drop of velocity during blank but meanwhile degrades the estimated motion even at the presence of stimulus.

An important issue was to answer to the question raised by the experimental study of Assad and Maunsell (1995). 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 (Newsome,

Wurtz, and Komatsu, 1988). 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 the cortex.

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 in fragmented trajectories. For instance, recent behavioral experiments imposing a blank during the straight trajectory of a tilted line shows complementing results in the light of results of Bogadhi, Montagnini, and Masson (2013). 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 level. 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 during blank.

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.

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. Complementing the previous work on the aperture problem (Perrinet and Masson, 2012), 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.

Summary

In this chapter, we have focused on the question of robust tracking and interrupted trajectories. With motion-based prediction model, we have studied motion estimation at fragmented trajectories, during transient absence of stimulus and after reappearance. MBP model simulations is accompanied by PX and PV models as control configurations. The scope of this chapter extends to the questions like how contrast and prior preference of slow speeds may affect motion estimation of fragmented trajectories. In the next chapter we will introduce a modified version of MBP model (so called diagonal model) to deal with neural delays.

Chapter 5

Diagonal model: delay compensation and anticipatory response

Abstract

In this chapter, we have focused on the problem of neural delays and the expected error in the position coding of moving objects. We have presented a modified version of the MBP model, so called diagonal model, that estimates the instantaneous position of moving objects, based on delayed motion measurements. With the diagonal model, we have introduced a delay compensation phase in trajectory processing and explained two main visual signatures: the trajectory-dependent anticipatory response and the Fröhlich effect. Our results highlight that motion information accumulated from the trajectory of a moving object may serve as a correction mechanism for positional errors caused by neural delays.

5.1 Problem statement: positional error caused by neural delays

It takes a definite time for the visual system to convey the luminous signals captured from visual scene to different layers of visual cortex. In humans, this time is estimated to be around 100 ms (Nijhawan, 1994). It is composed of two terms: a fixed delay caused by axonal transfer of sensory signals known as *neural transfer delay*, and the delay associated with processing time. Existence of these unavoidable delays suggests that we may always perceive moving scenes with a permanent lag and behind their actual position. This would introduce inefficiencies in the visual system for subtle tracking tasks and the accurate, on-time detection of the position and speed of objects. A tennis player is a good example: the successful motor action is produced to hit the moving target (the ball), in spite of delayed arrival of information (See Fig 5.1). Therefore, there may exist some compensatory mechanisms in the neural representation of motion.

Studying neural delays is also related with development of anticipatory response in neural populations. This may happen in populations which are very likely to have the stimulus in their

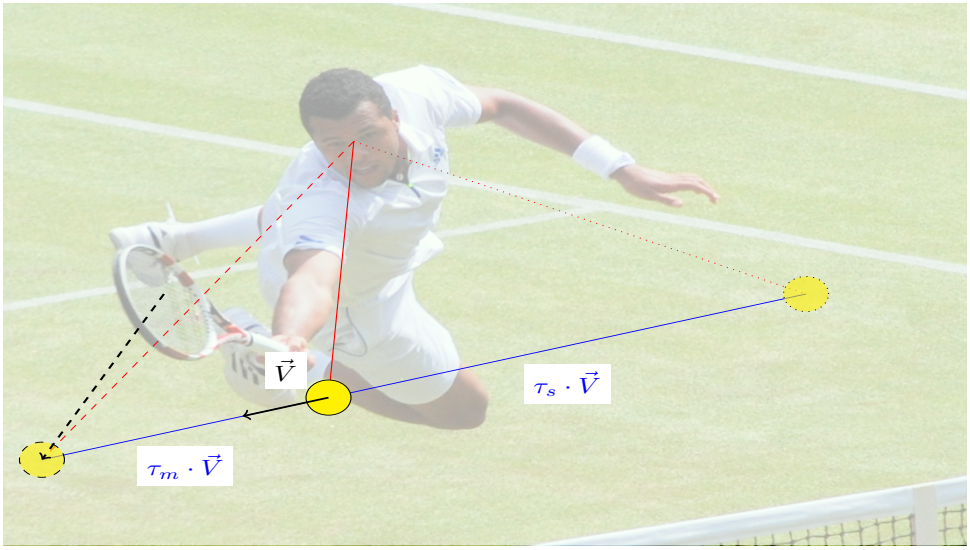

Figure 5.1: **The problem of neural delays for a tennis player**: player needs to track the instantaneous position of target and emit an accurate motor action to hit it somewhere ahead of the current position. In the image, $\vec{V}$ is the velocity of target ball and the circle with solid edges corresponds to the physical position of ball at a specific moment. The player appears to be looking at the solid circle and getting prepared to hit the target in the position of dashed circle. On the other hand, according to inevitable delays in the dynamic of sensory and motor systems, the instantaneous sensory information about the position of the ball arrives with a delay of (τ_s) and yet, there is another delay between sensory information delivery and emission of a suitable action (τ_m). How a tennis player is accurate enough 1) to look at the "correct" position of ball with delayed sensory information? 2) to be efficient enough to hit the ball on time?

receptive field in the near future, based on the history of motion.

In this chapter, we have studied predictive motion processing of smooth trajectories in a generic probabilistic framework for motion estimation (See chapter 2). Using the *motion-based prediction* algorithm for motion estimation and considering neural delays, we proposed a *diagonal model* for predictive motion estimation at current time, having access to the sensory information associated to some time steps before.

In particular, in line with experimental works on anticipatory aspects of motion integration in the visual cortex (Benvenuti et al., 2011; Guo et al., 2007), we investigated the dependence of estimated motion on the length of trajectory, at specific points of motion path. We found two distinct regimes in motion estimation of a smoothly moving target: first regime is in the early trajectory, in which position of stimulus is estimated with low confidence and with delay. In this phase, there is an evident anticipatory signature in estimated motion and it is dependent on the relative position of the estimated point. In the second regime which governs the late trajectory, tracking has already emerged and estimation confidence is maximum and all points in the late trajectory have very similar anticipatory profile.

Our findings on smooth trajectory estimation can be associated with a distinct size of population receptive fields in visual cortex: all specific positions laid on the early trajectory can represent the retinotopic position of neural activity which is in the range of horizontal

connections, while the positions in the late trajectory are pre-stimulated with the whole range of horizontal connections (limited length of horizontal connections, limited size of population receptive fields).

Furthermore, we studied the role of position estimation and velocity estimation separately and found that excluding velocity information from motion estimation process deteriorates the estimation significantly. In this regard, we are able to explain the Fröhlich effect as well: misperception of earliest parts of the trajectory may be caused by non-precise estimation of velocity.

This chapter is prepared in the following order: first we have a brief introduction on the study of trajectory prediction, and then section 5.3 provides description of MBP model and diagonal implementation of it. Section 5.4 includes our results in modeling of spatiotemporal processing of smooth trajectory. Finally, section 5.5 concludes our findings and explains advantages and limitations of our approach, in comparison with the studies reviewed in chapter1.

5.2 Smooth prediction of a trajectory and motion-based position coding

In this section, we briefly highlight the important aspects of trajectory prediction and the way that we have approached them in the following sections.

Bayesian framework and internal model of motion: In chapter 1, we reviewed some studies done on compensatory mechanisms for neural delays, motion anticipation and facilitatory mechanisms in processing of coherent predictable trajectories. Considering these studies, it is encouraging to define a generic, large-scale framework to explore the shared neural mechanism among all mentioned signatures of predictive coding. Some of previous works, addressed questions on how predictability of a stimulus may affect neural processing at different levels and various cortical areas (See chapter 1). Also there are some modeling efforts to define notion of internal model in visual perception (Changizi, 2001; Changizi and Widders, 2002; Erlhagen, 2003) and to explain the shared mechanisms in different processing aspects of trajectory motion.

Indeed, an internal model of visual perception can be well expressed in a Bayesian framework and one can explain successive steps of motion computations in terms of Bayesian **Priors** and **Likelihoods**. Then interactions of internal model (shaped by priors on smoothness and slowness of motion) with instantaneous sensory likelihood would result in **posterior** distribution. Estimated motion including probability distributions of position and velocity of stimulus, is posterior of such a Bayesian inference.

In the current framework, we have studied such an internal model for coherent trajectories. This allows investigations on predictive localization of stimulus and dependence of motion processing to the trajectory and contextual information. This is also related to the questions about localization of neural activity. For instance, characterization of stimulated neural populations,

directly from delayed arrival of sensory information and the knowledge of the trajectory. Particularly, we stressed the role of velocity information to explain quicker and facilitated processing of coherently moving objects. We keep internal predictive representation of motion as a central hypothesis and aim to explain facilitating neural phenomena that might result from that as well as contributions to delay compensation.

Accumulation of trajectory information: Our modeling framework provides precise probability distributions on the position and velocity of stimulus. As the stimulus moves along the trajectory, the spatiotemporal distribution of probabilistic position code is varied. Indeed, positions ahead of the current location of stimulus start to accumulate belief about approaching arrival of stimulation. For positions distributed on different parts of the trajectory, we surveyed the development of anticipatory response as a pre-activation alert from an approaching stimulus.

Motion extrapolation in early and late trajectory: To extrapolate the estimated response based on a delayed arrival of input, the internal representation is engaged to overcome uncertainty and to provide a more reliable and robust sensory encoding. For a reasonable range of delays, the model builds up the position of the stimulus at the current time based on the last observed input and ends up with a correct estimation, taking advantage of the motion signal. In results (section 5.4), we have focused on gradual increase of estimation confidence in the trajectory. Considering delays, a time lag is expected at estimation of the earliest part of the trajectory (Fröhlich effect) (Fröhlich, 1923), while for a highly detectable moving object far enough from its initial position the correct estimation is already there.

5.3 Diagonal model: motion-based prediction with axonal delays

As mentioned in the previous section, our main objective in this chapter is to study the role of predictability of trajectory on the dynamics of position coding and development of anticipatory response in neural populations. As we described in chapter 3, the internal Bayesian model reflects our expectancy from the visual scene based on prior information on coherency of motion (See chapter 2 for detailed description of model and implementation). To study the effect of delays, the framework integrates sensory information associated with some steps before, with what is suggested by internal model.

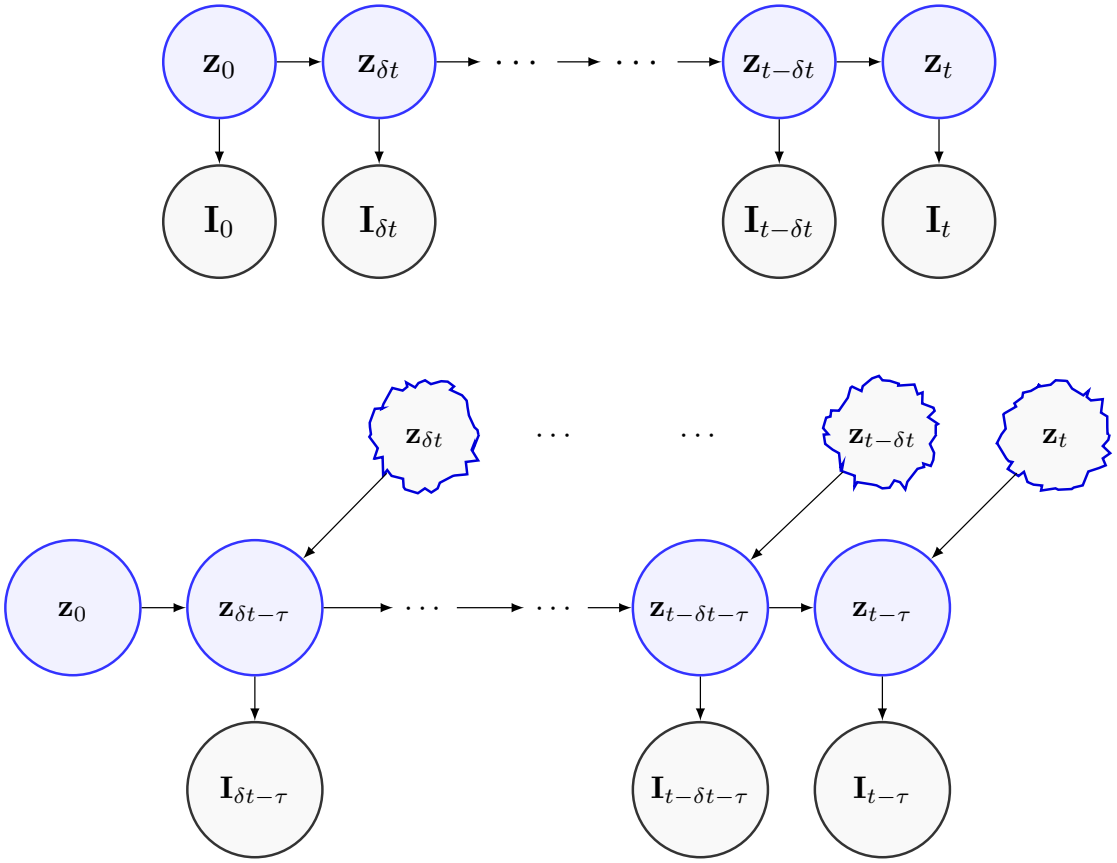

Figure 5.2: **Diagonal Markov chain**: **Top)** classic Markov chain for state transition of system: in current study, the estimated state vector is $z_t = \{x, y, u, v\}$ composed of position and velocity of moving stimulus and measurements are performed from the image at current time, I_t. **Bottom)** Diagonal model of motion-based prediction: considering τ as neural delay, there is no measurement at $t = 0$ and ultimate state estimation at any time step is z_t. Diagonal scheme of motion estimation includes two steps: first, the motion state $z_{t-\tau}$ is estimated based on smoothness constraint and delayed sensory information $I_{t-\tau}$, then according to a prior knowledge on late arrival of sensory information, the response is extrapolated for a period of virtual blank ($duration = \tau$), without sensory measurements.

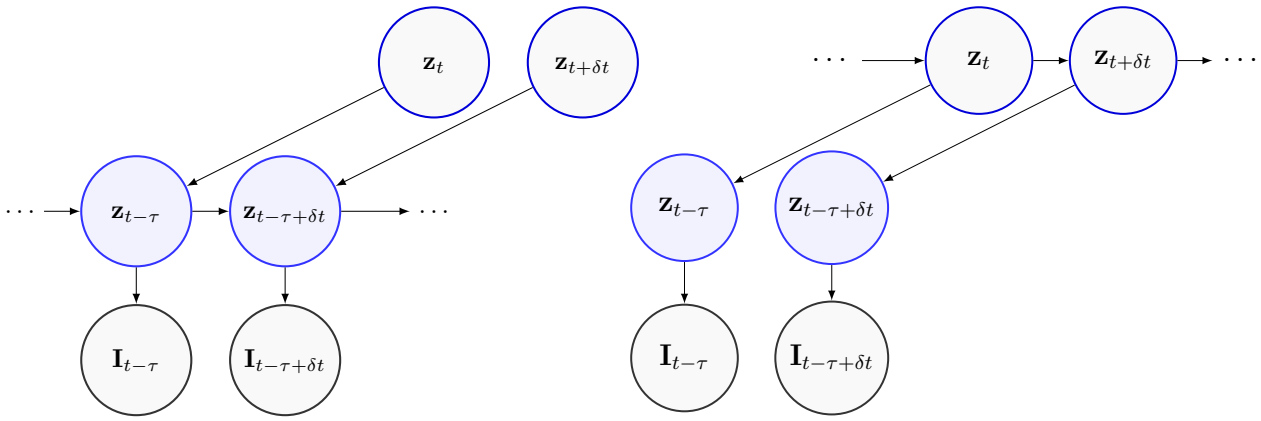

Figure 5.3: **Two different modes of diagonal Markov chain**: **Left)** diagonal model, pushing mode **Right)** diagonal model, pulling mode

5.3.1 Diagonal model and delayed access to sensory input

Considering neural delays, predictive position coding can be implemented by "pushing forward" the population response in the direction of motion such as to compensate the delay. Our model uses motion signal including position and velocity of moving object to extrapolate the trajectory response to the positions which are most likely to be covered by the stimulus. As illustrated in Fig 5.2, classical Markov chain for state estimation of stimulus can be redrawn in a diagonal fashion. Diagonal scheme of motion states illustrates the dependence of the extrapolated state $z(t)$ to the state suggested by delayed motion information $z(t-\tau)$, where τ is the (known) value of delay. Note that the earliest part of the trajectory is necessarily missed (motion estimation starts at $t = \delta_t - \tau > 0$, as there is no sensory information before), but the next states should become realistic estimations of the actual, present position of stimulus and trajectory and prediction overcomes the restrictive effect of delay. The diagonal model of motion extrapolation was originally proposed by Nijhawan and Wu (2009) to explain the detailed mechanism of motion extrapolation by retinal ganglion cells (See chapter 2). Here we use it as an abstract rule in predictive motion estimation.

As described in chapter 3, motion (position and velocity) is described by the state vector $z_t = \{x_t, y_t, u_t, v_t\}$, such that the master equations of the model are:

$$\textit{estimation:} \quad p(z_t|I_{0:t}) \propto p(I_{t-\delta t:t}|z_t) \cdot p(z_t|I_{0:t-\delta t}) \tag{5.1}$$

$$\textit{prediction:} \quad p(z_t|I_{0:t-\delta t}) = \int dz_{t-\delta t} \cdot p(z_t|z_{t-\delta t}) \cdot p(z_{t-\delta t}|I_{0:t-\delta t}) \tag{5.2}$$

Considering delayed access to sensory input in diagonal model (See Fig 5.2), the current state $p(z_t|I_{0:t-\tau})$ can be predicted using an extrapolation step, "pushing" the past state to the present.

As a consequence, master equations of diagonal model can be written as:

$$\textit{estimation:}\ \ p(z_{t-\tau}|I_{0:t-\tau}) \propto p(I_{t-\tau-\delta t:t-\tau}|z_{t-\tau}) \cdot p(z_{t-\tau}|I_{0:t-\tau-\delta t}) \tag{5.3}$$

$$\textit{prediction:}\ \ p(z_{t-\tau}|I_{0:t-\tau-\delta t}) = \int dz_{t-\tau-\delta t} \cdot p(z_{t-\tau}|z_{t-\tau-\delta t}) \cdot p(z_{t-\tau-\delta t}|I_{0:t-\tau-\delta t}) \tag{5.4}$$

$$\textit{extrapolation:}\ \ p(z_t|I_{0:t-\tau}) = \int dz_{t-\tau} \cdot p(z_t|z_{t-\tau}) \cdot p(z_{t-\tau}|I_{0:t-\tau}) \tag{5.5}$$

As can be seen from these equations, this mode is similar to Eqs. 5.1 and 5.2, except that these are delayed by τ, the known sensory delay. This information $p(z_{t-\tau}|I_{0:t-\tau})$ is then "pushed" forward in time using the extrapolation step. We finally obtain the estimate of motion at the current time, knowing the information acquired until $t-\tau$, that is $p(z_t|I_{0:t-\tau})$.

This "pushing" mode is consistent with the diagonal model (see Figure 5.3-Left) and is the most intuitive. It may be however more practical (in particular for neural network implementations) to derive a set of predictive steps that would directly act on the estimation on the state at the current time $p(z_t|I_{0:t-\tau})$. Note that:

$$p(z_t|I_{0:t-\tau}) = \int dz_{t-\tau} \cdot p(z_t|z_{t-\tau}) \cdot p(z_{t-\tau}|I_{0:t-\tau}) \tag{5.6}$$

$$\propto [\int dz_{t-\tau} \cdot p(z_t|z_{t-\tau}) \cdot p(I_{t-\tau-\delta t:t-\tau}|z_{t-\tau})] \cdot p(z_{t-\tau}|I_{0:t-\tau-\delta t}) \tag{5.7}$$

$$\propto [\int dz_{t-\tau} \cdot p(z_t|z_{t-\tau}) \cdot p(I_{t-\tau-\delta t:t-\tau}|z_{t-\tau})] \cdot \int dz_{t-\tau-\delta t} \cdot p(z_{t-\tau}|z_{t-\tau-\delta t}) \cdot p(z_{t-\tau-\delta t}|I_{0:t-\tau-\delta t}) \tag{5.8}$$

Regrouping terms, it comes:

$$p(z_t|I_{0:t-\tau}) \propto \int dz_{t-\tau} \cdot p(z_t|z_{t-\tau}) \cdot [\int dz_{t-\tau-\delta t} \cdot p(z_{t-\tau}|z_{t-\tau-\delta t}) \cdot p(I_{t-\tau-\delta t:t-\tau}|z_{t-\tau}) \cdot p(z_{t-\tau-\delta t}|I_{0:t-\tau-\delta t})] \tag{5.9}$$

Noting that the term within brackets can be written as a prediction from $t-\tau$ to t, we have:

$$p(z_t|I_{0:t-\tau}) \propto \int dz_{t-\delta t} \cdot p(z_t|z_{t-\delta t}) \cdot p(I_{t-\tau-\delta t:t-\tau}|z_t) \cdot p(z_{t-\delta t}|I_{0:t-\tau-\delta t}) \tag{5.10}$$

where

$$p(z_t|z_{t-\delta t}) \cdot p(I_{t-\tau-\delta t:t-\tau}|z_t) \cdot p(z_{t-\delta t}|I_{0:t-\tau-\delta t}) = \int dz_{t-\tau} \cdot p(z_t|z_{t-\tau}) \cdot [p(z_{t-\tau}|z_{t-\tau-\delta t}) \cdot p(I_{t-\tau-\delta t:t-\tau}|z_{t-\tau}) \cdot p(z_{t-\tau-\delta t}|I_{0:t-\tau-\delta t})] \tag{5.11}$$

Assuming $p(I_{t-\tau}|z_{t-\delta t}) = p(I_{t-\tau}|z_t)$ for vanishingly small δt:

$$\textit{prediction:}\ \ p(z_t|I_{0:t-\tau}) = \int dz_{t-\delta t} \cdot p(z_t|z_{t-\delta t}) \cdot p(z_{t-\delta t}|I_{0:t-\tau}) \tag{5.12}$$

$$\textit{estimation:}\ \ p(z_{t-\delta t}|I_{0:t-\tau}) \propto p(I_{t-\tau-\delta t:t-\tau}|z_t) \cdot p(z_{t-\delta t}|I_{0:t-\tau-\delta t}) \tag{5.13}$$

$$\textit{extrapolation:}\ \ p(I_{t-\tau-\delta t:t-\tau}|z_t) = \int dz_{t-\tau} \cdot p(z_t|z_{t-\tau}) \cdot p(I_{t-\tau-\delta t:t-\tau}|z_{t-\tau}) \tag{5.14}$$

We call this second mode the "pulling" mode. These modes share the same processing logic but they offer different implications about the manner that internal model and likelihoods might be implemented:

In pushing mode motion state $z_{t-\tau}$ is estimated based on delayed sensory input $I_{t-\tau-\delta t:t-\tau}$ and smoothness of motion. Duration δ_t is step size of estimation and τ represents the imposed delay. Equation 5.3 calculates the probability of a desired motion state, using likelihood of that state (measured by delayed sensory information), and predicted weight given by equation 5.4. In the next step, the estimated motion is extrapolated for a period of virtual blank (duration = τ), while there is no sensory measurements (see Figure 5.3-Right). Thus, extrapolative step shown by equation 5.5 is purely predictive with the smoothness constraint and prior information of the value of delay τ (see section 5.3.2).

In pulling mode the probabilistic representation is different, the current state is directly estimated based on the delayed measurements and extrapolative step is hidden in the probability $p(z_t|I_{0:t-\tau-\delta t})$. Under the stationarity assumption for prediction, both modes are mathematically equivalent and produce probabilistic representation of instantaneous state based on delayed measurements. Information about the estimate of motion (position, velocity) at time t knowing the sensory information observed between 0 and $t-\tau$ is contained in the pdf $p(z_t|I_{0:t-\tau})$. As we saw it can be computed using the diagonal model in push mode and summarized in the following master equations:

$$p(z_t|I_{0:t-\tau}) \propto \int dz_{t-\delta t} \cdot p(z_t|z_{t-\delta t}) \cdot p(I_{t-\tau-\delta t:t-\tau}|z_t) \cdot p(z_{t-\delta t}|I_{0:t-\tau-\delta t}) \tag{5.15}$$

$$p(I_{t-\tau}|z_t) = \int dz_{t-\tau} \cdot p(z_t|z_{t-\tau}) \cdot p(I_{t-\tau}|z_{t-\tau}) \tag{5.16}$$

Equations 5.15 and 5.16 are master equations of diagonal model which provide probabilistic distribution of estimated motion state z_t, based on delayed motion measurements $I_{t-\tau}$. In the following part we describe how the pdf functions $p(z_t|z_{t-\tau})$ and $p(z_t|z_{t-\delta t})$ are computed.

5.3.2 Motion-based prediction

Predictive representation of motion favors smooth trajectories in two steps: first, temporal coherency of motion at each step, represented by $p(z_{t-\tau}|z_{t-\tau-\delta t})$ or $p(z_t|z_{t-\delta t})$, is calculated:

$$\begin{aligned} x_{t-\tau} &= x_{(t-\tau-\delta t)} + u_{(t-\tau-\delta t)} \cdot (\delta t) + \nu_x \\ y_{t-\tau} &= y_{(t-\tau-\delta t)} + v_{(t-\tau-\delta t)} \cdot (\delta t) + \nu_y \end{aligned} \tag{5.17}$$

$$\begin{aligned} u_{t-\tau} &= \gamma \cdot u_{(t-\tau-\delta t)} + \nu_u \\ v_{t-\tau} &= \gamma \cdot v_{(t-\tau-\delta t)} + \nu_v \end{aligned} \tag{5.18}$$

$$\nu_x, \nu_y \propto \mathcal{N}(x, y; 0, D_X \cdot \delta t) \tag{5.19}$$

$$\nu_u, \nu_v \propto \mathcal{N}(u, v; 0, (\sigma_p^{-2} + D_V^{-1})^{-1} \cdot \delta t) \tag{5.20}$$

As we described in chapter 3, ν_x, ν_y are Gaussian distributions of position blurring where $D_X \cdot \delta t$ is blur value sampled at each time step. Blurring of velocity is done with a sample from ν_u and ν_v Gaussian distributions with standard deviation of $(\sigma_p^{-2} + D_V^{-1})^{-1} \cdot \delta t$, where σ_p is the standard deviation of prior information on slowness of motion.

Then $p(z_t|z_{t-\tau})$ is calculated by a predictive extrapolation of estimated motion (equations 5.21-5.22). The estimated motion is extrapolated forward, for the duration of τ and based on the knowledge from trajectory and a prior knowledge on the fixed delay τ :

$$\begin{aligned} x_t &= x_{t-\tau} + u_{t-\tau} \cdot (\tau) + \omega_x \\ y_t &= y_{t-\tau} + v_{t-\tau} \cdot (\tau) + \omega_y \end{aligned} \tag{5.21}$$

$$\begin{aligned} u_t &= \gamma \cdot u_{t-\tau} + \nu_u \\ v_t &= \gamma \cdot v_{t-\tau} + \nu_v \end{aligned} \tag{5.22}$$

$$\omega_x, \omega_y \propto \mathcal{N}(x, y; 0, D_X \cdot \tau) \tag{5.23}$$

$$\omega_u, \omega_v \propto \mathcal{N}(u, v; 0, (\sigma_p^{-2} + D_V^{-1})^{-1} \cdot \tau) \tag{5.24}$$

As defined in (Weiss and Fleet, 2001), prior information in slowness and smoothness of motion can be formulated by standard deviation as $(\sigma_p^{-2} + D_V^{-1})^{-1} \cdot dt$ on velocity. 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 σ_p. The update rule (see (Perrinet and Masson, 2007) for a derivation) assumes independence of the prior on slow speeds with respect to predictive prior on smooth trajectories.

5.3.3 Neural interpretation of diagonal model

Motion-based prediction in diagonal fashion can be implemented in a two-layered neural network. Fig 5.4 illustrates the simplest demonstration of the algorithm in a neuronal structure. Source

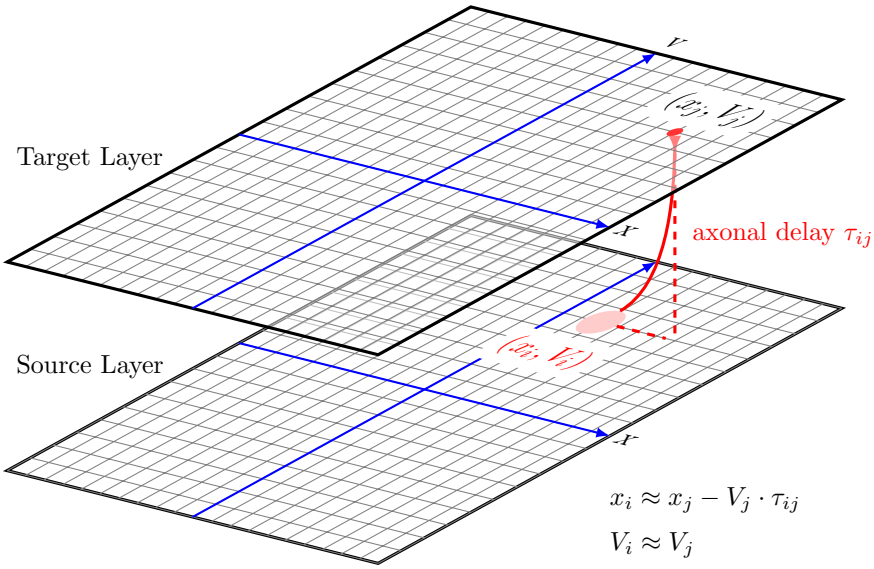

Figure 5.4: **Schematic of neural implementation of diagonal model,** diagonal delay compensation can be demonstrated in a two-layered neural network including source (input) and target (predictive) layers. Source layer receives delayed sensory information and encoded position and velocity in the source layer (x_i, V_i) stimulates the neural populations corresponding to the corrected position of stimulus (x_j, V_j) in the target layer. Degree of correction depends on the precision of encoded velocity and the prior knowledge of neural delay τ_{ij}

layer provides neural predictions and activates specific populations of the target layer. In particular, mapping between source and target layers is anisotropic, depending on the estimated velocity and neural delay τ. For instance, in the case of rightward motion, the target layer may be interpreted as a neural population which gets stimulated by sensory information received by some lefthand neurons. Note that diagonal representation does not change fundamentally the MBP model that we presented in chapter 3 and (Khoei, Masson, and Perrinet, 2013). This new representation, at abstract level, introducs a prior knowledge on delay τ. As a consequence, it provides distinct layers for demonstration for delayed arrival of stimulus and predictive neural activities.

We have implemented diagonal model in a spiking neural network (SNN) via anisotropic connections (Kaplan et al., 2014). The SNN model demonstrates the anisotropic diffusion of predicted neural activations and reproduces the same family of results that we present in the next section.

5.4 Results

This section includes our results in two sub-sections: first, we have studied the development of an anticipatory response for the delayed arrival of stimulus in diagonal MBP model, and in the second part, we have extended the previous results and studied the spatiotemporal profiles of predictive processing for smooth trajectories. In all results described in this chapter, we have

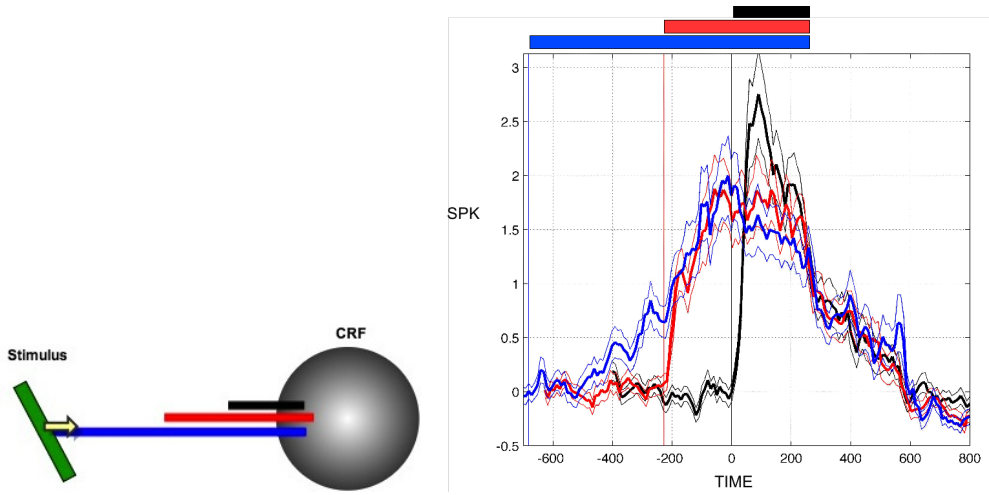

Figure 5.5: **Anticipatory response of V1 populations, studied by Benvenuti et al. (2011) Left)** Stimulation protocol of (Benvenuti et al., 2011). A coherently moving stimulus (green bar) is approaching to the CRF of target population at V1 of macaque monkey. The experiment is repeated for three different trajectory lengths $(1.5°, 3°, 6°)$, each shown with a different color, where the speed of motion is $6.6°/sec$. **Right)** Spiking response of neural population (CRF) to each stimulation condition has been illustrated by the corresponding color. Horizontal axis indicates time after arrival of stimulus in CRF. Anticipatory signature of this data is evident by development of response at time $t < 0$: in red and blue stimulation conditions, response to the moving bar started before arrival of stimulus to CRF. We have studied how MBP diagonal model may explain such a signature.

studied estimated position for a horizontally moving dot presented during 128 frames. All results have been averaged over 10 trials and estimated positions are calculated as histograms over the whole range of trajectory, where the range is divided into 400 positional bins.

5.4.1 Delayed arrival of stimulus and probabilistic anticipatory response

As described in chapter 1, there are experimental evidences to support existence of anticipatory neural mechanisms in processing of an object moving in a coherent trajectory (Guo et al., 2007; Roach, McGraw, and Johnston, 2011). In another study, the existence of similar mechanism has been reported in extra cellular recordings of V1 populations of macaque monkey (Benvenuti et al., 2011). In Fig 5.5 simple experimental protocol of this study has been shown as well as recorded profile of anticipatory response. In the experimental protocol, averaged population response has been recorded from an Utah array located in V1 of a macaque monkey. The stimulation protocol includes a moving bar approaching to the receptive field of recorded population with different trajectory lengths. As it is evident in neural responses, a neural population starts to develop an anticipatory response before arrival of stimulus to its CRF and having a longer trajectory before stimulation of CRF would result in earlier onset of rising phase of anticipatory response. Inspired by these data, we have estimated instantaneous position of stimulus by MBP and PX diagonal models.

As described in the previous section, the diagonal model processes delayed sensory information

by forward extrapolation of estimated positions and starts to prepare a response by extrapolating the delayed state, taking advantage of previous position and velocity of stimulus.

To study this experimental framework with our modeling approach, we have simulated a counterpart experiment within the model. In an electrophysiological experiment Benvenuti et al. (2011) recorded from a target neural population, while a moving stimulus approaches the CRF with different trajectory lengths. In modeling framework we have studied development of probabilistic confidence about arrival of stimulus in target positions ahead of current position of stimulus. From this view, having a long trajectory before arrival to CRF would be equivalent to studying a more distant position in the trajectory.

Figs 5.6 and 5.7 include results of delayed MBP and PX models with $\tau = 100\ ms$, where the stimulus is a horizontally moving dot and each frame of movie stimulus has been arbitrarily assigned to 10 ms of biological time. Fig 5.6 illustrates histograms of estimated position, during the period that the stimulus moves between -1 and 0. In MBP model, estimated position is corrected very early in the trajectory, just after receiving the first measurements ($\tau = 100\ ms$). Estimated position by PX model is less precise and stays lagged behind the veridical position of stimulus.

We have further explored the estimated position in the early part of trajectory. As illustrated in Fig 5.7, we have chosen three successive positions of trajectory to study temporal distribution of estimated response in those points. In progressive responses (left column) the time corresponded to the arrival of stimulus center in that position has been illustrated with dashed lines of matched color. In centered responses (right column), as zero in time axis matches to the arrival time of stimulus in the under study positions, development of belief about arrival of stimulus on target position is more evident.

Comparison between diagonal MBP and PX models reveals that, by progression of the stimulus in the length of trajectory, MBP model systematically grows the confidence of correct position estimation. While PX model provides a delayed and poor estimation of position, no matter how far the current position is from the beginning of the trajectory. In other words, excluding velocity information from the estimation algorithm results in a motion estimation model which equally processes the positions in early and late trajectory (See right bottom plot in Fig 5.7). Here, we have replicated the experimental data on macaque monkey (Benvenuti et al., 2011). The diagonal MBP framework, compensates the imposed delay τ, at the beginning of trajectory. Also the temporal distribution of estimation position shows dependency to the relative location of the point in the trajectory. In the next section, we further study this emergent property of this probabilistic model.

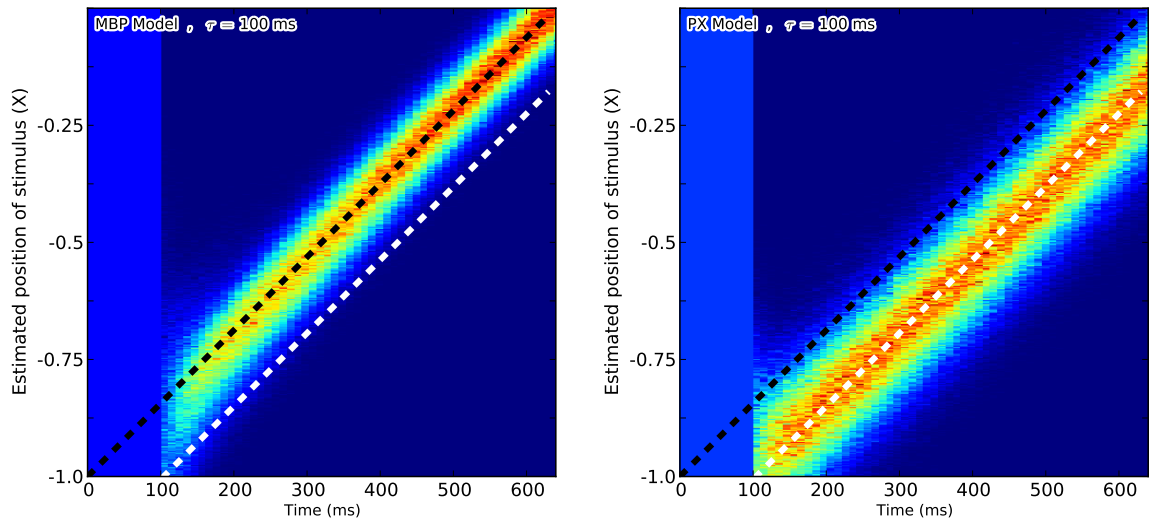

Figure 5.6: **Estimated position of stimulus by delayed MBP and PX models with** $\tau = 100\ ms$, the stimulus is moving between -1 and 0 and the hue shows the probability of estimated positions (dark blue and red respectively correspond to the lowest and highest probabilities). Instantaneous position of the stimulus is marked by dashed black line, where white line matches to the delayed measurements from position of stimulus.

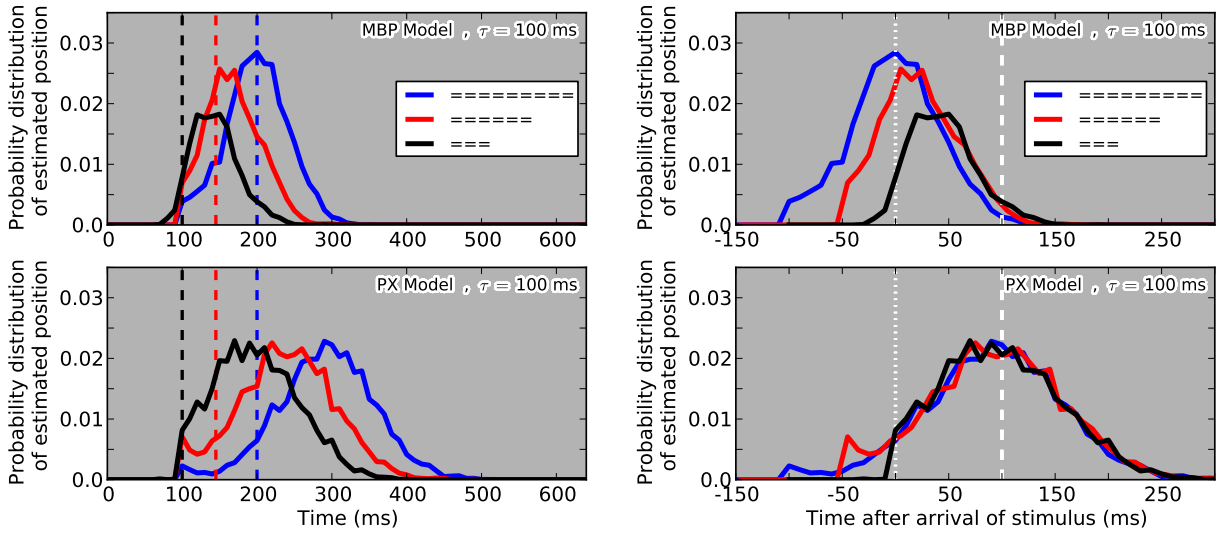

Figure 5.7: **Estimated position of stimulus by delayed MBP and PX models at three successive points of trajectory**, simulating the experimental protocol described in Fig 5.5. Stimulus is a dot moving in a coherent path and inset plots depict the relative position of studied points. **Left)** progressive response: estimated response has been illustrated as it develops over time, dashed lines are matched to the arrival time of stimulus to that specific position. **Right)** centered response: estimated response centered by arrival time of stimulus to the each of three positions (white dotted line), white dashed line corresponds to the time in which model receives measurements of stimulus being present in each of the three position ($\tau = 100\ ms$).

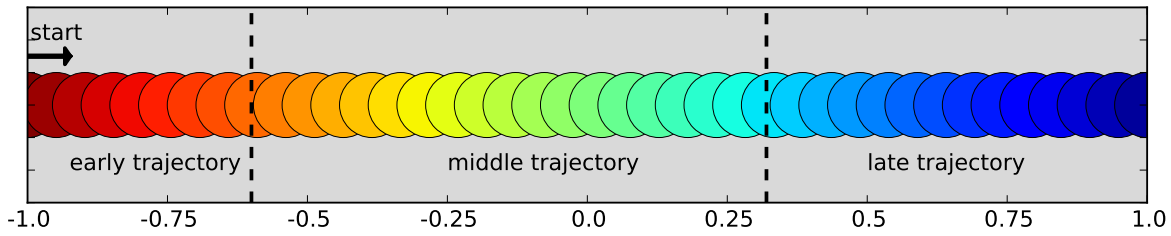

Figure 5.8: **Spatial range of trajectory sampled in number of discrete positions**: early, middle and late trajectories are color coded.

5.4.2 Predictive processing of smooth trajectory: spatiotemporal extrapolation of motion

According to Fig 5.7, the diagonal MBP model provides a predictive representation of motion by anisotropic diffusion of belief at future positions in the trajectory, where PX lacks such a mechanism. This model also suggests a large scale constraint on the anticipatory response of neural populations before arrival of stimulus into their CRF. In next figures, we have extended this result by elaborating the estimated position allover the trajectory, and for different delay values. As shown in Figure 5.8, the position of stimulus in the trajectory is color coded and in one period it moves from reddish positions toward blueish ones. Temporal profile of estimated positions at each point of the trajectory is illustrated with the corresponding color.

Fig 5.9 shows the dynamics of estimated response at sampled points of the trajectory by progression of stimulus along the path, for PX and MBP diagonal models with $\tau = 100\ ms$. As mentioned earlier, MBP model corrects the positional error at early trajectory by developing anticipatory response, where estimations of PX model lag behind the physical position of the stimulus, until the end of trajectory. MBP model also uses information from the trajectory to grow the estimation confidence and reaches its maximum just after middle of trajectory, while this effect is not visible in PX model.

In Figs 5.10 and 5.11, motion estimations of two models have been illustrated for three delay values (each column) and few samples from early, middle and late trajectory (each row). Black dashed lines in Figs 5.10 and 5.11 indicate the time in which the center of stimulus arrives in that specific position of the trajectory (actual physical position of stimulus). White dashed lines indicate the delayed position of stimulus, the one that is available for the model via measurments. In MBP model and in the case with delay $\tau = 0\ ms$, estimation confidence grows by time and by advancing in the trajectory. Also the time in which estimated position reaches to its peak matches to the arrival time of stimulus center to each position.

Imposing more delay results in misestimation of the earliest trajectory (the difference between peak time and black dashed line in couple of early positions), where this position mismatch between real position of stimulus and estimated position is corrected quickly for next positions and the effect of delay is compensated. In PX model, even in the case with no delay $\tau = 0\ ms$,

peak of estimated positions happens later than the time matched to the arrival of the stimulus in that position. For the cases with bigger delays, estimated position is matched to the delayed stimulus information and lags behind the actual position.

Comparison of MBP and PX models reveals robustness of MBP estimations for higher delay values as well as having a saturated response for late enough positions of the trajectory. What we have studied in previous part to simulate anticipatory response of neural populations in V1 can be seen in the reddish part of trajectory, where the width of the temporal profile is sensitive to the distance from the start point of the trajectory.

Figs 5.12-5.13 illustrate the spatial distributions of estimated positions in diagonal PX and MBP models with $\tau = 100\ ms$. The temporal path of the stimulus is shown in the inset plot and estimated position at each time step is of the same color. All traces are subtracted from the actual position of the stimulus, thus the black dashed line corresponds to the center of stimulus. On the other hand, having a delay equal to $\tau = 100\ ms$ means that at each time step the motion measurements would be associated to the position of stimulus at $t - \tau$. This delayed position has been marked with dashed white line. Allover the trajectory, estimated position by PX model is behind the physical center of stimulus, while MBP model after few initial frames corrects the positional lag.

5.4.3 Dependence of delay compensation with respect to the contrast

In Fig 5.14 we have studied MBP and PX model responses for varying contrast levels and in different parts of the trajectory. The aim of this experiments was to reveal the effect of contrast in delay compensation mechanism. In these figures the spatial distribution of the estimated response at three different segments have been illustrated: averaged spatial distributions over first 5 frames (delay correction phase), and over middle and late trajectories. Dashed black and white lines respectively correspond to the actual and delayed centers of stimulus. According to these results, in the early part of trajectory and for MBP model, higher contrast leads to quicker delay compensation (bigger spatial extrapolation), where it is not the case for PX model. In middle and late parts of the trajectory, as delay has been already compensated, the center of responses in all contrast levels is the same and matched to the actual center of stimulus. On the other hand, the spatial profile of response in PX model, as described in Fig 5.10 and 5.11, is always lagging behind the actual position of stimulus and delay is not reduced. Particularly, unlike the early trajectory estimations by MBP model, there is no contrast dependent lead in response.

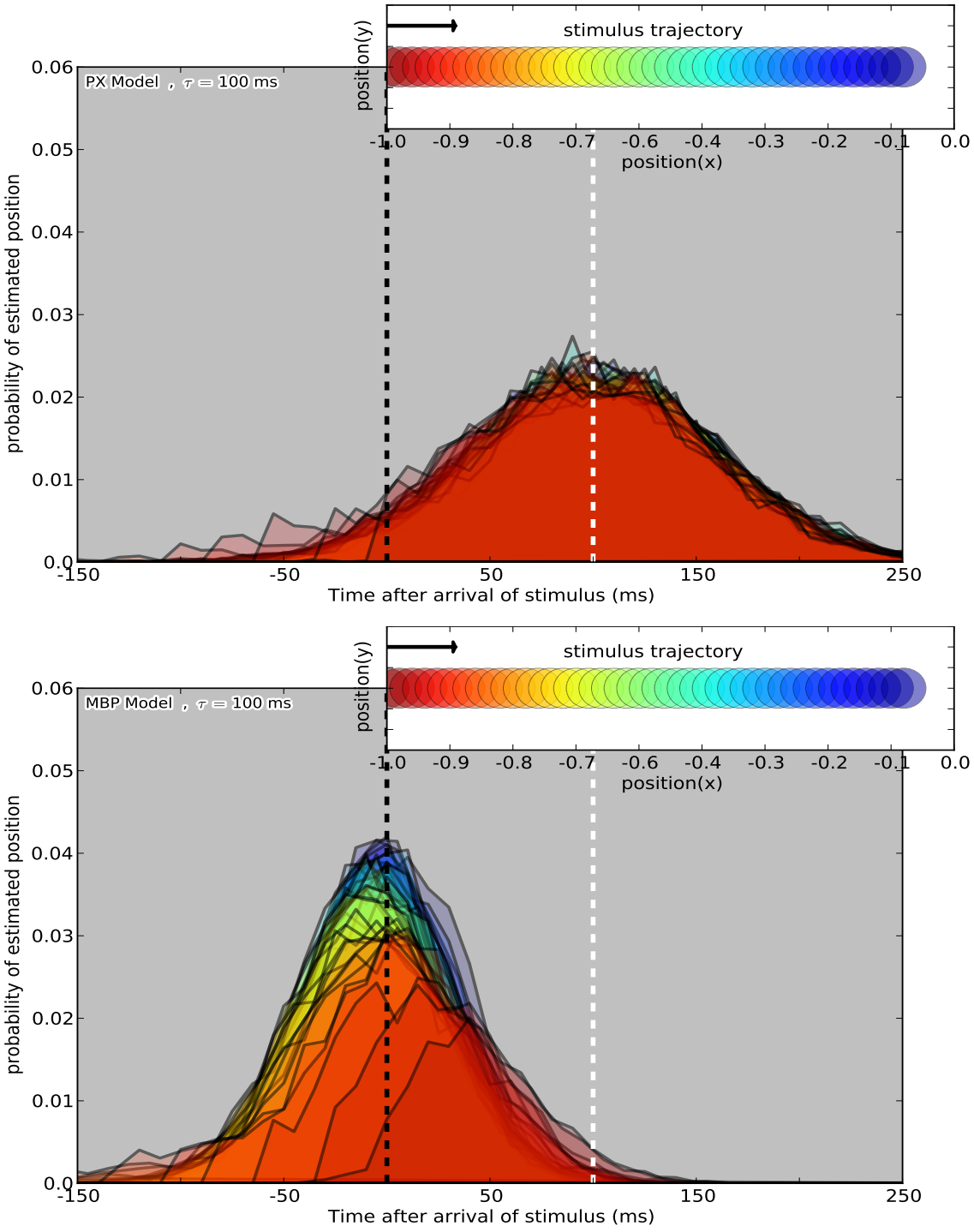

Figure 5.9: **Temporal distribution of estimated positions by diagonal models** ($\tau = 100\ ms$)**:** as shown in the inset plot, stimulus is a gaussian dot moving from reddish positions in the early trajectory toward bluish ones in the late trajectory. Estimations at sampled positions of trajectory have been illustrated with the corresponding color. The first temporal profile is slightly ahead of start point of trajectory, as the first estimation starts at $t = \tau$ (Fröhlich effect). All traces are centered by the time in which actual center of stimulus arrives to that specific position (black dashed line). The relative time in which model receives delayed motion information is indicated by white dashed line. **Top)** PX model: estimated positions allover the trajectory are delayed and of the same profile shape, **Bottom)** MBP model: estimated positions at early trajectory are delayed but gradually this delay is compensated, as it is evident by shift of peak time toward zero. Also the shape of temporal profile is dependent on the length of the trajectory behind each position.

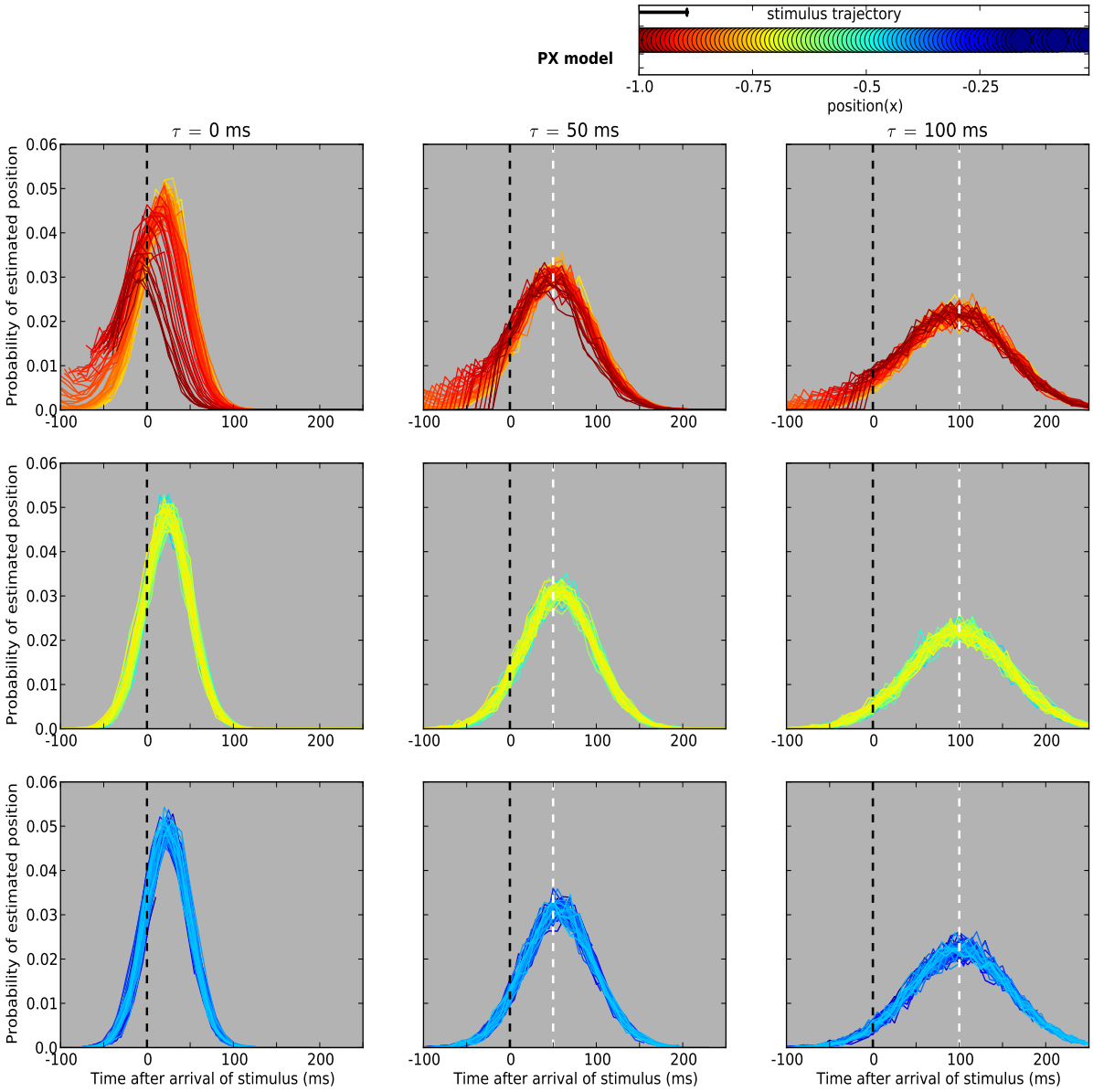

Figure 5.10: **Temporal profile of estimated positions by diagonal PX model**: each column illustrates estimated profiles from start to the end of trajectory for PX model with a delay value noted as τ. As shown in the inset plot, the spatial range of motion trajectory is sampled and shown with a spectrum of colors from red to blue. For clarity, trajectory is divided to three part and estimated profiles at sampled positions of each part are plotted. Profiles are centered by the time in which actual center of stimulus arrives to that position. Hence, temporal lag of estimated position is reflected in the distance between peak of profile and dashed black line. White dashed white line indicates the time in which model receives delayed arrival of sensory information ($t = \tau$). Results of PX model suggests that even at the case with $\tau = 0$ the positional response is slightly lagged behind the actual position of stimulus. For $\tau > 0$ peak of response is matched to the white dashed line. Estimated positions by PX model are poor and delayed and of the same shape for any arbitrary position in the trajectory. Note that in second and third columns where $\tau > 0$, plotted profiles are start from estimated position of stimulus at $t = \tau$ (just after Fröhlich effect)

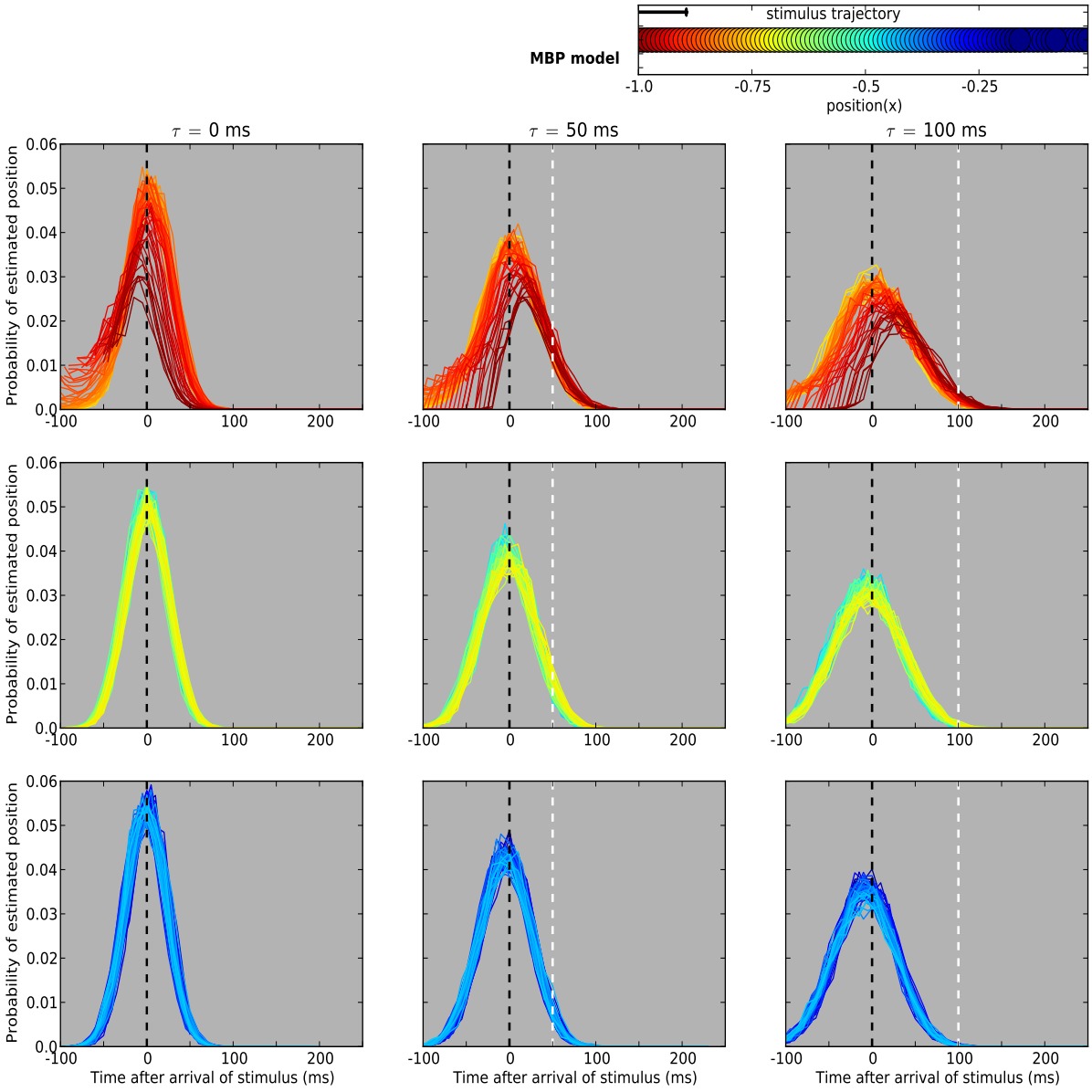

Figure 5.11: **Temporal profile of estimated positions by diagonal MBP model**: position of stimulus in the trajectory is color coded and estimated position by model at each sampled position is plotted by the corresponding color. Each column includes estimated profiles along the trajectory by MBP model with a delay noted as τ on the top. For clarity, the profiles have been plotted in three part. All profiles are centered by the time in which actual center of stimulus arrives to that position. White dashed line indicates the time in which model receives delayed arrival of sensory information. Note that in second and third columns where $\tau > 0$, plotted profiles are starting from position of stimulus at $time = \tau$ (just after Fröhlich effect). According to the results of MBP model, for $\tau > 0$, earliest estimations are lagged behind actual position of stimulus (reddish profiles), but the positional error is corrected quickly and in the middle and late parts of trajectory profiles are centered by the actual center of stimulus, despite delayed arrival of motion information.

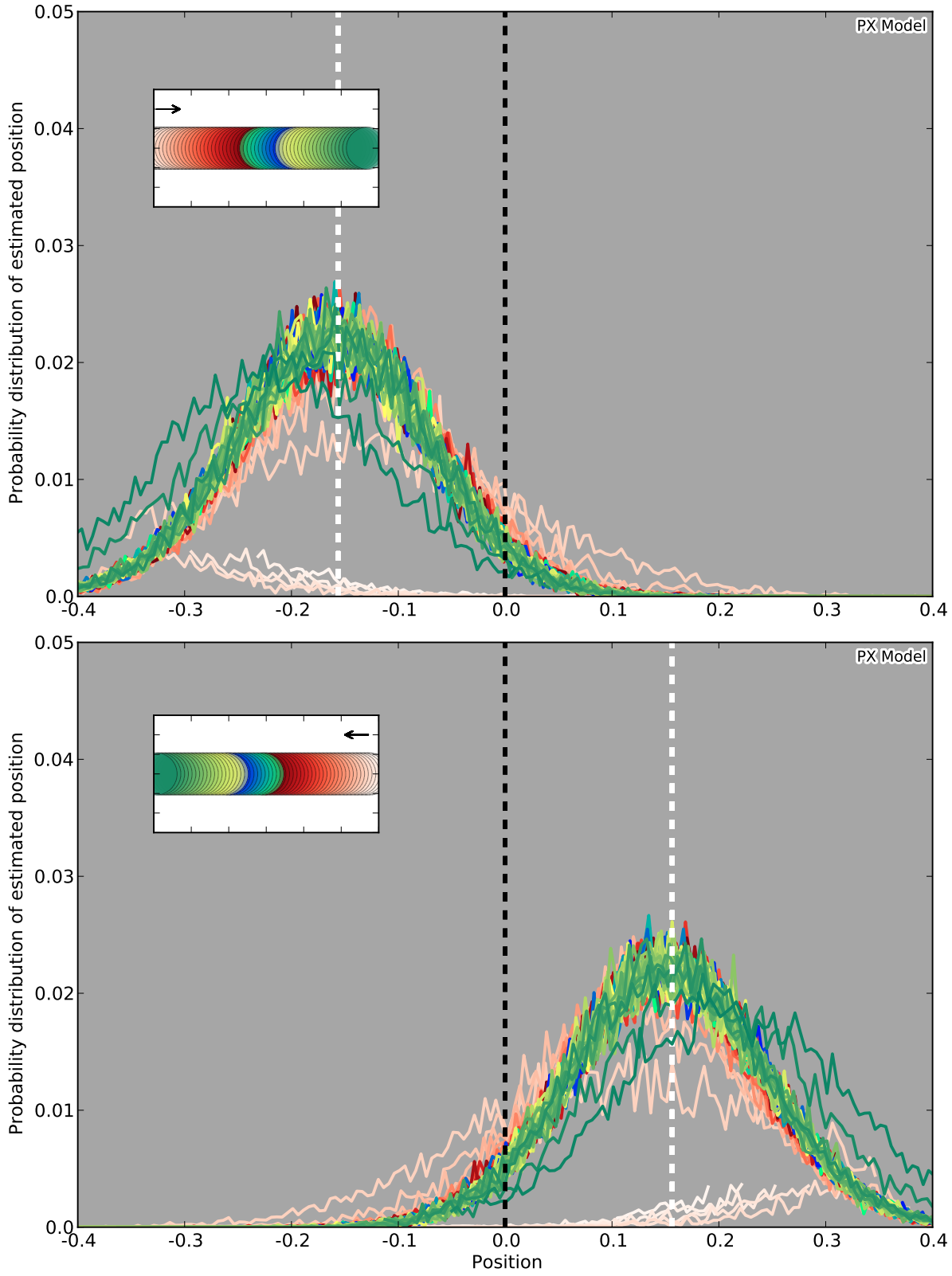

Figure 5.12: **Spatial distribution of estimated positions in the PX model with** $\tau = 100\ ms$: spatial distribution of estimated positions as stimulus moves along the trajectory shown in inset. Time steps of motion estimation are matched to the color spectrum from red to green. All traces are subtracted from the actual position of stimulus, thus black dashed line indicates the center of stimulus and white dashed line corresponds to delayed position of stimulus. **Top)**: rightward motion **Bottom)**: leftward motion

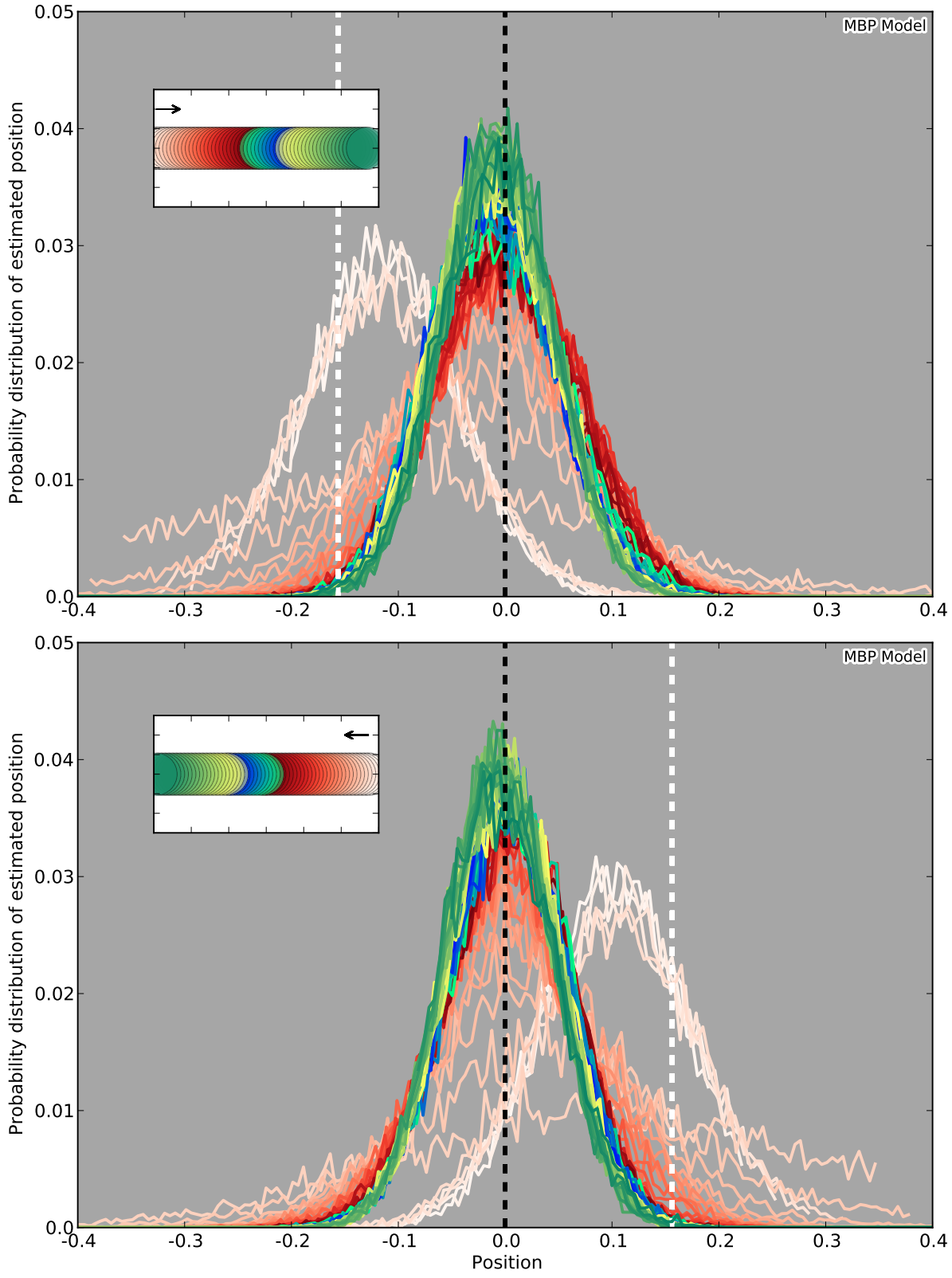

Figure 5.13: **Spatial distribution of estimated positions in the MBP model with** $\tau = 100$ ms: spatial distribution of estimated positions as stimulus moves along the trajectory shown in inset. Time steps of motion estimation are matched to the color spectrum from red to green. All traces are subtracted from the actual position of stimulus, thus black dashed line indicates the center of stimulus and white dashed line corresponds to delayed position of stimulus. **Top)**: rightward motion **Bottom)**: leftward motion

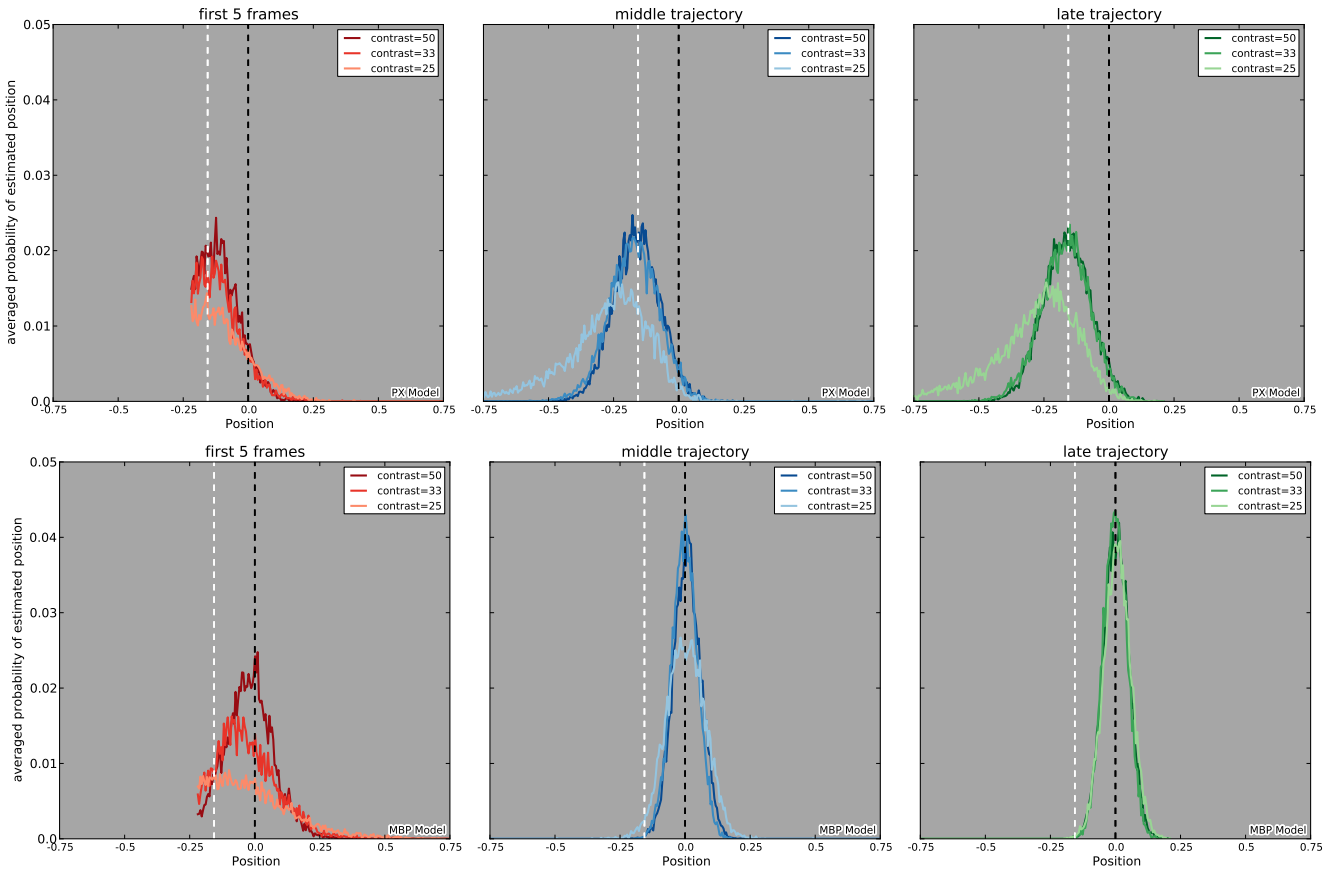

Figure 5.14: **Estimated positions at early, middle and late parts of the trajectory**, and at three different contrast levels. Stimulus is a dot moving in rightward direction and each panel corresponds to averaged response over corresponding temporal areas. Contrast levels are color coded and black and white dashed lines respectively indicate the actual and delayed positions of the stimulus. **Top)** PX model, **Bottom)** MBP model

5.5 Conclusion and Discussion

In this chapter, we have introduced the diagonal MBP model, as an effective mechanism to compensate positional error caused by neural delays. The questions that we have addressed are progressively fit in three following categories, each corresponding to one experimental setup on the MBP and PX diagonal models. First we have simulated delayed arrival of visual input and studied the role of delays in position coding of smoothly moving objects. As a specific case, we have simulated the anticipatory response of neural populations in V1 to a simple straight trajectory and its dependence on the length of the trajectory. Second, we have extended the same experiments to explore how predictive position coding may be changed at different parts of trajectory and with respect to a range of delay values. In the third part we have explored how delay compensation phase might be affected by contrast.

Herein, we will conclude and discuss the implications of our results.

Predictive extrapolation of response to delayed visual input In the first part of results we have simulated the anticipatory signature of the population response in V1. Our results highlight the importance of a large scale and trajectory dependent constraint serving as delay compensation mechanism. That may be implemented as diffusion of belief about appearance of stimulus in trajectory positions distant from retinal or cortical position of stimulus. The Diagonal MBP model that we have proposed for trajectory processing is a potential candidate to modulate spatiotemporal distribution of neural code for anticipatory objectives.

To this aim, we have examined the dynamical development of position coding at three successive positions of the trajectory, which are laid at different distances from the beginning point. This would be the counterpart to the experimental setup in which stimulus enters to the target CRF with three different trajectory lengths. These results are in qualitative accordance with extracellular responses recorded from populations in V1 (Benvenuti et al., 2011). In the next step, to study temporal profile of estimated response at different parts of the trajectory, we have imposed different delay values for both MBP and PX models. In the MBP model, knowledge about the velocity of the stimulus, allows to extrapolate the position code from the position provided by delayed information. However, as shown in Figure 5.11 there is a delay compensation phase at the earliest part of the trajectory.

According to Figure 5.2, motion is defined by its state vector $z = (x, y, u, v)$ and at $t = \tau$, first information from trajectory (corresponding to the position and velocity of stimulus at time = 0) arrives to the system. Then $z(\tau)$, as the first position estimation, is an extrapolated response from delayed information. For very early positions in the trajectory, as the estimated velocity is not accurate enough, the extrapolated position estimations are still temporally lagging behind the center of the stimulus. This delay is gradually corrected by improvement of velocity estimation (note the shift of peak position in early trajectories of Figure 5.11 and Figure 5.9). The correction of positional error by MBP model is also evident in Figure 5.13.

Furthermore, the dependence of temporal profile of response with respect to the length of

trajectory, as experimentally observed in V1 populations, is seen in early trajectory of Figure 5.11. Our results highlight two distinct temporal phases in the processing of smooth trajectories: an early phase and a late phase. In the early phase, the delay is compensated by spatial extrapolation of delayed sensory information. Then for the rest of trajectory, extrapolated position estimations are matched to the actual position of the stimulus. In the early phase, the temporal distribution of estimated position is sensitive to the distance of actual position of stimulus from the start point of the trajectory. In late phase, tracking has already emerged at its maximum probability and position estimations are transferred in space with a saturated shape.

The trajectory dependent anticipatory signature of estimated motion is in accordance with the experimental results from (Guo et al., 2007) and (Benvenuti et al., 2011). On the other hand, the most commonly accepted mechanism for anticipatory response of neurons in the early visual cortex is diffusive role of horizontal connections (Angelucci et al., 2002; Stettler et al., 2002). However, these connections have a limited spatial distribution and conduction speeds. For motion trajectories longer than a threshold, an increase in the length of trajectory would not change the modulatory effect of lateral connections on population response. In MBP model, this effect can be explained by saturated profile of position estimations in the late trajectory.

***Delay compensation and Fröhlich effect*:** Delay compensation in the early trajectory and misestimation of earliest positions is consistent with Fröhlich effect (Fröhlich, 1923). With Motion-based position coding, perceptual miss of earliest positions in the trajectory may be explained by initial delay time, as there is no sensory information before time = τ. Extrapolated estimations in MBP model begin at time= τ, in which system starts to compensate the effect of delay.

As we reviewed in chapter 1, there are various experimental evidences from the retina, primary visual cortex and MT+ area supporting motion-based position coding of moving objects. Our MBP model provides a generic abstract framework implementing predictive extrapolation of smooth trajectories, without any specification about neural properties of one region in the visual path. Indeed, considering prediction as a large scale mechanism, provides a powerful framework to simulate crude and abstract population responses and to study a range of questions. In this context prediction is able to modulate spatiotemporal properties of neural responses. Internal model of trajectory motion, implemented by motion-based and systematic extrapolation of responses to delayed inputs results in a coherent representation of motion, even for the cases of stimuli with short transient absence (as results in chapter 4).

It is worth to mention that if in theory, the internal MBP model is able to overcome neural transfer delays, in reality this delay is just significantly reduced and probably never reaches to zero. Though, there is no way to make sure that if our perception is completely matched to the actual motion or just it is an efficient enough approximation of it.

Considering the retinotopic map, a stimulus moving in a coherent path (a straight line) triggers successive neural hills of activity in neurons whose receptive fields are centered on the

trajectory. In this chapter, we have explored how the motion of the stimulus may impose a large scale constraint on the modulation and shift of the neural code in the cortical space. Therefore, we have addressed questions on correspondence between physical position of the stimulus and probabilistic significance of the neural code by populations specialized for that specific position. Our framework, by anisotropic diffusion of the probabilistic confidence to following positions, may explain the spread of neural activity in cortical areas, based on the relative distance between each area and the cortical position of the stimulus.

We have studied the effect of contrast in characteristics of position code in early, middle and late trajectories in PX and MBP models with $\tau = 100\ ms$. The MBP model suggests that higher contrast would lead to quicker delay compensation in early trajectory (see the peaks' order in Figure 5.14), which is consistent with results of Berry et al. (1999) on dependence of motion extrapolation to contrast in retinal cells. Although this dependence only holds in the early trajectory and for late enough positions there is no effect of contrast on peak location of position code.

In all parts of study we have used PX model as a control to MBP model, to highlight the effect of motion-based position coding versus position-based models such as (Tlapale, Masson, and Kornprobst, 2010).

Summary

In this chapter, we have studied dependence of position coding of smoothly moving objects on their trajectory. In continuation to the MBP model described in chapter 3, we have introduced the diagonal model to simulate neural transfer delays. In the next chapter, we will use this model to study flash lag effect as a motion-induced position shift.

Chapter 6

Motion-induced position shifts as consequence of prior information on smooth trajectories

Abstract

Our perception from smooth motion is rooted in the continuous flow of sensory information, coming from systematic displacement of eyes to follow the moving object in the visual scene. Some of experimental studies on motion processing, provide various evidences to highlight the difference in position coding of moving objects versus stationary transient stimuli (See chapter 1). Evidences of motion induced-position shifts and the flash lag effect, as a prototypical example of this domain, suggest the facilitatory role of predictable motion in position coding of moving object and compensation of neural delays. This category of studies disentangles the contribution of motion and more precisely the velocity signal in coherency of predictive neural activity. In this chapter, we have used the diagonal model to simulate the flash lag effect and three known variants of it (standard FLE, flash initiation FLE and flash termination FLE). Our results highlight that motion-induced position shifts like FLE may be considered as a direct consequence of the facilitatory and extrapolative aspect of predictive position coding.

6.1 Problem statement: Motion-induced position shifts

As we studied in chapter 5, there may exist some compensatory mechanisms in the neural representation of motion. The flash lag effect (FLE), from category of motion-induced position shifts, is a well studied visual illusion linked with the concept of neural delays.

Motion-induced position shifts may be caused by facilitatory motion processing mechanisms. As reviewed in chapter 1, various behavioral and experimental studies on FLE suggest existence of a facilitatory and predictive mechanism in motion processing to compensate the neural delays by extrapolation of motion in smooth trajectories. Visual perception is supposed to stay unaltered, unless the neural system receives enough evidences about a considerable change in the world state. In other words, the internal representation of the world will be kept or updated

by *perception of changes* (MacKay, 1958). In processing of smooth motion trajectories, as there is no significant change in the state of the world (namely direction of motion), the internal presentation is likely to be progressive to code for the most expected positions of the stimulus. From this view, FLE can be regarded as a motion-induced position shift, caused by a position corrective mechanism for trajectory motion.

There are evidences on occurrence of a counterpart effect on color, luminance and spatial frequency. Similar to motion-induced position shifts, coherent changes in some other features of visual stimulus like color, luminance and spatial frequency can lead to a continuous and progressive change in the visual perception and cause observations similar to FLE (Sheth, Nijhawan, and Shimojo, 2000). These visual signatures highlight that *perception of smooth change* in the sensory data can be reflected in a coherent neural representation and facilitated neural processing, for instance by delay compensations mechanisms.

In chapter 5, we found that excluding velocity information from the motion estimation process deteriorates it significantly. Our findings in estimation of smooth and predictable motions can also explain the sensory origins of FLE: in accordance with the hypothesis of Nijhawan and Wu (2009) and experimental studies of FLE (Berry et al., 1999; Jancke et al., 2004), we suggest that predictive extrapolation of trajectory motion via a simple internal model may be responsible to the delay compensation and FLE.

This chapter is prepared in the following order: first we have a brief introduction to FLE and its various explanations, and then section 6.3 includes our results in modeling of FLE. Finally in the section 6.4, we will conclude our findings and explain advantages and limitations of our approach in comparison with the studies reviewed in chapter1.

6.2 Flash lag effect (FLE) and different theories

As we reviewed in chapter 1, studies on FLE suggest that positional coding of objects is likely to be affected by some contextual information like contrast or being in motion or static. Thus, for a stationary and a moving stimulus, having the same retinal position at the same time will result in different perceived positions. That is likely to be induced by motion signal. The FLE was first discovered by MacKay (1958) and after that it did not attract much research attention until 1994, when Nijhawan started to study similar questions (Nijhawan, 1994). In an experimental setup which assured perfect alignment of both moving and stationary stimuli at the same time, this study reported that most subjects perceive the moving object to lead in space. Nijhawan (1994) proposed motion extrapolation as a hypothesis to explain FLE. According to this hypothesis, the visual system is predictive and takes advantage of trajectory information to be prepared for upcoming flow of moving stimulus and also to be able to correct positional errors caused by neural delays.

In the last two decades, this approach has been criticized by various challenging experimental setups by Nijhawan (Nijhawan, 2002; Nijhawan and Wu, 2009; Nijhawan et al., 2004) and others (Eagleman and Sejnowski, 2000; Krekelberg et al., 2000; Schlag et al., 2000; Whitney

and Murakami, 1998; Whitney, Murakami, and Cavanagh, 2000) (See chapter 1 for a complete review).

The main flaw of motion extrapolation theory is that it can not account for some particular experimental observations on FLE, for instance when the flash appears at the end of the trajectory of a moving object. In this case, motion extrapolation predicts a perceptual overshoot same as standard FLE, while there is no report of such an effect in the conducted experiments. In general, motion extrapolation acts as an inertial constraint in detection of sudden changes in the status of motion, like motion termination or motion reversal. Therefore it can not explain the relevant experimental data.

One alternative hypothesis for FLE is based on different processing latencies of the moving and stationary objects (Jancke et al., 2004; Purushothaman et al., 1998; Whitney and Murakami, 1998; Whitney, Murakami, and Cavanagh, 2000). This account assumes that the visual system processes the moving object with more efficient mechanisms and this causes the percept of lag in FLE. The improved processing of moving objects may have multiple origins but there is no evidence to support any significant difference in processing latency of moving and stationary objects. Also, sensory information of both stimuli are supposed to be transferred via the same physical path and therefore there can not be a big difference in transfer and processing time. Instead, the efficiency in processing of moving stimuli can be rooted in the primed state of cortical areas for predictable stimulus.

Some other studies have explored FLE from the aspect of position persistence for the flashing object (Krekelberg, 2001; Krekelberg and Lappe, 2000). According to this hypothesis, motion information are slowly averaged within a 500 ms window, therefore, the perception at the position of flash is persisted while the averaged position for moving object will be ahead. The main flaw of this account is that, regarding the dynamics of motion integration $500ms$ is unrealistically long.

Finally, one other hypothesis that have been in discussion as a potential mechanism underlying FLE is postdiction (Eagleman and Sejnowski, 2000). This account assumes that visual system collects information not only from the past of a visual event but also from its future. In other words, the neural response associated with position of a moving stimulus at a certain time is modulated by the position of stimulus during a temporal window centered by that time. Postdiction theory can explain why FLE does not happen in flash terminated cycle but is not successful in accounting for the standard FLE. In postdiction model, the flash is regarded as a reset for motion integration, thus observers are expected to perceive a change in the speed of moving object as well as the spatial lag of the flash. If we consider both moving and flashing stimuli to have the same neural delay as Δt, then flash occurs at $t = t_0$ and is observed at $t = t_0 + \Delta t$. During this time, the sensed position of moving object is moved between $x = \Delta s$ (behind the position of flash) to $x = \sigma s$ (where σs is measured value of the flash-lag effect). Thus, to cause the effect, the constant speed of moving object must change from $V = \frac{\Delta s}{\Delta t}$ to $\acute{V} = \frac{\Delta s + \sigma s}{\Delta t}$. Humans are good enough in the perception of speed increments, but in the case of

FLE such a perception of speed increment was never reported by any subject (Nijhawan, 2002).

Our theory to explain FLE is consistent with the motion extrapolation account. Results presented in the next section illustrate our simulations on FLE and its various experimental aspects by diagonal models (See chapter 5), and in particular by stressing on the role of velocity-based information.

6.3 Results

This section includes our results on modeling FLE with motion-based prediction (MBP) and position-based prediction (PX) models. In all results described in this chapter, we have studied estimated position for a horizontally moving dot, presented during 128 frames. All results have been averaged over 10 trials and estimated positions are calculated as histograms over the whole range of trajectory, where the range is divided into 400 positional bins.

6.3.1 Motion based prediction and the flash lag effect (FLE)

Our results in chapter 5 highlight the crucial role of motion-based position coding, as an effective delay compensation mechanism for trajectory motion. In the same context, one could provide an explanation for potential sensory origins of FLE as a well studied visual illusion. Our modeling approach emphasizes different manipulation of stationary and trajectory motion by the sensory system, and explores advantages of motion-based position coding in a simple and generic framework.

Our experimental procedure to simulate FLE in the model can be explained by Fig 6.1. Similar to frequently used experimental setup in the previous studies, we have simulated standard and half cycle flash lag demonstrations: standard FLE experiment is composed of a moving stimulus and a stationary flash that appears in perfect alignment with it in the middle of the trajectory. In the case of half cycle FLEs, the flash appears either in the beginning or at the end of trajectory (Nijhawan, 2002). As shown in the Fig 6.1, we have done six experiments in the model: three described FLEs for rightward and leftward motion.

Results of rightward motion experiment on MBP model are illustrated in Fig 6.2. It includes estimated positions of flash and moving stimulus, at standard FLE setup as well as flash initiation and flash termination half cycles of FLE.

For rightward motion, the dot starts from $x = -0.8$ (the position of initial flash) and moves toward $x = 0.8$ and disappears in this position (the position of termination flash). For leftward motion, the dot starts from $x = 0.8$ (the position of initial flash) and moves toward $x = -0.8$ and disappears in this position (the position of termination flash). In all experiments, flash lasts for %4 of trajectory length (5 frames out of 128) and estimated responses are histograms with 400 bins over the range of $(-1, 1)$. Each row in Fig 6.2 illustrates estimated positions for the flash or moving stimulus, for the 5 frames that the flash lasts. As it is evident in the figures, during the time that response to the stationary flash is developed, estimated position for the

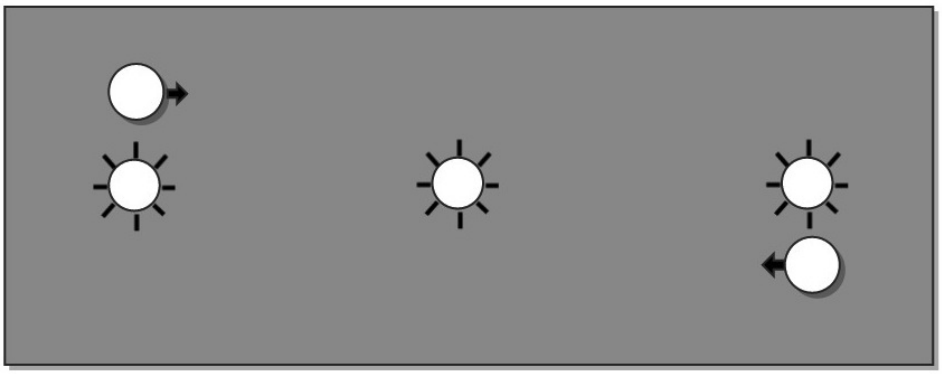

Figure 6.1: **Experimental setup of FLE model**: in six different experiments, each of two moving dots (in rightward and leftward directions) are aligned with each of stationary flashing dots. Based on the position of flash experiments are called flash initiated FLE, standard FLE and flash terminated FLE.

moving object is shifted in the direction of motion. For flash initiation half cycle (red plot), where flash appeared in the beginning of trajectory of moving dot, during flashing frames the confidence of estimation is increased and it is centered at the center of the flashing dot. For the moving object at the beginning of its trajectory, during the same time frames, confidence of estimated positions is growing along with a smooth shift of peak in the direction of motion. The similar effect is seen for middle and late trajectories (blue and green plots). Except that, for these positions, estimations have already reached to their maximum confidence.

Figs 6.3-6.6 summarize results of standard FLE from spatial and temporal aspects:

FLE in position: In Figs 6.3 and 6.4 estimated positions for stationary flashing stimulus, and stimuli moving in rightward and leftward directions are illustrated. Responses are obtained from diagonal models, respectively with $\tau = 0\ ms$ and $\tau = 100\ ms$, and averaged over frames of flashing period. In the MBP model, just after the end of flash, position estimations of moving stimuli are ahead of flash position. Estimated positions by PX model, as it is expected from previous results, are even slightly behind the center of stimulus.

Spatial lead of moving object is more evident in Fig 6.4 with $\tau = 100\ ms$. This figure shows that, by the time that response to flash reaches to its maximum ($t = \tau$ + end frame of flash), the response to the moving stimuli is considerably shifted in the direction of motion. Indeed, as the moving object is in the middle of its trajectory, in the response of diagonal MBP model the positional error has been already corrected. Therefore, at the starting frame of flash, the position of moving object aligned with flash is correctly estimated. While the response to the flash starts to develop at ($t = \tau$ + start frame of flash) and grows until ($t = \tau$ + end frame of flash). During this time, estimated positions of moving object moves forward. This result of diagonal MBP model highlights the importance of the motion signal and velocity of motion in differentiation of two stimuli appearing in the same location at the same time.

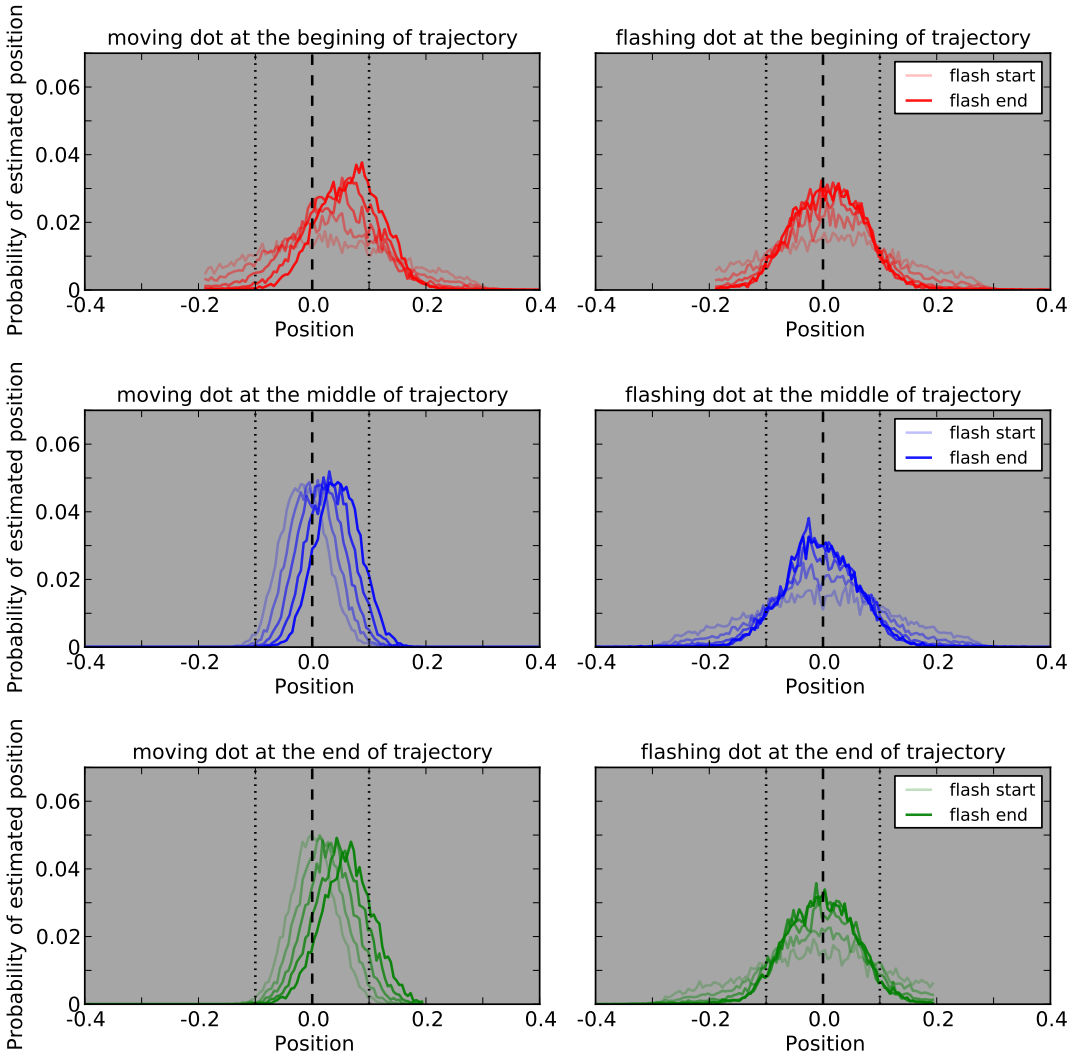

Figure 6.2: **FLE in MBP model** ($\tau = 0\ ms$)**: Top)** flash initiated FLE, **Middle)** standard FLE and **Bottom)** flash terminated FLE in rightward motion. Each row includes estimated position for flash and moving stimulus at five frames of flash. Frames from start to the end of flash are plotted with light to dark colors. Dashed and dotted lines respectively match to the center and radius of flashing dot.

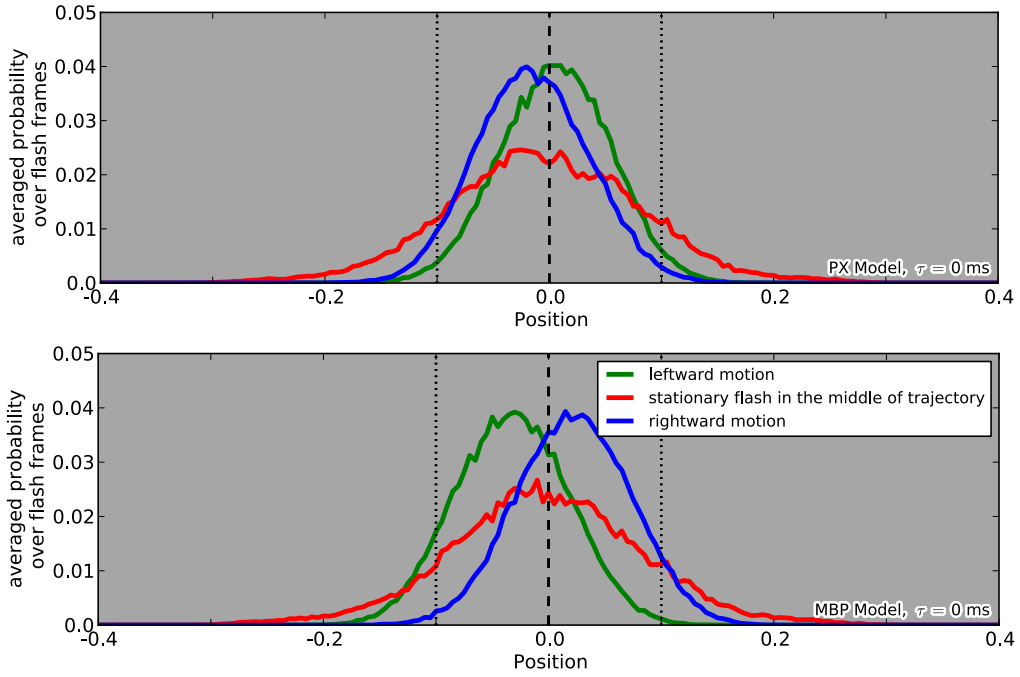

Figure 6.3: **Spatial response in PX and MBP models of standard FLE** with $\tau = 0$ ms, for rightward and leftward motions: each trace illustrates spatial distribution of estimated position averaged over flashing frames. Dashed and dotted lines respectively match to the center and radius of flashing dot and each curve is average of estimated positions over five flashed frames.

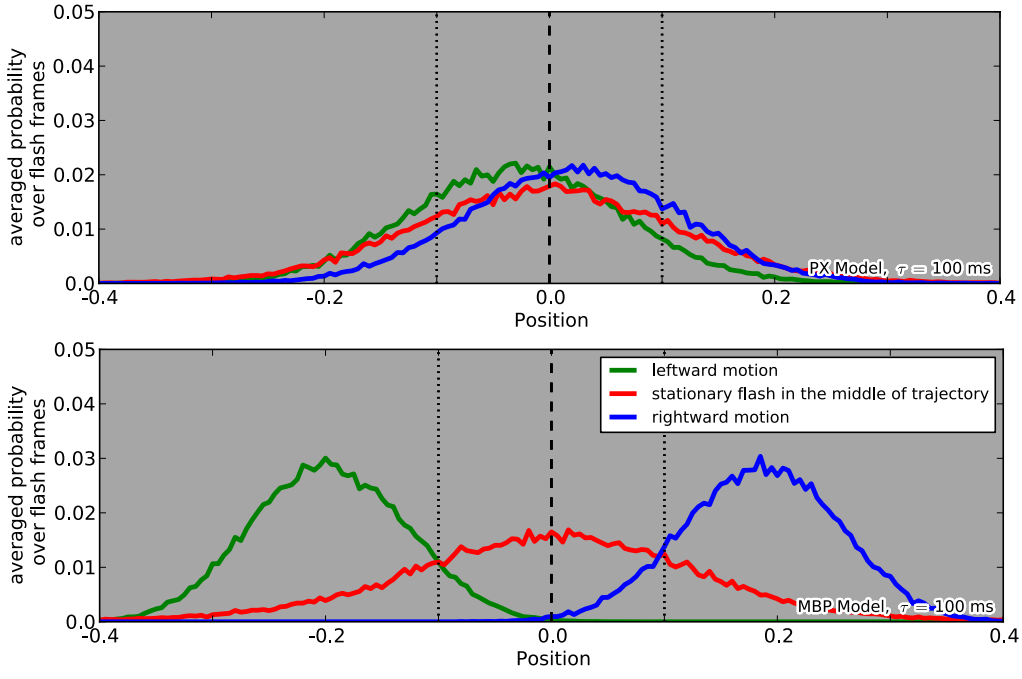

Figure 6.4: **Spatial response in PX and MBP models of standard FLE** with $\tau = 100$ ms: each trace illustrates spatial distribution of estimated position averaged over temporal range (flash start + τ, flash end + τ. Dashed and dotted lines respectively match to the center and radius of flashing dot and each curve is the average of estimated positions over five frames of the flash.

FLE in time: In Figs 6.5-6.6, the temporal dynamics of the estimated response in the location of the flash are illustrated. The temporal profiles are averaged over all positional bins covering the location of the flash. For the moving object, the responses are bell shaped, as the confidence hill is supposed to move to the next positions in the trajectory. Where, responses to the flash are damping with a slow dynamics, depending on the strength of σ_p, prior on slow speeds (See Equation 5.24). This inertia effect can be considered as position persistence of flashing stimulus. In Figure 6.5, for MBP model with $\tau = 0\ ms$, at time corresponding to the beginning of flash (left dotted line), estimation confidence of flashing object starts to rise, while response to the moving object is already at its maximum confidence.

Standard FLE versus flash initiated half cycle: In Figs 6.7 and 6.8 we have compared MBP models of standard FLE and flash initiated FLE for $\tau = 0\ ms$ and $\tau = 100\ ms$. The spatial lead in both FLEs is in the same range, but the effect can be interpreted to be stronger in the standard FLE, as the estimation confidence of moving object in the middle of trajectory is higher than beginning.

Figs 6.9-6.10 illustrate the temporal profile of flash initiation FLE by MBP models with $\tau = 0\ ms$ and $\tau = 100\ ms$. In the case with $\tau = 0\ ms$, after beginning of flash, the positional response for both stimuli starts to grow similarly. Then, in the position of the flash, response of moving stimuli drops and estimation confidence moves toward next positions in the trajectory, while the confidence of estimation for flashing stimulus damps slowly. Imposing $\tau = 100\ ms$ in the MBP model of flash initiated FLE (Fig 6.10) leads to different temporal profile: at time $t = \tau$ after beginning of flash, the positional response for both stimuli is the same, but response of moving stimulus drops quicker. As diagonal MBP model of moving object aims to correct positional error at the earliest part of trajectory, the estimation confidence is quickly shifted from the position of the flash toward the positions supported by the velocity signal.

Positional inertia of stationary flash: Fig 6.11 illustrates the dependence of temporal dynamics of flash response to σ_p (preference to slow speeds). According to the equation 5.24, imposing a stronger prior information of slowness of motion will result in a more temporally resistant response for flashing stimulus. Comparing this sluggish damp in temporal response of flash with the quick transfer of response to the moving stimulus reveals a position persistence effect similar to what has been discussed in some studies (Krekelberg, 2001; Krekelberg et al., 2000).

Dependence of FLE with respect to the speed: In Fig 6.12 dependence of spatial lead of moving stimulus to the speed has been shown. In the reference speed ($v = 1$) the lead is very small and then it is linearly increased with speed.

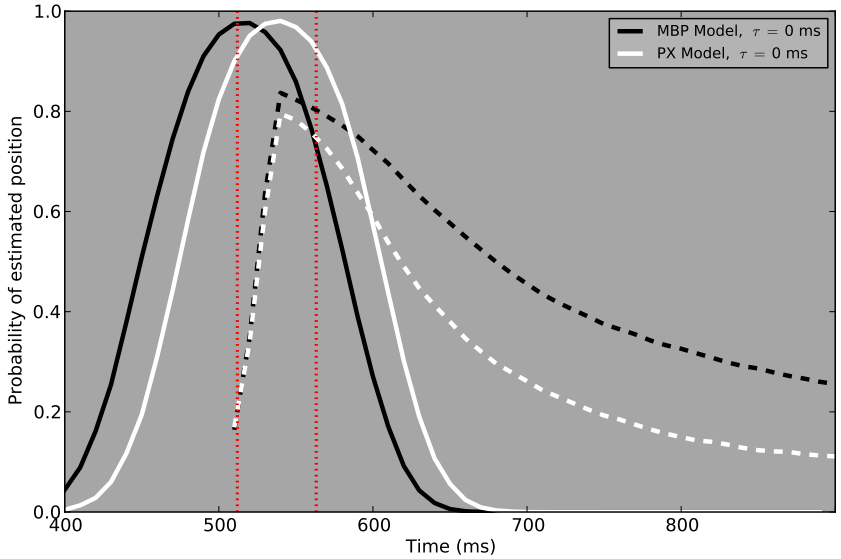

Figure 6.5: **Temporal response in MBP and PX models of standard FLE with** $\tau = 0$ *ms* : each trace illustrates temporal distribution of estimated position, summed over position bins corresponding to the location of the flash. Solid and dashed lines respectively correspond to moving and flashing stimuli. Duration of flash has been marked with red dotted lines.

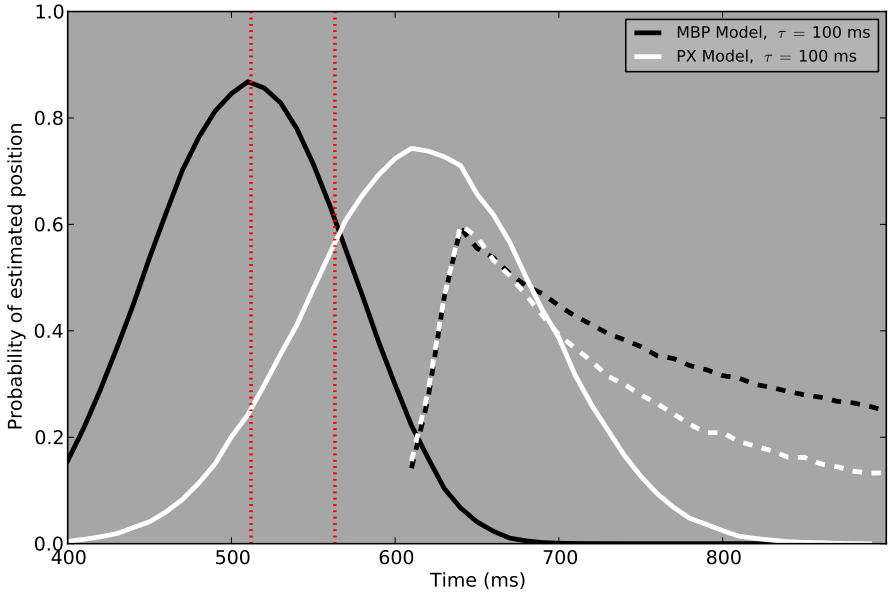

Figure 6.6: **Temporal response in MBP and PX models of standard FLE with** $\tau =$ 100 *ms*: each trace illustrates temporal distribution of estimated position, summed over the position bins corresponding to the location of flash. Solid and dashed lines respectively correspond to moving and flashing stimuli. Duration of flash has been marked with red dotted lines. As the flash happens in the middle of trajectory, positional error caused by delay is already corrected, and response of MBP model for moving object at the position of flash is accurate.

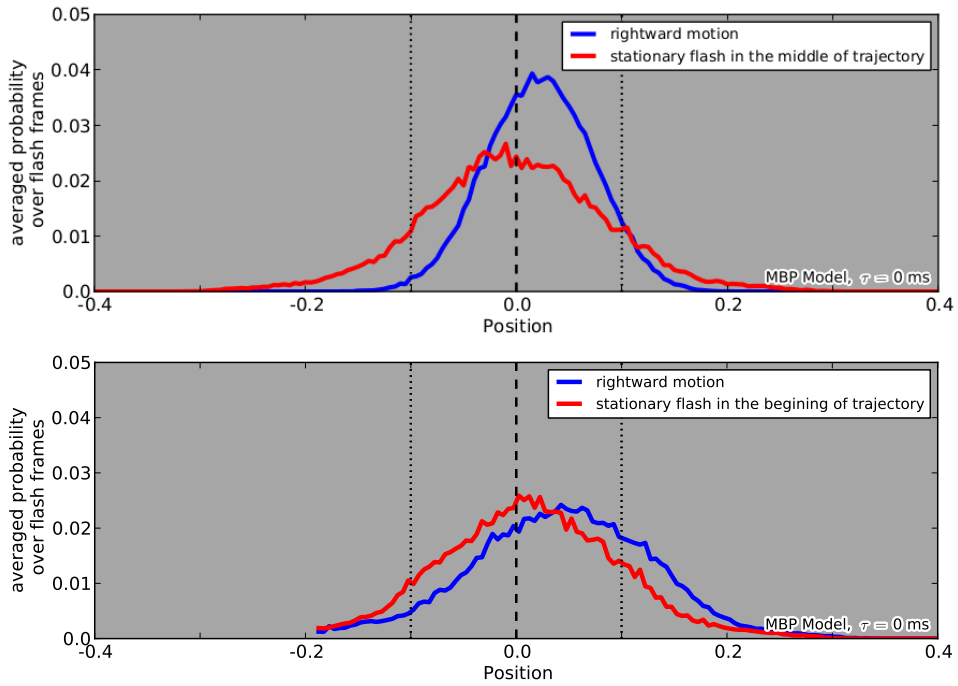

Figure 6.7: **Comparison of standard FLE and flash initiated FLE:** MBP model with $\tau = 0\ ms$ **(Top):** standard FLE **(Bottom):** flash initiated FLE. Each trace illustrates spatial distribution of estimated position averaged over flashing frames. Dashed and dotted lines respectively match to the center and radius of flashing dot.

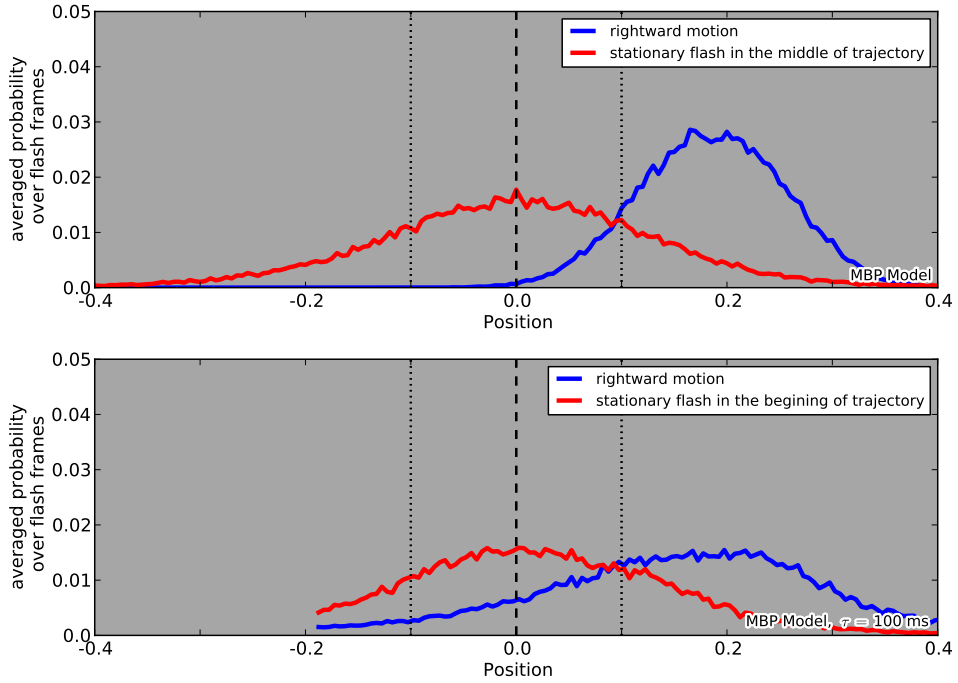

Figure 6.8: **Comparison of standard FLE and flash initiated FLE:** MBP model with $\tau = 100\ ms$, each trace illustrates spatial distribution of estimated positions averaged over temporal range (flash start + τ, flash end + τ), dotted and dashed lines correspond to the radius and center of dot stimulus. **(Top)** standard FLE, **(Bottom)** flash initiated FLE.

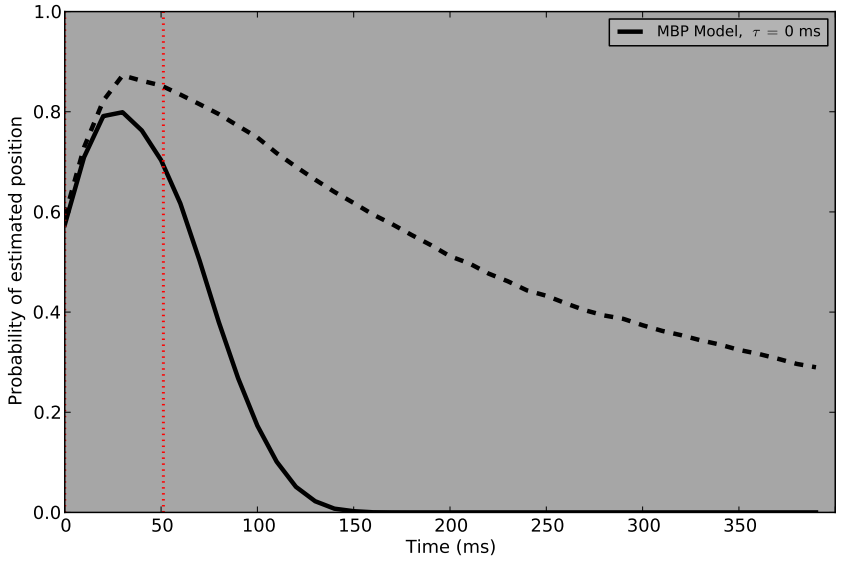

Figure 6.9: **MBP model of flash initiated FLE with** $\tau = 0$ ms: each trace illustrates temporal distribution of estimated position averaged over the position bins corresponding to the location of flash, solid and dashed lines respectively correspond to moving and flashing stimuli. Duration of flash has been marked with red dotted lines.

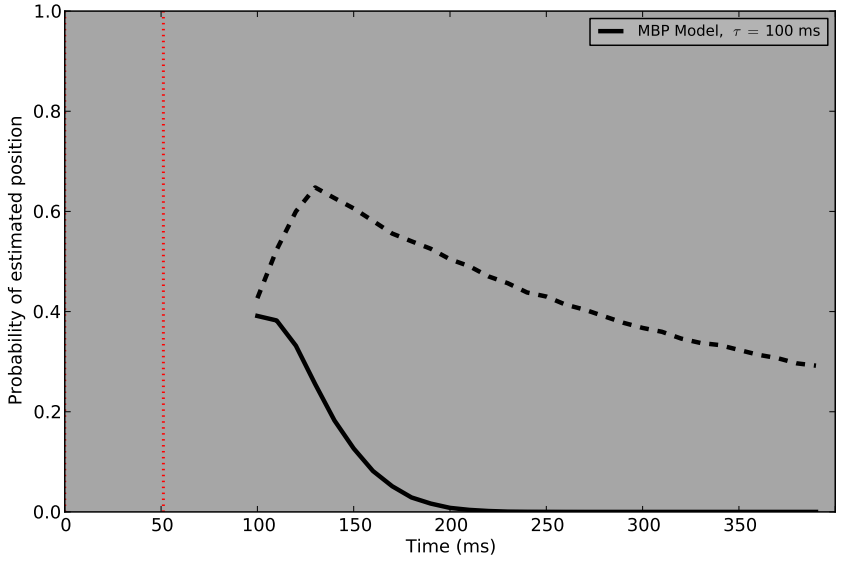

Figure 6.10: **MBP model of flash initiated FLE with** $\tau = 100$ ms: duration of flash has been marked with red dotted lines and each trace illustrates temporal distribution of estimated position, averaged over the position bins corresponding to the location of flash, solid and dashed lines respectively correspond to moving and flashing stimuli. At the time that response to flash starts to develop, the estimated response for the moving object at the starts to drop. It implies that in the position estimation of moving object, the confidence of estimation is pushed to the more forward positions, as it is evident in Fig 6.8

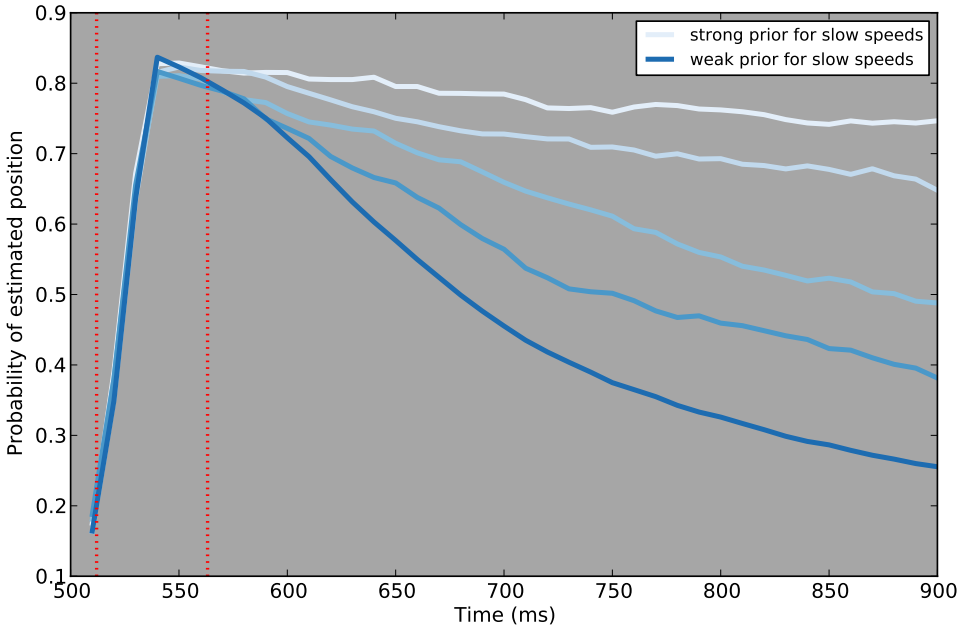

Figure 6.11: **Dynamics of estimated position for flash stimulus**: each trace shows estimation confidence on the position matched to center of flash. After the end of flash, the estimation confidence drops with a dynamics dependent on the prior preference for slow speeds.

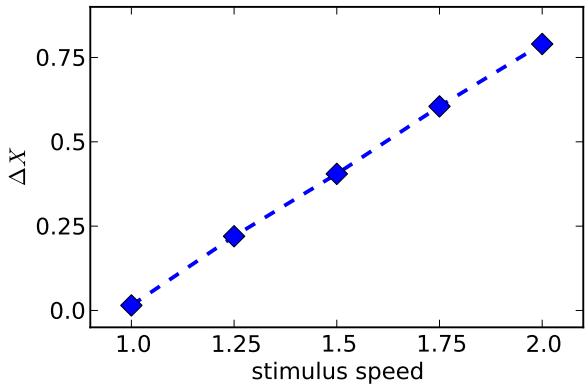

Figure 6.12: **Dependence of positional lead of moving stimulus to speed**: for the reference speed the lead is nearly zero and it increases with increase in the speed

Dependence of FLE with respect to the contrast of flashing stimulus: As we mentioned earlier, decreasing the contrast of moving stimulus will result in the later emergence of tracking state (See Fig 5.14). Therefore, decreasing the contrast of both stimuli in the case of standard FLE is equivalent to decreasing only the contrast of flashing stimulus. We have explored dependence of FLE to the contrast of flash and the results are summarized in Figs 6.13, 6.14 and 6.15.

Fig 6.13 illustrates estimated response for a stationary flash stimulus at three different contrast levels and for flash duration%4. The spatial profile of responses has been averaged over flashing frames. Evidently, the peak of response depends to the contrast of stimulus.

Fig 6.14 includes the peaks of flash response in MBP model, in a grid of different contrasts and durations for the flash. These are peaks of positional response (similar to the peaks in the Fig 6.13). If we consider a certain value as detectability threshold of flash (marked for each contrast), the grid shows that in lowest contrast threshold is reached for a longer flash and causes the bigger positional lead.

Fig 6.15 illustrate development of estimation confidence in the position of flash, at 3 conditions marked in Fig 6.14. If we consider a detectability threshold for the response of the flash (0.7, for example), in lower contrast we need to keep the flash for longer duration to achieve the threshold.

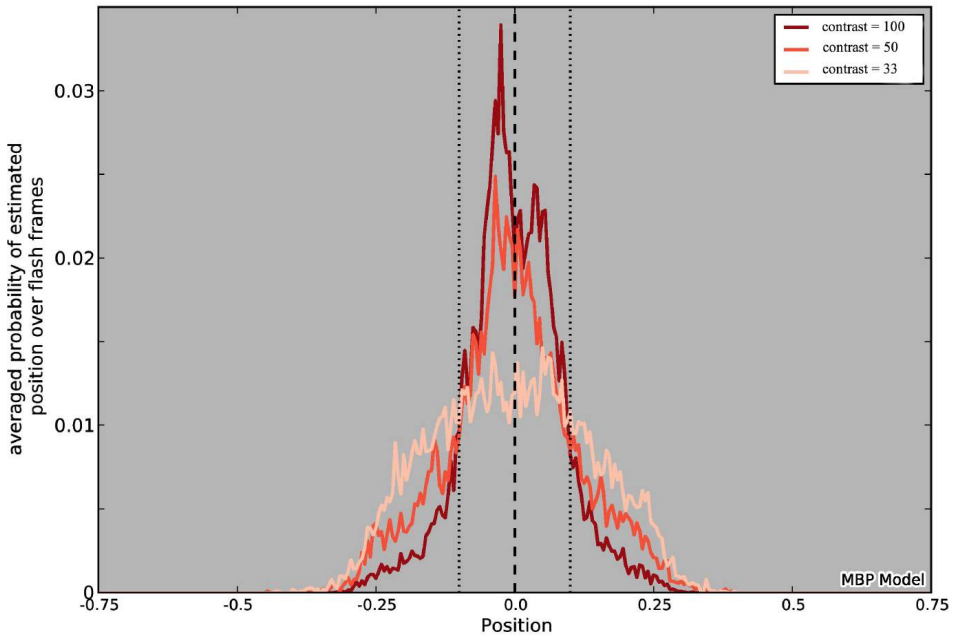

Figure 6.13: **MBP model, spatial distribution of estimated response to a stationary flash dot**: response has been averaged over flashing frames and dotted lines indicate the size of stimulus with a center shown by dashed line.

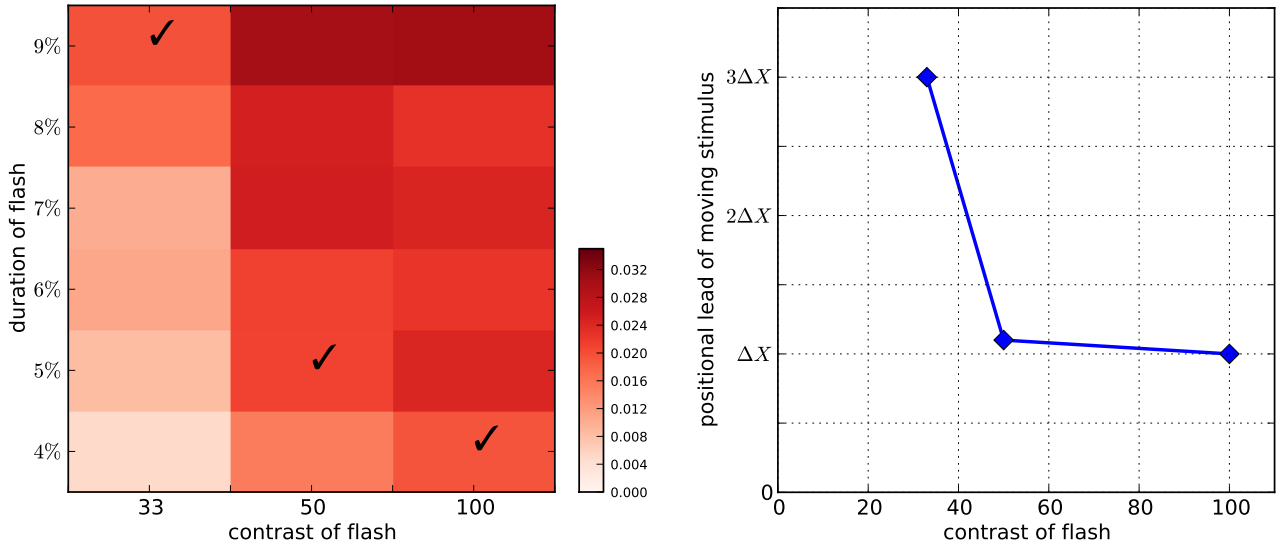

Figure 6.14: **MBP model, positional lead of moving stimulus at different contrast**: **Left)** color level at each cell of grid represents the peak of model response, at the position matched to the center of flash and averaged over the flashing frames (integral of response at each condition is 1). Defining a peak threshold for detectability of flash reveals that in low contrast conditions the threshold is achieved for longer durations of the flash (compare marked cells). **Right)** positional leads of moving stimulus for 3 marked conditions of the grid

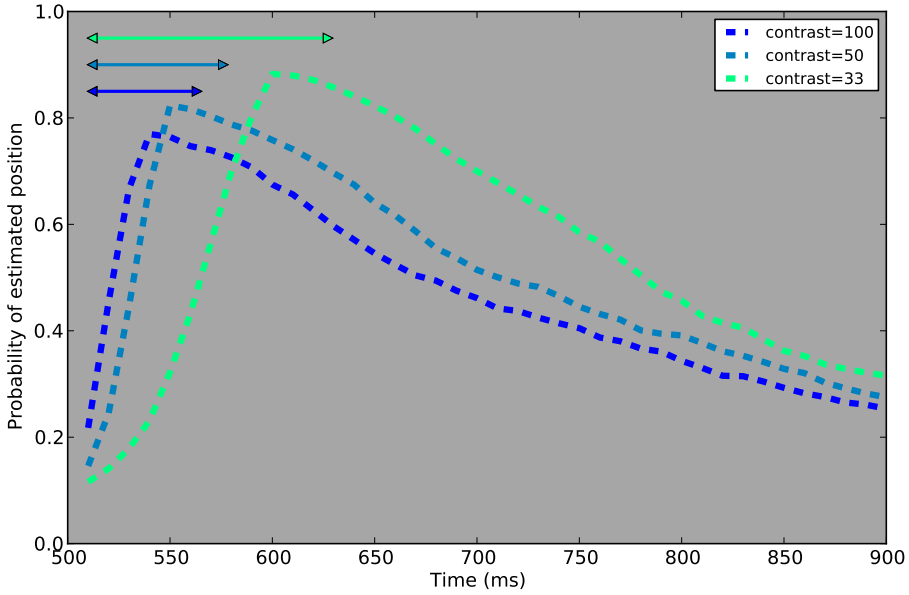

Figure 6.15: **Response of MBP model to flashing stimulus at 3 different contrasts**: each trace illustrates development of confidence at positional bins matched to the position of flash. Defining a threshold for detectability of flash (for instance 0.7) highlights the later detectability for the stimulus with lower contrast.

6.4 Conclusion and Discussion

In this chapter, we have studied the role of prediction in spatiotemporal explanation of motion-induced position shifts. As a well known framework we have modeled flash lag effect (FLE) in three experimental setup including: standard, flash-initiated and flash-terminated. Experimental and modeling studies on FLE are closely related to investigations on delay compensation and predictive position coding mechanisms. Based on our results on motion based processing of trajectory motion (See chapter 5), we have modeled FLE from the aspect of predictability of trajectory.

As reviewed in chapter 1, there are few main accounts for explanation of FLE: motion extrapolation, latency difference account which assumes existence of different latencies for flashing and moving objects, temporal averaging and postdiction mechanism. Each of these accounts have their own strengths and limitations. Our approach is in the direction of motion extrapolation account which was first proposed by Nijhawan (1994) and for about two decades has been under debate and discussions. Our model takes advantage of motion-based position coding to differentiate moving and stationary flashing objects by velocity of motion signal. The MBP model is a generic motion estimation algorithm that by using position and velocity information of trajectory, gives insights about the position and quality of neural response to flashing and moving stimuli.

Our results on FLE suggest how motion properties (mainly velocity) of stimulus may shift neural code in the corresponding population of neurons. For instance, according to our results in delay compensation, knowledge about the velocity of the stimulus may effectively shift the neural code from the populations that just received delayed input, toward the populations that are predicted to be stimulated at time $t + \tau$. In simulation of FLE on the MBP model we have imposed a short flash (lasting for %4 of trajectory length) at early, middle and end of trajectory. These cases would be matched to the three main demonstration of FLE in previous studies known as flash initiated FLE, standard FLE and flash terminated FLE. MBP model explains the spatial lead of a moving stimulus observed in FLE by shift in the positional response of moving object in the direction of motion.

The summary in Figs 6.3 and 6.6 reveals the difference of MBP and PX model (using or not using velocity signal in motion estimation) in accounting for the spatial lead of moving object. Estimations by PX model, are always slightly behind the position of stimulus, even for the case with no imposed delay ($\tau = 0$).

All FLE experiments are simulated under two circumstances for each of MBP or PX models: $\tau = 0\ ms$ and $\tau = 100\ ms$. As shown in Figs 6.2 and 6.3, the spatial lead in position code of moving stimulus is evident in the case with $\tau = 0\ ms$. Although according to Figure 6.4 this spatial shift is increased for the case with $\tau = 100\ ms$. In the temporal domain, at the start time of flash and in the positional bins matched to the position of flash, estimation confidence by MBP model is at the maximum level, where the response to flash starts to develop (See Figure 6.5). Imposing a delay as $\tau = 100\ ms$, as illustrated Figure 6.6, reveals that in

standard FLE and in the middle of trajectory, estimated position of moving object is matched to the actual position of stimulus, where response to the flashing stimulus starts to rise with delay.

Experimental evidences of FLE: Our approach in modeling FLE is consistent with experimental evidence of Berry et al. (1999) about motion extrapolation in retina. Their experiment on ganglion populations of retina provides insights to neural origins of FLE at the retinal level. They have reported that, during development of the response to flashing stimulus, response to moving object is shifted in the direction of motion. We have reproduced their experimental data with an abstract model without any neural specification. In addition, another experimental study (Jancke et al., 2004) provides evidence about the similar shift in the neural populations of cat visual cortex. We suggest that, predictive advancement of neural code towards the neural populations which their receptive field is slightly ahead of current position of stimulus, can be a general processing rule in the visual cortex.

Flash initiated cycle of FLE: Indeed, one may expect a smaller spatial lead for moving object in flash initiation cycle, as in earliest frames of moving object position estimation by MBP model is still lagged behind the actual position of stimulus. Considering the dynamics of trajectory processing with MBP model, this is not the case anymore and few frames of flash duration is enough for position code of moving object to compensate the delay and match to the actual position of stimulus. From this perspective, the MBP model does not predict any significant difference in the amount of spatial lead simulated in flash-initiated, standard and flash-terminated cycles of FLE. On the other hand, obviously direct relation of spatial lead with the length of flash is suggested by our model. The results that we have shown in Figs 6.7- 6.8 may also be interpreted in another way: the standard FLE can be considered to be stronger than flash initiated FLE, as estimation confidence is higher for moving object in the late enough parts of trajectory (far from start point).

FLE and dependence on contrast: Another important aspect in experimental studies of FLE is dependence of spatial lead on the contrast. In fact, as MBP model works based on belief accumulation from trajectory and without any specific contrast gain control mechanism, having low contrast would only result in later emergence of tracking state (saturated profile shape). From this perspective, MBP model highlights two points in contrast dependence of FLE: 1- at flash initiation cycle, higher contrast would result in bigger spatial lead of moving object. 2- at standard and flash terminated cycle (late enough positions in trajectory of moving object), dependence of spatial lead on contrast can be explained by contrast of flashing stimulus.

In Figs 6.14 and 6.15, we have summarized dependence of FLE on the contrast of flashing stimulus. Consistent with the relevant experimental evidences (Kanai, Sheth, and Shimojo, 2004), model response to a flashing stimulus with low contrast will take longer to reach detectability threshold. Thus, moving stimulus will advance further in the trajectory and cause a bigger

positional lead.

Shortcomings of motion extrapolation theory: In the literature, two main shortages of motion extrapolation account in full explanation of FLE are raised: first, it was found experimentally that the spatial lead of moving object in flash initiation cycle is comparable with the spatial lead observed in standard FLE setup. While one excepts that, at very early trajectory two objects should not be differentiable enough. Our results on existence of the same spatial lead amount in flash initiation cycle answers this apparent contradiction: duration of flash is enough for moving object to take advantage of motion signal in its trajectory and to correct the positional error caused by neural delays.

Second shortage of motion extrapolation is in accounting for flash termination cycle: experimentally, no spatial lead or perceptual overshoot is observed for flashes at the end of trajectory, while motion extrapolation hypothesis would predict such an effect. Indeed, this is a limitation of our MBP model too: the internal model of motion develops an inertia after gaining confidence about the trajectory, possibly filling a blank in the trajectory (Khoei, Masson, and Perrinet, 2013). Thus, it is less sensitive to changes of the stimulus status, such as a sudden stop. Motion extrapolation is a successful explanation for maintenance of coherent motion perception (See chapter 3), but meanwhile can produce unrealistically long trajectory estimation, after disappearance or any status change of moving object. This shortage can be observed in very slow drop of estimated position of flash in Figs 6.5- 6.6.

As argued by Nijhawan (2002), we also agree that predictive motion extrapolation is a powerful hypothesis for motion estimation of smoothly moving objects, but as soon as having a significant status change in the stimulus some other supplementary mechanisms may come to play and cease extrapolation of the trajectory. For instance, modifications of the model to integrate a hierarchical generative model including a higher level representation of the end of a trajectory (or of an occluder) would allow to integrate both cases (Adams, Perrinet, and Friston, 2012) and is compatible with an implementation including axonal delays such as (Perrinet, Adams, and Friston, 2014).

It is worth to mention that if in theory, the internal MBP model is able to overcome neural transfer delays, in reality this delay is just significantly reduced and probably never reaches to zero. From this perspective, one can imagine that we perceive the moving object ahead of a flashing one because trajectory motion supports reduction of neural transfer delays for moving object, and not for stationary one. Though, there is no way to make sure that if our perception is completely matched to the actual motion or just it is an efficient enough approximation of it. Thus, not having perceptual overshoot in the end of trajectory is justified: with a more realistic motion extrapolation, when a moving stimulus stops, in a very short while our perception catches it in the real position without any spatial lead. This approach matches the findings of Jancke et al. (2004): moving (predictive) objects are indeed processed quicker than flashing (unpredictive) objects but still with a fixed, axonal delay.

Part IV

Conclusion and Discussion

Chapter 7

Conclusion and discussion

Scope of the study

In this thesis, we have investigated visual signatures emerging from motion-based position coding of moving objects. We have used a Bayesian motion estimation framework (implemented by a particle filtering method) with a central hypothesis on temporal coherency of motion and we have explored the complementary role of velocity based information in efficient coding of position.

Motion-based position coding theory proposes a predictive coding mechanism, based on an internal model of motion and smooth trajectory. As it is generally believed, in retinotopic visual areas there is a systematic correspondence between stimulated receptive fields in the physical world and activation of neural populations. Therefore, the location of neural activity and specifications of activated populations is predictable. Here, using a motion estimation framework, we have explained this predictability by mapping the motion trajectory to the flow of neural activity. To this aim we have estimated the position of neural activity by a motion-based prediction model (MBP, predictive in position and velocity). Two other configurations have been used mainly as control models, to highlight the difference in accuracy of estimations by excluding one of motion components: position-based prediction model (PX, predictive only in position) and velocity-based prediction model (PV, predictive only in velocity).

In this chapter we will summarize our theoretical findings and will discuss implications and limitations of our results in the context of recent literature.

7.1 Modulation of position coding by temporal coherency of motion

According to the classical knowledge about retinotopy and the existence of hyper-columns (Hubel and Wiesel, 1968), motion is mapped into a systematic diffusion of neural activity in the visual cortex. The neural activity codes the instantaneous position and velocity of moving objects, by a dynamic and smooth scan of corresponding neural populations. Basic anatomical properties of

visual areas along with a hypothesis on temporal coherency of motion, impose some constraints on relevant theoretical and modeling studies. Temporal coherency of motion proposes a model for investigating neural mechanisms behind the processing of short and long term motion. It has been successful in explaining some visual signatures and psychophysical evidences, including motion cooperativity (Yuille and Grzywacz, 1989), motion outliers (Watamaniuk, McKee, and Grzywacz, 1995b), motion inertia and occlusion (Burgi, Yuille, and Grzywacz, 2000; Khoei, Masson, and Perrinet, 2013) and aperture problem (Perrinet and Masson, 2012; Yuille and Grzywacz, 1989).

Studies with a hypothesis on temporal coherency of motion (Grzywacz, Watamaniuk, and McKee, 1995; Watamaniuk, McKee, and Grzywacz, 1995b; Yuille and Grzywacz, 1989) have considered motion estimation as an optimization problem. As we reviewed in chapter 2, in these frameworks velocity is estimated by a smoothing constraint on speed and direction domains, which is compatible with neural structure of visual areas and their optimal function. Mathematical description of this hypothesis bridges between physical origin of sensory information and specifications of neural codes, which can be implemented into neural networks simulating specific region of the visual system (Burgi, Yuille, and Grzywacz, 2000).

The MBP model used in this study, implements a type of temporal coherency of motion and provides a framework to explore fine grained maps of estimated position and velocity. In addition it highlights the contribution of the velocity signal in position coding. Our results show how excluding velocity information from motion estimation may lead to impaired or delayed position coding.

7.1.1 Motion extrapolation in blanked or occluded trajectories

In chapter 4 we have challenged the robustness of estimated motion in conditions with low sensory confidence, like having interruption in the trajectory or high levels of background noise. We have used motion extrapolation theory (Nijhawan, 1994) to explain the coherency of estimated motion, its dynamics during transient absence of the stimulus and motion catch-up after reappearance. As we reviewed in chapter 1 and chapter 4, motion extrapolation and motion inertia (Anstis and Ramachandran, 1987) are similar theories in line with motion coherency. Indeed, extrapolation of moving objects and inertia of estimated motion may be built-in visual mechanisms, to respect motion coherency as a natural regularity.

In this part of the study, we have proposed that motion extrapolation can be supported by an internal model of trajectory and we have investigated the efficiency of MBP and PX models in holding the efficient internal representation. Also we have used the PV model and according to our results, the MBP model supports quicker catch-up of motion after transient absence of the stimulus.

In all model experiments, we have used a Gaussian dot as stimulus. To study the specific role of each motion component, we have kept the stimulus simple and controlled. In fact, one may suppose that a Bayesian motion estimator would be efficient enough in dealing with simple and

easy-to-track stimuli, even by having a weak and incomplete internal model for motion. This theoretical expectation arises from an assumption on strong effect of sensory data (Bayesian likelihoods) in driving motion estimations. Our findings in chapter 4 and (Khoei, Masson, and Perrinet, 2013) answers to this family of questions. According to our results, the difference between PV and MBP models is slight in simple tracking of stimulus (though estimation by MBP model is more precise), where PX model provides a poor estimation of position and velocity. This suggests that velocity information is a fundamental ingredient for precise position coding.

7.1.2 Velocity dependent representation of invisible motion

To study the extrapolation of the trajectory during a short transient absence of the stimulus, we have compared the behavior of the three above mentioned models (MBP, PX and PV). The implication of the MBP model results is compatible with experimental findings on motion extrapolation, while it is not the case for PX and PV models. Estimated position and velocity by MBP model during the blanking period of stimulus is kept largely coherent, in the sense that estimated position is transported in the direction of motion. For longer durations of blank estimated position and velocity start to diffuse slowly, resulting in a decrease in estimation confidence.

According to experimental evidences, neural populations in the area MST keep their firing activity during short absence of the stimulus (Newsome, Wurtz, and Komatsu, 1988). Such a sustained activity is also reported in the parietal posterior cortex of monkeys (Assad and Maunsell, 1995) during transient occlusions of the target, in OSR (omitted stimulus response) in the retina of the tiger salamander (Schwartz and Berry, 2008), and in flicker electroretinogram (ERG) of human cone system (McAnany and Alexander, 2009).

Some other experimental findings in different visual areas report that neural activities associated with tracking of occluded stimuli are very similar to tracking of visible stimuli. Makin, Poliakoff, and El-Deredy (2009) have analyzed event related potentials (ERP) of human subjects, for two experiments including tracking of visible and occluded stimuli. They report a high overlap between results of two experiments and suggest that stored velocity information may be responsible for the coherency of the recorded response during occlusion. Their findings is consistent with some earlier fMRI studies of visual areas during occluded motion. Olson et al. (2004) have investigated the neuronal representation of occluded objects in the human brain and have suggested that occluded motion is treated similarly to the real motion and is processed by similar cortical areas. More precisely, their fMRI recordings show strong activation of IPS (Interparietal Sulcus) during motion occlusion and less significant activity in human homologous of areas MT/MST. These findings highlight the particular role of IPS in keeping neural representation of inferred or expected (not seen) motion, which is potentially part of a cortical circuit for planning of future actions.

It also has been shown by fMRI studies that MT/MST areas are responsive to static images of objects with implied motion, and this recorded activity is significantly less for images without

motion implications (Kourtzi and Kanwisher, 2000; Senior et al., 2000). Neural activity recorded from area FEF (Frontal Eye Field) also implies a speed-dependent update in the internal representation of a trajectory (Barborica and Ferrera, 2003). In this study, monkeys were trained to make saccades to the estimated positions of invisible stimuli and recorded activity from FEF seemed to be modulated by the speed of target (before disappearance).

Our modeling results in MBP configuration provides a generic account for representation of invisible motion, independent from anatomical and neural specifications of each area and only by taking advantage of velocity-related information. The knowledge of velocity and the way that it contributes in keeping internal representation of trajectory can be widely varied in different stages of the motion processing stream. For example, in lower level areas like V1 velocity information may be fed back from higher areas like MT, to stabilize the internal model of motion.

7.1.3 Tracking of invisible motion

To simulate eye velocities and to be able to compare our results with behavioral data, we have used velocity readouts from a simple oculomotor plant. Regarding behavioral studies on tracking of an interrupted trajectory by human and monkey subjects (Becker and Fuchs, 1985; Bennett and Barnes, 2003; Bogadhi et al., 2011; Eckmiller and Mackeben, 1978), smooth pursuit is maintained during a blanking period, though eye velocities have been reported to be decreased in first 200 ms after disappearance of the stimulus and stabilized in a level around $40\% - 60\%$ of target velocity.

Our modeling results are not directly comparable with these data. According to our results, estimated velocity during blank either can remain at the same level as it was before blank, or it can drop with a slope. The slope of velocity decrease will be determined by the strength of prior information on slowness of motion. Weak prior information (a Gaussian in velocity domain, centered by zero and with high variance) leads to a plateau in velocity during blank, while a strong prior (a Gaussian in velocity domain, centered by zero and with low variance) results in a decreasing estimated velocity in the absence of stimulus.

A rather strong prior may explain a reasonable drop of estimated velocity consistent with experimental evidence, but that will also affect the catch-up of velocity right after appearance of the stimulus. In general, imposing a very strong prior on slowness of motion will push estimated velocities toward zero, in a way that, despite strong sensory measurements, the model will fail to track the stimulus.

Our theoretical investigations suggest that, in the absence of stimulus and sensory evidence, prior assumption on the slowness of motion may come into play and cause a gradual drop in the estimated velocity for invisible stimulus. Where by reappearance of the stimulus sensory information plays the main role and determines the estimated speed. This interpretation is compatible with the approach of Montagnini et al. (2007), where they have proposed a recurrent Bayesian framework with updating priors.

7.1.4 Complementary role of position information in robust tracking at low contrasts

We have studied dependence of estimated velocity with respect to the contrast in PV and MBP models. At very high contrasts both models provide an accurate estimation of velocity but their behavior differs as soon as background noise increases. The MBP model shows a more robust response with respect to the noise: increasing noise up to a rather high threshold only delays emergence of tracking and crossing that level will cause no-tracking state. This binary response in velocity estimation is not observed in the PV model. It seems that excluding position information from internal model of motion and relying only on velocity deteriorates the estimated velocity proportional to the noise level.

To combine two sources of uncertainty, we have simulated interrupted trajectories in a range of contrasts. Our results in this part follows a binary pattern similar to contrast-tracking figures. Results of MBP model suggests a robust and quick catch-up of stimulus after reappearance for a wide range of contrasts. While estimated velocity by PV model at corresponding contrast levels is deteriorated with a slower catch-up dynamics.

To assess the role of the slowness prior on the quality and robustness of motion tracking in low contrast conditions, we have used different internal models of trajectory (PX, PV and MBP). In these cases, histograms of estimated position for stimulus are better criteria than averaged velocity estimations. In summary, according to our results, a strong prior on slowness of motion ($\sigma_p = 10$) will render a poor tracking even in high contrast conditions in the PV model. Estimated positions by PX model in high contrast starts to develop, but gradually lags behind the target positions and ultimately stops. In lower contrast the stopped position happens earlier in the trajectory. The MBP model provides with a more accurate estimation, in the sense that estimated position stops very late in the trajectory, and only at very low contrast levels.

These findings altogether complement the whole implication of our hypothesis: a motion-based predictive coding with strong prior on slowness of motions accounts for efficient and robust tracking of moving targets, even during their temporal absence.

7.2 A theoretical framework to investigate the effect of neural delays on position coding

In chapter 5, we have simulated neural transfer delays of motion information. In a modified Markov chain for state space of motion (so called diagonal model), we have investigated motion estimations based on motion measurements at time steps behind the current time, and with an internal model on smoothness of motion and delayed arrival of sensory information. Inspired by a diagonal motion extrapolation mechanism in ganglion cells of retina, proposed by Nijhawan (Nijhawan and Wu, 2009), we have implemented diagonal delay compensation as an abstract rule in motion estimation. The advantage of our model is that it provides a generic account for partial correction of errors in position coding, at any stage of motion

processing. Being independent from neural specifications of a particular visual area, our large scale motion estimation model emphasizes a strong influence of the velocity signal in modulation and correction of position coding, as it has been experimentally reported in various stages of motion processing machinery of visual system (Berry et al., 1999; Jancke et al., 2004; Maus, Fischer, and Whitney, 2013; Maus et al., 2010; Maus et al., 2013; Whitney et al., 2003).

As our simulations suggest, in diagonal MBP model, the positional error starts to decrease as velocity estimation gets more precise. Theoretically, in a steady state tracking with an accurate velocity estimation the positional error is zero. The Fröhlich effect (Fröhlich, 1923) refers to misperception of the earliest part of motion trajectory and it is believed to be linked with the problem of neural delays. In the diagonal MBP model a similar misestimation of position in the early trajectory is observed. For example, by imposing a delay as $\tau = 100\ ms$, the first estimation of position will start at $time = 100\ ms$ and will correspond to the position of stimulus at $time = 0$. However, after a few estimation steps and by having a more accurate estimation of velocity, extrapolation of trajectory during the delay period will catch up with the actual position of stimulus. We suggest that the Fröhlich effect is likely to be caused by the misestimation of velocity at the beginning of the trajectory. At one level this can be tested by behavioral experiments, for example by measuring the velocity of eyes and the value of positional error. In electrophysiological studies, this may be translated to an experiment to careful investigation of dynamics between development of positional code and velocity response.

In addition, the transfer delay of sensory information is a shared issue among all modalities and it is likely that the brain recruits predictive mechanisms to ease and accelerate perception and decision making. According to our investigations, motion-based predictive coding can be a way to compensate neural delays.

7.2.1 Anticipation of approaching stimulus

The diagonal MBP model, unlike PX configuration, is able to reproduce the experimental data on dependence of positional code with respect to the length of the trajectory. According to relevant experimental evidences (Benvenuti et al., 2011; Guo et al., 2007) response of neural populations in the early visual cortex is more distributed in time, for the stimulus that arrives to their receptive fields with a longer trajectory behind. Lateral connections in V1 are generally believed to contribute in propagation of sensory belief among neighbor and co-functional neural populations.

In simulations of the model, we have investigated an equivalent question: does the temporal distribution of estimated position depend on its distance from the starting point of the motion? In the other words, we have replicated this neural signature by considering neural delays and probabilistic accumulation of evidence about appearance of stimulus in a certain position in the trajectory. However, this signature is visible in our simulations only in the early part of the trajectory. Meaning that, in early points of the trajectory the internal model of motion is not fully stabilized and after some point (as soon as having correct velocity estimation) the temporal

distribution of estimated position keeps the same profile. This may be explained by the limited range of lateral connections: for trajectories longer than a threshold length, no sensitivity with respect to the length is expected.

As a main highlight, only the MBP model reproduced the trajectory dependent signature of position coding. Excluding velocity information (as our results in the PX model) leads to the same temporal shape for position code in all points of the trajectory. Also, the position estimated by PX model remains delayed, matching to delayed motion measurements. Difference between these two models stresses the potential interference of velocity code on development of the anticipatory position coding, as well as compensation for delays. One possible neural implementation of this theoretical insight can be modulatory effect of feedback from MT to V1. This can correct the positional error caused by delays and to further anticipate activation of populations whose receptive field are about to be stimulate, by a complementary knowledge on the coded velocity.

7.2.2 Flash lag effect as a basic visual signature

In chapter 5 we have proposed a model for flash lag effect based on the motion extrapolation account of Nijhawan (Nijhawan, 1994). Our approach stresses a facilitated processing of trajectory motion and highlights the difference between estimated position for an stationary, flashed and a smoothly moving stimulus.

In one set of simulations we have investigated the response of the diagonal MBP and PX models to standard FLE by imposing $\tau = 0$. First, we determined the shortest duration for flash which creates a significant probabilistic confidence for estimated position (about 4% of trajectory duration). Then we showed that for the MBP model, this duration is enough to shift the response of moving stimulus to a more forward position in the trajectory, matching to the instantaneous position of the moving stimulus. There is no such lead in estimated position of moving stimulus by PX model, implying that this model deals trajectory motion as a packed set of stationary flashes, without velocity and trajectory.

These results suggest that, without considering neural transfer delays, and having the same position for two stimuli at the same time, estimation of trajectory motion in the middle of its trajectory is facilitated by velocity information. Temporal dynamics of estimated position for moving stimulus illustrates that, in the MBP model the peak of the response matches the start time of the flash, when the flash response starts to develop. Just after the start of the flash, the response to the moving stimulus falls off, as estimation belief is shifted to more forward positions.

Imposing a nonzero delay to the diagonal model results in a bigger shift in the positional code of moving object (more delayed response for the flash). As we discussed earlier, the diagonal MBP model of moving stimulus corrects the positional error in the beginning of the trajectory and later on provides an accurate estimation of actual position of stimulus. Simulating FLE with a nonzero delay (τ) highlights that total latency of flash is composed

of τ plus the time needed for accumulation of belief about occurrence of flash in a certain position.

Inertia of flash response: As we mentioned earlier, taking advantage of velocity information, estimated position of moving stimulus is smoothly shifted in the direction of motion. This causes a quick displacement of probabilistic belief from the position co-aligned with the flash. Unlike the facilitated estimation of motion, the positional response of the model to a stationary flash is developed and vanished in a more sluggish manner. This main difference in the positional code of moving and flashing stimulus can potentially be a cause for the lagged perception of flash. Our results suggests that, after disappearance of the flash, the belief on occurrence of flash in a specific position starts to drop with a rather slow dynamics. Considering this effect, along with a rapid and smooth shift in the estimated position of moving stimulus, puts forward inertia of flash response as a potential cause of FLE: flash is perceived later than moving stimulus and the perception stays for a longer duration. The inertia of estimated position for flash may be interpreted as positional persistence (Krekelberg, 2001; Krekelberg and Lappe, 2000).

Our model explains this effect with a rather unrealistically slow dynamics. Indeed, in the absence of motion, as it is the case for the stationary flash, estimated position is highly shaped by prior information on slowness of motion. After disappearance of the flash, as there is no sensory evidence, the MBP model keeps predicting about having a slow motion (or no motion as a flash) and this prediction stays longer for a stronger prior on slow motions (lower variance around velocity = 0)

Dependence of FLE on velocity and contrast : According to various experiments, the value of positional lead of moving stimulus in FLE is linearly increased with velocity (Krekelberg and Lappe, 1999; Krekelberg and Lappe, 2000; Nijhawan, 1994). This dependence has been reproduced by the MBP model.

Our results on dependence of FLE to the contrast of flashing stimulus is consistent with experimental results of Kanai, Sheth, and Shimojo (2004). Indeed, as the MBP model works based on the accumulation of evidences about existence of an stimulus at a specific position, in low contrast conditions of moving stimulus only the earliest positions of the trajectory are misestimated. Thus, in the MBP model and for the case of standard FLE, the position code of a moving stimulus will be almost independent from the contrast. In this context, the positional lead of a moving object will only depend on the detection time of the flash. A flash with lower contrast will need longer duration to be detected and during this time the moving stimulus will advance more in the trajectory. This suggests that processing of an unpredictable stimulus is more affected by contextual effects like contrast.

7.2.3 Comparison with other models for FLE

Neural field model of Erlhagen: As we discussed earlier, the notion of internal representation of moving objects provides an explanation for the coherency of perceived motion despite random

interruption in the visual stream, and also for extrapolative and anticipatory aspects of motion processing. According to this theory, ultimate neural response is shaped by a combination of top-down feedback signal and bottom-up sensory information and the internal model of stimulus stored in higher visual areas supports a more reliable interpretation of visual input.

Our modeling study explores the functional consequences of having such an internal model, using position information, velocity information or both. As we reviewed in chapter 2, the neural field model of Erlhagen (2003) is the most relevant study on the emergence of FLE from interplay of internal model and stimulus stream. In this setup, composed of excitatory and inhibitory populations of position coding cells, extrapolation of the trajectory emerges from lateral interactions and network dynamics, in an appropriate parameter regime. Also, this model shows that priming of the position field is caused by accumulation of sub-threshold activities of excitatory populations.

Our model also highlights the role of the internal model of the trajectory, built from motion information. The main difference between our model and the model of Erlhagen, apart from having different structures, is that we have investigated the consequences of velocity information on shaping of position coding, where that model includes only position-tuned cells. On the other hand, results from both models stress the critical role of sub-threshold neural activities on creation of anticipatory or extrapolative signatures: in Erlhagen's model that is achieved in a certain parameter settings of neural field model and in our model it is raised by probabilistic accumulation of motion-based position estimation, in the predicted positions of the moving stimulus.

Neural network models of FLE: Few neural network models have addressed questions on delay compensation, motion extrapolation and FLE. For example the model of Baldo and Caticha (2005) demonstrates motion extrapolation and FLE, arising from a simple feedforward network of leaky integrate and fire neurons. Other studies (Lim and Choe, 2006; Lim and Choe, 2008) have discussed motion extrapolation at single neuron level and have explained FLE and delay compensation in a network with facilitating synapses, with sensitivity to the rate of change in various aspects of stimulus (for instance: direction, orientation, luminance).

All these works have investigated spatial priming of neurons via lateral and facilitatory connections, ignoring the facilitatory effect that may arise from velocity coding.

Postdiction model and motion extrapolation: Our approach is partially consistent with the theory behind postdiction model of FLE (Eagleman and Sejnowski, 2000). Though, taking a flash stimulus as a reset in motion integration process (as they proposed) is against our hypothesis based on probabilistic accumulation of trajectory information. The diagonal MBP model builds motion estimations at *time* t based on sensory informations corresponding to the past (*time* $t - \tau$) and proposes how neural delays may be compensated in visual processing. We emphasize the complementary role of velocity information in this compensation while postdiction theory proposes that coded position of stimulus is always pushed forward based on trajectory

information received in a short time interval after time t.

As we discussed in chapter 1, postdiction hypothesis accounts for experimental data in the flash terminated cycle, as well as cases in which there is an abrupt reversal in the trajectory. Nevertheless, postdiction theory in standard FLE predicts a velocity increment for moving stimulus which has not been experimentally observed by careful investigations (Nijhawan, 2002). In addition, according to postdiction, the Fröhlich effect is a consequence of a latency in position interpretation, as the earliest part of trajectory is missed because postdictive system needs information from the future.

Our approach and motion estimation theory is based on the motion extrapolation account proposed by Nijhawan (1994), even though both hypotheses do not oppose in the final outcome: motion extrapolation pushes the estimated position forward using trajectory information, and postdiction associates a position to the moving object which is influenced by future positions of it. As we reviewed (See sections 1.4 and 6.2), both accounts can be considered as a corrective mechanism against delays.

Our approach is in agreement with the theory of Krekelberg et al. (2000) on temporal integration. They have suggested that FLE is caused by motion of the moving stimulus after the flash, as different trajectories of two stimuli (trajectory versus no trajectory in the case of flash) leads in different temporal integration.

7.3 Perspectives and future work

Relational biology is a research field founded by Rosen (2012) to bridge between physical rules in the world and succession of neural activities in the neural system. Clear cause-effect relationships in the environment, known by causality, physics and other scientific rules, dramatically narrows down our expectations from future states of the world. Rational biology investigates that how evolutionary environment and causality in the physical world may be encoded into the organization of all living systems, or even impose a parallel causal entailment in neural responses.

Our study shares this point of view and fits into the same field of research: we have studied trajectory motion as a sequence of events in space and time and have challenged estimations of it with a knowledge of restrictions of the estimator system. Our study, by surveying motion-based prediction in the visual system, emphasizes that regularity and predictability of the sensory world may impose a large scale constraint of neural processing paradigms to obtain a more stable and reliable neural encoding, despite intrinsic restrictions and delays.

7.3.1 Predicting the present not predicting the future

Theoretical and experimental studies of neural delays are somewhat linked with hypotheses on the anticipatory nature of sensory encoding. According to the anticipation theory of Rosen (2012):

> ‘An anticipatory system is a natural system that contains an internal predictive

> model of itself and of its environment, which allows it to change state at an instant in accord with the models predictions pertaining to a later instant"

Application of this definition to biology and neuroscience implies that, the brain may permanently rely on its expectations from real time sensory input, to generate the instantaneous neural responses. To this aim a knowledge of itself is required, which is the outdated arrival of the sensory stream. Hence, if we take the time in the physical world as a reference, the internal representation of the external world is lagged behind it, although the response to the sensory input at *time t* may be created by probabilistic belief about future states of the stimulus.

The internal predictive model can provide the most expected states of the external world and based on this theory, it is likely that the brain keeps instantaneous encoding of the sensory world by a continuous update of probabilistic sensory representations. Our approach in diagonal model of motion-based prediction is consistent with the theory of Rosen. Instantaneous motion is estimated based on delayed arrival of sensory input and an internal model composed of knowledge on restrictions of the system (imposing a delay on sensory input) and knowledge on the regularities of the world (temporal coherency of motion). To obtain an instantaneous motion estimation the response to delayed sensory information is extrapolated during a virtual blank period (obviously without sensory measurements). Thus, the estimated motion is a response to expected state of the stimulus at the current time.

The theory of Rosen (2012) addresses a subtle point in the nature of biologic anticipation, and our study is also in accordance with that: the internal predictive model may have knowledge of its delayed access to the sensory world. By this knowledge, it pulls the expectations from future states of the input stream, to estimate the present state. Thus, the internal predictive model only catches up the present state of the external world and does not anticipate the future.

7.3.2 How might motion-based prediction be implemented?

Our study provides theoretical insights on the critical role of motion signal and velocity based sensory information in position coding of moving objects. We have investigated the complementary role of velocity information on correct position coding particularly by considering neural delays. According to various experimental findings (Berry et al., 1999; Jancke et al., 2004; Maus, Fischer, and Whitney, 2013; Maus et al., 2010; Maus et al., 2013; Nijhawan, 1994), the visual system distinguishes between moving and stationary stimuli which happen to be at the exact same location at the same time. This is reflected in displacement of position code for the moving target in direction of motion, so called motion extrapolation. Evidences of motion extrapolation suggest that neural populations engaged in position coding are likely to be modulated by velocity processing machinery, as it is observed by shifted neural activity in the corresponding direction or by dependence of positional lead in FLE to the speed.

Different modeling studies have reproduced motion induced position shifts mainly in neural networks composed of position tuned cells and via dynamical effect of lateral connectivities (Baldo and Caticha, 2005; Erlhagen, 2003; Lim and Choe, 2006; Lim and Choe, 2008) and provided

promising experimental predictions. Nevertheless, little research effort have been dedicated to the interaction of position coding and velocity coding cells. Our study, without being restricted to neural specifications of a particular area in the visual system, explores the consequences of including velocity information in position coding.

So far, motion extrapolation has been assumed to be caused by asymmetric structure of cells in retina (Nijhawan and Wu, 2009) or well known lateral connections in the early visual cortex. Our results highlights that, for instance, interaction between direction selective and position selective populations in retina or feedback from MT area to V1 can also play a critical role in distinguishing stationary and moving objects and generating motion induced position shifts.

7.3.3 Future work: neuro-mimetic motion computations

The current computational framework, implemented with particle filtering, has great capabilities in highlighting various aspects of motion processing in the visual system. Particle filtering and the CONDENSATION algorithm (Isard and Blake, 1998) are generalization of Kalman filter to deal with non-Gaussian perturbations in process and measurement models, such as multimodal probability distributions. Furthermore, a particle filter includes the most important samples of probabilistic distributions as an effective approximation method.

Samples of estimated motion by particle filter represent local motion detectors concentrated on the instantaneous position of moving stimulus, as an abstract indication of the stimulated receptive fields. These detectors move by motion of stimulus and have their own estimated velocity attached to them, with a certain relative weight. This way of representing local motion estimators allows careful investigations on emergence of a global motion in space and time, from a pool of local estimations. Also the resultant fine grained spatiotemporal map makes it straight forward to study the link between precision of position code and precision of estimated velocity (Khoei, Masson, and Perrinet, 2013).

Coherence constraint on estimated motion results in anticipatory responses in the next positions of the stimulus, as it might be diffused by anisotropic lateral connections in early visual areas. Also, continuos update in weight of local motion estimators based on sensory likelihoods may simulate the role of feedback from higher areas like MT, to stabilize the sensory processing in early areas.

Current implementation is also particularly advantageous to study dynamics of motion integration for stimuli with non asymmetric shape or multiple stimuli in the same scene. As investigated in an earlier work (Perrinet and Masson, 2012), the aperture problem as a classical question in motion integration is studied in the similar motion estimation framework, with an emergent solution arising from spatiotemporal dynamic of local motion estimators. This example highlights the capacities of the current computational framework to investigate the emergent neural signatures from local to global integration process, for instance different tracking of edges and surfaces of a bar stimulus.

7.3.4 Future work: motion-based prediction

This study is centered on the main theory of motion-based prediction. The theory suggests that, physical rules governing the world, might cause a parallel cause-effect entailment in the neural world.

A vast body of literature exists with focus on the structural organization of the visual system in terms of development, neural connections and learning toward optimal processing of the visual world. Our study highlights that, along with a well organized machinery, the visual system may benefit from an internal model for motion, composed of position and velocity informations, pooled from neural activities distributed over multiple areas.

We have elaborated this hypothesis in an abstract framework and also on a spiking neural network with anisotropic connectivities (Kaplan et al., 2014). Implementation of MBP and PX models in a spiking neural network and via different neural connectivity patterns also confirms the crucial role of velocity-based information in delay compensation and development of anticipatory responses. Further work in this direction is in progress to study the Fröhlich effect and FLE in spiking neural network.

Our study puts forward new experimental ideas to be conducted by different methods and in different visual areas. A recent study according to our hypothesis reports a significant reduction in FLE by TMS disruption of area MT+, where the same disruption on V1/V2 does not influence the perceived position (Maus et al., 2013). This study provides the first direct evidence for the involvement of MT+ in perceived position and leaves some other research questions to be further investigated. For instance, along with our theoretical predictions on the influence of velocity coding on position code, it would be promising to block the feedback connections from higher areas to MT (by lesion, pharmacological methods etc..) and to study the consequences on the flow of activity in the early visual cortex. Equivalently, the contribution of velocity information in motion anticipation in retina (Berry et al., 1999) can be investigated by deactivating the direction selective ganglion cells.

Bibliography

Adams, Rick A., Laurent U. Perrinet, and Karl Friston (Oct. 26, 2012). "Smooth Pursuit and Visual Occlusion: Active Inference and Oculomotor Control in Schizophrenia". In: *PLoS ONE* 7.10, e47502+ (cit. on p. 135).

Adelson, Edward H. and James R. Bergen (Feb. 1, 1985). "Spatiotemporal energy models for the perception of motion". In: *Journal of Optical Society of America, A.* 2.2, pp. 284–99 (cit. on pp. 29–32, 46, 52).

Alink, Arjen et al. (Feb. 24, 2010). "Stimulus Predictability Reduces Responses in Primary Visual Cortex". In: *The Journal of Neuroscience* 30.8, pp. 2960–2966 (cit. on pp. 19, 20).

Allman, J., J. Kaas, and R. Lane (July 1973). "The middle temporal visual area (MT) in the bushbaby,Galago senegalensis". In: *Brain Research* 57.1, pp. 197–202 (cit. on p. 23).

Angelucci, Alessandra et al. (2002). "Circuits for Local and Global Signal Integration in Primary Visual Cortex". In: *The Journal of Neuroscience* 22.19, pp. 8633–8646 (cit. on p. 117).

Anstis, Stuart and V. S. Ramachandran (Jan. 1987). "Visual inertia in apparent motion". In: *Vision Research* 27.5 (cit. on pp. 26, 138).

Assad, John A. and John H. R. Maunsell (Feb. 9, 1995). "Neuronal correlates of inferred motion in primate posterior parietal cortex". In: *Nature* 373.6514, pp. 518–521 (cit. on pp. 23, 24, 68, 70, 71, 85, 92, 139).

Baldo, Marcus and Nestor Caticha (2005). "Computational neurobiology of the flash-lag effect." In: *Vision research* 45.20, pp. 2620–2630 (cit. on pp. 46, 145, 147).

Barborica, Andrei and Vincent Ferrera (2003). "Estimating invisible target speed from neuronal activity in monkey frontal eye field". In: *Nat Neurosci* 6.1, pp. 66–74 (cit. on p. 140).

Barlow, HB and WR Levick (1965). "The mechanism of directionally selective units in rabbit's retina." In: *The Journal of physiology* 178.3, pp. 477–504 (cit. on p. 3).

Barnes, G. R. and P. T. Asselman (Aug. 1991). "The mechanism of prediction in human smooth pursuit eye movements." In: *The Journal of physiology* 439, pp. 439–461 (cit. on p. 24).

Barthélemy, Frédéric V. et al. (Feb. 2008). "Dynamics of distributed 1D and 2D motion representations for short-latency ocular following." In: *Vision research* 48.4, pp. 501–522 (cit. on p. 91).

Becker, W. and A. F. Fuchs (1985). "Prediction in the oculomotor system: smooth pursuit during transient disappearance of a visual target." In: *Experimental brain research* 57.3, pp. 562–575 (cit. on pp. 24, 25, 69, 92, 140).

Beers, Robert J. van, Daniel M. Wolpert, and Patrick Haggard (May 2002). "When Feeling Is More Important Than Seeing in Sensorimotor Adaptation". In: *Current Biology* 12.10, pp. 834–837 (cit. on p. 27).

Bennett, Simon J. and Graham. Barnes (Oct. 2003). "Human ocular pursuit during the transient disappearance of a visual target." In: *Journal of Neurophysiology* 90.4, pp. 2504–2520 (cit. on pp. 22, 26, 92, 140).

Benucci, Andrea, Robert Frazor, and Matteo Carandini (2007). "Standing Waves and Traveling Waves Distinguish Two Circuits in Visual Cortex". In: *Neuron* 55.1, pp. 103–117 (cit. on p. 21).

Benvenuti, Giacomo et al. (2011). "Building a directional anticipatory response along the motion trajectory Building a directional anticipatory response along the motion trajectory Building a directional anticipatory response along the motion trajectoryin monkey area V1". In: *Abstracts of Society For Neuroscience* (cit. on pp. 18, 96, 105, 106, 116, 117, 142).

Berry, Michael et al. (1999). "Anticipation of moving stimuli by the retina". In: *Nature* 398.6725, pp. 334–338 (cit. on pp. 11, 13, 23, 40, 43, 44, 118, 120, 134, 142, 147, 149).

Blasdel, Gary and Guy Salama (1986). "Voltage-sensitive dyes reveal a modular organization in monkey striate cortex". In: *Nature* 321.6070, pp. 579–585 (cit. on p. 5).

Bogadhi, Amarender R. et al. (Apr. 23, 2011). "Pursuing motion illusions: A realistic oculomotor framework for Bayesian inference". In: *Vision Research* 51.8, pp. 867–880 (cit. on pp. 25, 49, 140).

Bogadhi, Amarender, Anna Montagnini, and Guillaume Masson (2013). "Dynamic interaction between retinal and extraretinal signals in motion integration for smooth pursuit". In: *Journal of Vision* 13.13, p. 5 (cit. on pp. 27, 51, 68, 69, 85, 92, 93).

Bringuier, V et al. (1999). "Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons." In: *Science (New York, N.Y.)* 283.5402, pp. 695–699 (cit. on p. 21).

Burgi, Pierre Y., Alan L. Yuille, and Norberto Grzywacz (Aug. 2000). "Probabilistic Motion Estimation Based on Temporal Coherence". In: *Neural Comput.* 12.8, pp. 1839–67 (cit. on pp. i, iii, 27, 35–39, 49, 53, 62, 85, 91, 92, 138).

Busse, Laura, Alex Wade, and Matteo Carandini (2009). "Representation of concurrent stimuli by population activity in visual cortex." In: *Neuron* 64.6, pp. 931–942 (cit. on p. 22).

Changizi, Mark (2001). "'Perceiving the present' as a framework for ecological explanations of the misperception of projected angle and angular size". In: *Perception* 30.2, pp. 195–208 (cit. on p. 97).

Changizi, Mark and David Widders (2002). "Latency correction explains the classical geometrical illusions". In: *Perception* 31.10, pp. 1241–1262 (cit. on p. 97).

Churchland, Mark M., I-Han H. Chou, and Stephen G. Lisberger (Oct. 2003). "Evidence for object permanence in the smooth-pursuit eye movements of monkeys." In: *Journal of neurophysiology* 90.4, pp. 2205–2218 (cit. on p. 26).

d'Avossa, Giovanni et al. (2007). "Spatiotopic selectivity of BOLD responses to visual motion in human area MT." In: *Nature neuroscience* 10.2, pp. 249–255 (cit. on p. 16).

de Xivry, Jean-Jacques O. and Philippe Lefèvre (Oct. 1, 2007). "Saccades and pursuit: two outcomes of a single sensorimotor process". In: *The Journal of Physiology* 584.1, pp. 11–23 (cit. on pp. 24, 92).

DeLucia, Patricia R. (2004). "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, pp. 243–285 (cit. on p. 23).

Deneve, Sophie, Jean-Rene Duhamel, and Alexandre Pouget (May 23, 2007). "Optimal Sensorimotor Integration in Recurrent Cortical Networks: A Neural Implementation of Kalman Filters". In: *The Journal of Neuroscience* 27.21, pp. 5744–5756 (cit. on p. 51).

Dubner, R. and S. Zeki (Dec. 1971). "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, pp. 528–532 (cit. on p. 23).

Eagleman, David M. and Terrence J. Sejnowski (Mar. 17, 2000). "Motion Integration and Postdiction in Visual Awareness". In: *Science* 287.5460, pp. 2036–2038 (cit. on pp. 10, 120, 121, 145).

Eckmiller, R. and M. Mackeben (Oct. 18, 1978). "Pursuit eye movements and their neural control in the monkey." In: *Pflügers Archiv : European journal of physiology* 377.1, pp. 15–23 (cit. on pp. 24, 140).

Erlhagen, Wolfram (2003). "Internal models for visual perception". In: *Biological cybernetics* 88.5, pp. 409–417 (cit. on pp. 40–42, 97, 145, 147).

Fischer, Jason and David Whitney (2009). "Precise discrimination of object position in the human pulvinar." In: *Human brain mapping* 30.1, pp. 101–111 (cit. on p. 15).

Fröhlich, F.W. (1923). "Über die Messung der Empfindungszeit". In: *Zeitschrift für Sinnesphysiologie* 54, pp. 58–78 (cit. on pp. 11, 98, 117, 142).

Fu, Yu-Xi, Yaosong Shen, and Yang Dan (Oct. 15, 2001). "Motion-Induced Perceptual Extrapolation of Blurred Visual Targets". In: *The Journal of Neuroscience* 21.20, RC172 (cit. on pp. 11, 23).

Fu, Yu-Xi et al. (2004). "Asymmetry in Visual Cortical Circuits Underlying Motion-Induced Perceptual Mislocalization". In: *The Journal of Neuroscience* 24.9, pp. 2165–2171 (cit. on pp. 12, 14).

Goldreich, D., R. J. Krauzlis, and S. G. Lisberger (Mar. 1992). "Effect of changing feedback delay on spontaneous oscillations in smooth pursuit eye movements of monkeys." In: *Journal of neurophysiology* 67.3, pp. 625–638 (cit. on p. 26).

Gollisch, Tim and Markus Meister (Jan. 28, 2010). "Eye Smarter than Scientists Believed: Neural Computations in Circuits of the Retina". In: *Neuron* 65.2 (cit. on pp. 22, 23).

Graf, Erich, Paul A. Warren, and Laurence T. Maloney (Oct. 22, 2003). "Extrapolation of motion paths behind an occluder". In: *Journal of Vision* 3.9, p. 791 (cit. on p. 22).

Grinvald, A et al. (1994). "Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 14.5 Pt 1, pp. 2545–2568 (cit. on p. 21).

Grzywacz, N. M., S. N. Watamaniuk, and S. P. McKee (Nov. 1995). "Temporal coherence theory for the detection and measurement of visual motion." In: *Vision research* 35.22 (cit. on pp. 34, 35, 138).

Guo, Kun et al. (2007). "Spatio-temporal prediction and inference by V1 neurons." In: *The European journal of neuroscience* 26.4, pp. 1045–1054 (cit. on pp. 18, 19, 96, 105, 117, 142).

Harrison, L. M. et al. (Feb. 1, 2007). "Extra-classical receptive field effects measured in striate cortex with fMRI." In: *NeuroImage* 34.3, pp. 1199–1208 (cit. on p. 20).

Hedges, James H., Alan A. Stocker, and Eero P. Simoncelli (May 19, 2011). "Optimal inference explains the perceptual coherence of visual motion stimuli". In: *Journal of Vision* 11.6 (cit. on p. 27).

Hick, We (1950). "The threshold for sudden changes in the velocity of a seen object." In: *The Quarterly journal of experimental psychology* 2.1, pp. 33–41 (cit. on p. 10).

Hogendoorn, Hinze, Thomas A. Carlson, and Frans A. Verstraten (Mar. 2008). "Interpolation and extrapolation on the path of apparent motion." In: *Vision research* 48.7 (cit. on p. 23).

Hubel, DH and TN Wiesel (1962). "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex." In: *The Journal of physiology* 160, pp. 106–154 (cit. on p. 3).

— (1968). "Receptive fields and functional architecture of monkey striate cortex." In: *The Journal of physiology* 195.1, pp. 215–243 (cit. on pp. 3, 4, 137).

Hunter, John D. (May 2007). "Matplotlib: A 2D Graphics Environment". In: *Computing in Science and Engineering* 9.3, pp. 90–95 (cit. on p. 62).

Hürlimann, Felix, Daniel C. Kiper, and Matteo Carandini (Sept. 2002). "Testing the Bayesian model of perceived speed." In: *Vision research* 42.19, pp. 2253–2257 (cit. on pp. 49, 91).

Ilg, U. J. and Peter P. Thier (May 7, 2003). "Visual Tracking Neurons in Primate Area MST Are Activated by Smooth-Pursuit Eye Movements of an "Imaginary" Target". In: *Journal of Neurophysiology* 90.3, pp. 1489–1502 (cit. on p. 24).

Ilg, Uwe J. (Oct. 1997). "Slow eye movements". In: *Progress in Neurobiology* 53.3, pp. 293–329 (cit. on pp. 23, 25).

Isard, Michael and Andrew Blake (1998). "Condensation – conditional density propagation for visual tracking". In: *International Journal of Computer Vision* 29.1, pp. 5–28 (cit. on pp. 48, 51, 52, 57, 62, 148).

Jancke, Dirk and Wolfram Erlhagen (2010). ""Bridging the gap: a model of common neural mechanisms underlying the Fröhlich effect, the flash-lag effect, and the representational momentum effect"". In: *Space and Time in Perception and Action. 1st ed, Cambridge: Cambridge University Press* 422-440 (cit. on p. 42).

Jancke, Dirk et al. (2004). "Shorter latencies for motion trajectories than for flashes in population responses of cat primary visual cortex". In: *The Journal of Physiology* 556.3, pp. 971–982 (cit. on pp. 7, 10, 12, 40, 120, 121, 134, 135, 142, 147).

Kalman, Rudolph, and Emil (1960). "A New Approach to Linear Filtering and Prediction Problems". In: *Transactions of the ASME–Journal of Basic Engineering* 82.Series D, pp. 35–45 (cit. on p. 50).

Kanai, Ryota, Bhavin R. Sheth, and Shinsuke Shimojo (Oct. 2004). "Stopping the motion and sleuthing the flash-lag effect: spatial uncertainty is the key to perceptual mislocalization." In: *Vision research* 44.22, pp. 2605–2619 (cit. on pp. 134, 144).

Kaplan, Bernhard A. et al. (2014). "Signature of an anticipatory response in area V1 as modeled by a probabilistic model and a spiking neural network". In: Proceedings of the 2014 International Joint Conference on Neural Networks; Beijing, China (cit. on pp. ii, iv, 104, 149).

Kawano, K. (1999). "Ocular tracking: behavior and neurophysiology". In: *Current Opinion in Neurobiology* 9 (cit. on p. 25).

Kerzel, Dirk and Karl R. Gegenfurtner (Nov. 11, 2003). "Neuronal processing delays are compensated in the sensorimotor branch of the visual system." In: *Current biology : CB* 13.22, pp. 1975–1978 (cit. on p. 6).

Khoei, Mina, Guillaume Masson, and Laurent Perrinet (2013). "Motion-based prediction explains the role of tracking in motion extrapolation". In: *Journal of Physiology-Paris* 107.5, pp. 409–420 (cit. on pp. i, 104, 135, 138, 139, 148).

Kitano, M et al. (1994). "Retinotopic and nonretinotopic field potentials in cat visual cortex." In: *Visual neuroscience* 11.5, pp. 953–977 (cit. on p. 22).

Kourtzi, Z and N Kanwisher (2000). "Activation in human MT/MST by static images with implied motion." In: *Journal of cognitive neuroscience* 12.1, pp. 48–55 (cit. on p. 140).

Krauzlis, R. J. and S. G. Lisberger (Mar. 1989). "A Control Systems Model of Smooth Pursuit Eye Movements with Realistic Emergent Properties". In: *Neural Computation* 1.1, pp. 116–122 (cit. on p. 26).

Krekelberg (Feb. 2001). "The persistence of position". In: *Vision Research* 41.4, pp. 529–539 (cit. on pp. 6, 8, 121, 126, 144).

Krekelberg and Lappe (1999). "Temporal recruitment along the trajectory of moving objects and the perception of position." In: *Vision research* 39.16, pp. 2669–2679 (cit. on pp. 16, 17, 144).

— (Jan. 2000). "A model of the perceived relative positions of moving objects based upon a slow averaging process". In: *Vision Research* 40.2, pp. 201–215 (cit. on pp. 6, 9, 121, 144).

— (June 2001). "Neuronal latencies and the position of moving objects." In: *Trends in neurosciences* 24.6, pp. 335–339 (cit. on p. 6).

Krekelberg et al. (Aug. 18, 2000). "The Position of Moving Objects". In: *Science* 289.5482, p. 1107 (cit. on pp. 6, 120, 126, 146).

Lim, Heejin and Yoonsuck Choe (2006). "Compensating for neural transmission delay using extrapolatory neural activation in evolutionary neural networks". In: *Neural Information Processing Letters and Reviews*, pp. 147–161 (cit. on pp. 46, 145, 147).

Lim, Heejin and Yoonsuck Choe (Oct. 2008). "Extrapolative delay compensation through facilitating synapses and its relation to the flash-lag effect." In: *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council* 19.10, pp. 1678–1688 (cit. on pp. 46, 145, 147).

Lisberger, S. G., E. J. Morris, and L. Tychsen (1987). "Visual motion processing and sensory-motor integration for smooth pursuit eye movements." In: *Annual Review of Neuroscience* 10 (cit. on p. 25).

MacKay, D. M. (Feb. 15, 1958). "Perceptual Stability of a Stroboscopically Lit Visual Field containing Self-Luminous Objects". In: *Nature* 181.4607, pp. 507–508 (cit. on pp. 7, 120).

Madelain, Laurent and Richard J Krauzlis (Aug. 2003). "Effects of learning on smooth pursuit during transient disappearance of a visual target." In: *Journal of neurophysiology* 90.2, pp. 972–982 (cit. on pp. 25, 26).

Makin, Alexis D. and Ellen Poliakoff (July 8, 2011). "Do common systems control eye movements and motion extrapolation?" In: *The Quarterly Journal of Experimental Psychology* 64.7, pp. 1327–1343 (cit. on pp. 23, 25).

Makin, Alexis D., Ellen Poliakoff, and Wael El-Deredy (Mar. 2009). "Tracking visible and occluded targets: Changes in event related potentials during motion extrapolation". In: *Neuropsychologia* 47.4, pp. 1128–1137 (cit. on pp. 23, 25, 139).

Masson, Guillaume S., Anna Montagnini, and Uwe J. Ilg (2010). "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. Chap. 8, pp. 161–188 (cit. on p. 25).

Masson, Guillaume S. and Laurent U. Perrinet (Jan. 21, 2012). "The behavioral receptive field underlying motion integration for primate tracking eye movements". In: *Neuroscience & Biobehavioral Reviews* 36.1, pp. 1–25 (cit. on p. 91).

Maunsell, J. H. and D. C. Van Essen (May 1983). "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 (cit. on p. 24).

Maus, Gerrit, Jason Fischer, and David Whitney (2013). "Motion-Dependent Representation of Space in Area MT+". In: *Neuron* 78.3, pp. 554–562 (cit. on pp. 5, 7, 9, 16, 40, 142, 147).

Maus, Gerrit et al. (2010). "Does Area V3A Predict Positions of Moving Objects?" In: *Frontiers in psychology* 1 (cit. on pp. 15, 142, 147).

Maus, Gerrit et al. (2013). "The perceived position of moving objects: transcranial magnetic stimulation of area MT+ reduces the flash-lag effect." In: *Cerebral cortex (New York, N.Y. : 1991)* 23.1, pp. 241–247 (cit. on pp. 5, 15, 40, 142, 147, 149).

McAnany, JASON J. and Kenneth r. Alexander (Mar. 9, 2009). "Is there an omitted stimulus response in the human cone flicker electroretinogram?" In: *Visual Neuroscience* 26.02, pp. 189–194 (cit. on pp. 24, 139).

Mikami, Akichika et al. (1986). "Motion selectivity in macaque visual cortex. III. Psychophysics and physiology of apparent motion". In: *J. Neurophysiol*, pp. 1340–1351 (cit. on p. 16).

Montagnini, Anna et al. (Jan. 2007). "Bayesian modeling of dynamic motion integration". In: *Journal of Physiology-Paris* 101.1-3, pp. 64–77 (cit. on pp. 27, 49, 140).

Naka, K. I. and W. A. Rushton (Aug. 1966). "S-potentials from luminosity units in the retina of fish (Cyprinidae)." In: *Journal of Physiology* 185.3 (cit. on p. 75).

Nauhaus, Ian et al. (2008). "Stimulus contrast modulates functional connectivity in visual cortex". In: *Nature Neuroscience* 12.1, pp. 70–76 (cit. on p. 22).

Newsome, W. T., R. H. Wurtz, and H. Komatsu (Aug. 1988). "Relation of cortical areas MT and MST to pursuit eye movements. II. Differentiation of retinal from extraretinal inputs." In: *Journal of neurophysiology* 60.2, pp. 604–620 (cit. on pp. 24, 25, 92, 139).

Nijhawan, Romi (July 28, 1994). "Motion extrapolation in catching". In: *Nature* 370.6487 (cit. on pp. 7, 8, 16, 23, 40, 95, 120, 133, 138, 143, 144, 146, 147).

— (Sept. 1, 2002). "Neural delays, visual motion and the flash-lag effect." In: *Trends in cognitive sciences* 6.9 (cit. on pp. 5, 8–11, 120, 122, 135, 146).

Nijhawan, Romi and Si Wu (2009). "Compensating time delays with neural predictions: are predictions sensory or motor?" In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367.1891, pp. 1063–1078 (cit. on pp. 44, 45, 100, 120, 141, 148).

Nijhawan, Romi et al. (2004). "Compensation of neural delays in visual-motor behaviour: No evidence for shorter afferent delays for visual motion". In: *Visual cognition* 11.2-3, pp. 275–298 (cit. on pp. 6, 120).

Nowlan, Steven J. and Terrence J. Sejnowski (1995). "A selection model for motion processing in area MT of primates". In: *Journal of Neuroscience* 15, pp. 1195–1214 (cit. on p. 27).

Oliphant, T. E. (May 2007). "Python for Scientific Computing". In: *Computing in Science and Engineering* 9.3, pp. 10–20 (cit. on p. 62).

Olson, Ingrid et al. (2004). "Neuronal representation of occluded objects in the human brain". In: *Neuropsychologia* 42.1, pp. 95–104 (cit. on p. 139).

Pack, C. C. and R. T. Born (2001a). "Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain". In: *Nature* 409 (cit. on p. 5).

Pack, Christopher C. and Richard T. Born (Feb. 22, 2001b). "Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain". In: *Nature* 409, pp. 1040–1042 (cit. on p. 24).

Pavel, M., H. Cunningham, and V. Stone (Nov. 1992). "Extrapolation of linear motion." In: *Vision research* 32.11, pp. 2177–2186 (cit. on p. 26).

Perrinet, Laurent U., Rick A. Adams, and Karl Friston (2014). "Active Inference, eye movements and oculomotor delays". In: *Biological Cybernetics* (cit. on p. 135).

Perrinet, Laurent U. and Guillaume S. Masson (Jan. 2007). "Modeling spatial integration in the ocular following response using a probabilistic framework". In: *Journal of Physiology-Paris* 101.1-3, pp. 46–55 (cit. on pp. 73, 75, 103).

Perrinet, Laurent U. and Guillaume S. Masson (Oct. 31, 2012). "Motion-Based Prediction Is Sufficient to Solve the Aperture Problem". In: *Neural Computation* 24.10, pp. 2726–2750. arXiv: 1208.6471 (cit. on pp. i, iii, 52, 63, 75, 87, 92, 93, 138, 148).

Pulfrich, C (1922). "Die Stereoskopie im Dienste der isochromen und heterochromen Photometrie". In: 10.25, pp. 553–564 (cit. on p. 6).

Purushothaman, Gopathy et al. (Dec. 3, 1998). "Moving ahead through differential visual latency". In: *Nature* 396.6710, p. 424 (cit. on pp. 6, 121).

Ramachandran, V. and S. Anstis (Jan. 1983). "Extrapolation of motion path in human visual perception". In: *Vision Research* 23.1, pp. 83–85 (cit. on p. 26).

Rao, Rajesh and Dana Ballard (1999). "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects". In: *Nat Neurosci* 2.1, pp. 79–87 (cit. on p. 20).

Ray, Supratim and John Maunsell (2011). "Network Rhythms Influence the Relationship between Spike-Triggered Local Field Potential and Functional Connectivity". In: *The Journal of Neuroscience* 31.35, pp. 12674–12682 (cit. on p. 22).

Reynaud, Alexandre, Guillaume S. Masson, and Frédéric Chavane (Sept. 5, 2012). "Dynamics of Local Input Normalization Result from Balanced Short- and Long-Range Intracortical Interactions in Area V1". In: *The Journal of Neuroscience* 32.36, pp. 12558–12569 (cit. on p. 91).

Roach, Neil, Paul McGraw, and Alan Johnston (2011). "Visual motion induces a forward prediction of spatial pattern." In: *Current biology : CB* 21.9, pp. 740–745 (cit. on pp. 20, 21, 105).

Robinson, D. A. (Dec. 31, 1973). "Models of the saccadic eye movement control system." In: *Kybernetik* 14.2, pp. 71–83 (cit. on p. 26).

Robinson, D. A., J. L. Gordon, and S. E. Gordon (Oct. 1, 1986). "A model of the smooth pursuit eye movement system". In: *Biological Cybernetics* 55.1, pp. 43–57 (cit. on p. 26).

Rosen, Robert (2012). *Anticipatory Systems, Philosophical, Mathematical, and Methodological Foundations*. 2nd. Vol. 1. IFSR International Series on Systems Science and Engineering. Springer (cit. on pp. 146, 147).

Rosenbaum, D. A. (Nov. 1975). "Perception and extrapolation of velocity and acceleration." In: *Journal of experimental psychology. Human perception and performance* 1.4, pp. 395–403 (cit. on p. 23).

Rust, Nicole C. et al. (2006). "How MT cells analyze the motion of visual patterns". In: *Nature Neuroscience* 9.11, pp. 1421–31 (cit. on pp. 75, 91).

Schlag, J. and M. Schlag-Rey (Mar. 1, 2002). "Through the eye, slowly: delays and localization errors in the visual system." In: *Nature reviews. Neuroscience* 3.3, pp. 191–215 (cit. on p. 8).

Schlag, John et al. (Jan. 6, 2000). "Neuroscience: Extrapolating movement without retinal motion". In: *Nature* 403.6765, pp. 38–39 (cit. on p. 120).

Schwartz, G. and M. J Berry (Jan. 23, 2008). "Sophisticated temporal pattern recognition in retinal ganglion cells." In: *Journal of neurophysiology* 99.4, pp. 1787–1798 (cit. on pp. 12, 23, 24, 139).

Senior, C et al. (2000). "The functional neuroanatomy of implicit-motion perception or representational momentum." In: *Current biology : CB* 10.1, pp. 16–22 (cit. on p. 140).

Sheth, Bhavin R., Romi Nijhawan, and Shinsuke Shimojo (May 2000). "Changing objects lead briefly flashed ones". In: *Nat Neurosci* 3.5, pp. 489–495 (cit. on p. 120).

Simoncelli, Eero P. and David J. Heeger (Mar. 1998). "A model of neuronal responses in visual area MT". In: *Vision Research* 38.5, pp. 743–761 (cit. on pp. 75, 91).

Slovin, Hamutal et al. (2002). "Long-term voltage-sensitive dye imaging reveals cortical dynamics in behaving monkeys." In: *Journal of neurophysiology* 88.6, pp. 3421–3438 (cit. on p. 21).

Sotiropoulos, Grigorios, Aaron R. Seitz, and Peggy Seriès (Nov. 2011). "Changing expectations about speed alters perceived motion direction". In: *Current Biology* 21.21, R883–R884 (cit. on p. 27).

Spering, Miriam et al. (2005). "Effects of contrast on smooth pursuit eye movements". In: *Journal of Vision* 5 (cit. on p. 91).

Stettler, Dan et al. (2002). "Lateral Connectivity and Contextual Interactions in Macaque Primary Visual Cortex". In: *Neuron* 36.4, pp. 739–750 (cit. on p. 117).

Stocker, Alan A. and Eero P. Simoncelli (Mar. 19, 2006). "Noise characteristics and prior expectations in human visual speed perception". In: *Nature Neuroscience* 9.4, pp. 578–585 (cit. on p. 27).

Tlapale, Emilien, Guillaume Masson, and Pierre Kornprobst (Aug. 6, 2010). "Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism." In: *Vision research* 50.17, pp. 1676–1692 (cit. on pp. 65, 118).

Watamaniuk, S. N., S. P. McKee, and Norberto M. Grzywacz (Jan. 1995a). "Detecting a trajectory embedded in random-direction motion noise." In: *Vision Research* 35.1, pp. 65–77 (cit. on pp. 72, 74, 77).

Watamaniuk, Scott N. J., Suzanne P. McKee, and NM. Grzywacz (Jan. 1995b). "Detecting a trajectory embedded in random-direction motion noise". In: *Vision Research* 35.1 (cit. on pp. 34, 138).

Weiss, Yair and David J. Fleet (2001). "Velocity likelihoods in biological and machine vision". In: *In Probabilistic Models of the Brain: Perception and Neural Function*, pp. 81–100 (cit. on pp. 49, 103).

Weiss, Yair, Eero P. Simoncelli, and Edward H. Adelson (June 20, 2002). "Motion illusions as optimal percepts". In: *Nature Neuroscience* 5.6, pp. 598–604 (cit. on pp. 27, 52, 62, 65, 73, 91).

Welch, Greg and Gary Bishop (1995). *An Introduction to the Kalman Filter*. Tech. rep. Chapel Hill, NC, USA (cit. on p. 27).

Whitney, David and David Bressler (2007). "Spatially asymmetric response to moving patterns in the visual cortex: re-examining the local sign hypothesis." In: *Vision research* 47.1, pp. 50–59 (cit. on p. 12).

Whitney, David, Patrick Cavanagh, and Ikuya Murakami (Dec. 2000). "Temporal facilitation for moving stimuli is independent of changes in direction". In: *Vision Research* 40.28, pp. 3829–3839 (cit. on p. 6).

Whitney, David and Ikuya Murakami (Dec. 1998). "Latency difference, not spatial extrapolation". In: *Nat Neurosci* 1.8, pp. 656–657 (cit. on pp. 6, 9, 120, 121).

Whitney, David, Ikuya Murakami, and Patrick Cavanagh (Jan. 2000). "Illusory spatial offset of a flash relative to a moving stimulus is caused by differential latencies for moving and flashed stimuli". In: *Vision Research* 40.2, pp. 137–149 (cit. on pp. 6, 121).

Whitney, David et al. (2003). "Flexible retinotopy: motion-dependent position coding in the visual cortex." In: *Science (New York, N.Y.)* 302.5646, pp. 878–881 (cit. on pp. 12, 142).

Wuerger, Sophie et al. (Jan. 22, 2010). "Motion extrapolation of auditory-visual targets". In: *Information Fusion* 11.1, pp. 45–50 (cit. on pp. 23, 27).

Xivry, Jean-Jacques O. de et al. (Oct. 30, 2013). "Kalman Filtering Naturally Accounts for Visually Guided and Predictive Smooth Pursuit Dynamics". In: *Journal of Neuroscience* 33.44, pp. 17301–17313 (cit. on p. 51).

Yuille, Alan L. and Norberto M. Grzywacz (June 1989). "A mathematical analysis of the motion coherence theory". In: *International Journal of Computer Vision* 3.2, pp. 155–175 (cit. on pp. i, iii, 32–34, 71, 138).