

A look into the black box: eye movements as a probe of visual motion processing*

Guillaume S. Masson¹, Daniel R. Mestre¹

¹ Centre de Recherche en Neurosciences Cognitives. CNRS. Marseille

Feb 11, 1998

Abstract

The ability of a living organism to detect and process changes in the visual world and to use visual motion information for the control of behavior is remarkable. In primates, several cortical areas organized in a hierarchical way are necessary to compute local measurements of visual motion and then to extract higher order information about object- and self-motion. Such processing allows us to deal with complex optical flow resulting from our displacement in a tridimensional environment. During such displacement, we can relocate our gaze and lock it on a newly acquired visual target. Visual stabilization of the eyes requires a complex processing of the translational component of the visual flow in which the motion of the object of interest is embedded. By recording eye movements triggered by visual motion, we can dissect the various stages of this visual motion processing and distinguish between the relative contributions of low-level, automatic filtering and higher-order, cognitive, selection of the motion of interest for the oculomotor system. In this review, we claim that the analysis of eye movements can be used as a probe of visual motion processing. By recording the very first part of tracking responses, we are able to identify some elementary processing modules corresponding to early linear filters implemented in the early stages of the cortical motion pathway. By studying gradual changes of eye movement behaviors over time, as a function of perceptual or attentional states of the observers, we can understand the dynamical relationships between perception and action. Within this experimental framework, we may start to open the "black box" in which visual information processing was kept closed in most of oculomotor models for many years. Furthermore, this might be a starting point towards the investigation of complex interplays between perception and action in the control of spatial behaviors.

1 Introduction

The visual world around us is constantly changing. Such changes in the optic array not only result from external events such as moving objects or animals but also from the motion of our own eyes, head or whole body relative to its environment. Gibson (1950) first suggested that the optical flow was an adequate way of describing the visual input to a living organism interacting with the environment. First, the optical flow describes inputs as continuous changes in the visual array rather than as a succession of static images, emphasising the key role of visual motion information. Second, optic flows carry

*To be published in *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*. Please refer to the journal version for citation

information related to both the observer's self-motion and the 3D layout of the environment (see Figure 1). Therefore, the concept of optic flow is particularly seminal when one tries to understand the relationships between a living organism and its surroundings: global optic flow fields are generated by the organism's behavior and help the seeing organism control its actions.

In the last decade, many studies have been devoted to the investigation of the mechanisms and neural substrates that the brain uses to process the optic flow field. Psychophysical evidence suggests that motion processing occurs at least over two successive stages. At the first stage, spatio-temporal filtering mechanisms compute motion measurements within local regions of the visual space. At the second stage, these local motion signals are combined together, in order to extract higher order information about object- and/or self-motion (see Smith and Snowden, 1994). Such higher order visual signals allow human subjects to report the direction of self-motion (or 'heading') with high accuracy or to reconstruct the tridimensional form of a visual object defined solely by the relative differential motions of individual elements (see Cornilleau-Perez & Gielen, 1997, for a recent review). Many neurophysiological data have been gathered on brain mechanisms of visual processing (see Albright & Stoner, 1995 for a review). A classical assumption is that the complex mosaic of more than 20 extrastriate visual areas is organized into hierarchical pathways (Zeki, 1994) which implement the logical steps of visual processing, from the extraction of visual primitives to the emergence of a coherent percept (e.g. Movshon *et al.*, 1985; Rodman & Albright, 1989; Salzman & Newsome, 1994). For instance, the primary visual cortex is assumed to deal with the problem of local motion measurements (Movshon & Newsome, 1996), while "higher" visual areas, such as MT or MST, contain neurons selectively activated by different types of optic flow fields (see Wurtz & Duffy, 1992).

Such general purpose visual processing is not only involved in conscious perception. Vision is essential for controlling actions, especially eye movements, and a lot of evidence indicates that oculomotor control involves various stages along these hierarchical pathways (see Krauzlis, 1994). For instance, reflexive smooth eye movements are driven by the motion of the visual projection of an object onto the retina. A walking observer can relocate his/her gaze towards an object of interest located in an eccentric part of the visual field and/or in another depth plane. As his/her own displacement relative to the object logically results in a retinal slip of the newly acquired visual target, visual stabilization mechanisms act, in synergy with vestibulo-ocular reflexes (see Waespe & Henn, 1987), to keep this image in the highest-resolution part of the retina: the fovea. Moreover, to improve the visual analysis of this object, the two images falling onto the two retinas must be kept in close correspondence. This is done by adjusting the binocular point of fixation on the target (see Howard & Rogers for a comprehensive review). To achieve such complex motor coordination of the two eyes, the oculomotor system needs precise information about the motion and the location in depth of the selected object. This information must be unambiguous and non relevant motion signals from other parts of the visual scene must be filtered out.

In this mainstream, we would like to present here some new results suggesting that simple, automatic processing of the optic flow field are implemented into the visual system, that allow the oculomotor system to initiate appropriate eye movements with very short latency. As originally suggested by Miles (1993), early "machine-like" responses might indicate what are the elementary, build-in, modules of optic flow processing in the primate visual system. Such processes seem to exist prior to any coherent percept (e.g. Masson *et al.*, 1997). Nevertheless, a great challenge is to understand how several, elementary modules are put together in order to allow a steady-state control of smooth eye movements in a flexible way. Thus, it becomes essential to understand the nature of the relationships between a simple motor system, such as the smooth eye movement control system and the perceptual state of the subject. A large number of studies indicate that

eye movements also depend on attentional and perceptual factors¹. The clarification of these relationships is crucial, due to the reciprocal nature of the linkage between eye movements and visual inputs: visual inputs are an important drive for eye movements, and, in return, eye movements transform the visual inputs and, *in fine* visual perception. Thus, relationships between perception and eye movements provide an exquisite approach to understand the complex problem of the so-called "perception-action coupling". Such framework calls for the accurate description of the dynamical aspects of both the visual and oculomotor systems and of the dynamics of their interaction in the visual stabilization of gaze in space for instance.

In this review, we will describe two examples illustrating how the analysis of eye movements can provide a very useful insight into the understanding of how the "visual" brain works. First we will describe one example of the elementary early filtering modules that constrain visual motion processing. Next, we will show that a common representation of the 3D visual surrounding can be used by both the oculomotor and perceptual systems. Finally, we will discuss the functional consequences of the coupling between visual perception and eye movements in order to suggest that early smooth eye movements which are initiated before the emergence of any stable percept might indeed contribute to the build-up of a conscious percept, by transforming the visual flow field.

2 Selection of motion: automatic stereoscopic filtering

We previously mentioned that different types of optic flow exist, depending on the type of self motion. In the absence of any compensatory eye movement, a pure rotation of the vantage point results in a rigid rotation of the optic array, with flow lines resembling lines of latitude on a globe whose poles are defined by the axis of rotation (Figure 1A). The main characteristic of such an optic flow field is that the visual direction of motion and visual velocity of all elements are completely dictated by the angular velocity of the observer's rotation. An example of this situation is the rotation of the head in space. It is admitted that the resulting retinal image motion can be fully compensated by an appropriate eye counter-rotation in the orbits. Notice however that such a description results from a crude approximation of the actual retinal flows. As the eyes axes-of-rotation are located slightly in front of the head axis-of-rotation, head rotation in space results in slight translations of the retinal images. That is to say that pure rotational optic flows are generated by, and only by, eye movements (Koenderink, 1986). The actual consequence of such an approximation requires further investigation. In particular, the ability of the visual and oculomotor systems to distinguish between rotational and translational components of an optical flow field remains unclear. By contrast, a pure translation of the observer results in a more serious challenge for the visual stabilization mechanisms. Now, the flow pattern consists of motion along the lines of longitude on the globe, whose poles are defined by the direction of self-motion, with flow lines emerging from the pole ahead (the focus of expansion or focus of outflow) and disappearing into the pole behind (the focus of contraction) (Figure 1B). The optical velocity of an element at a given eccentricity depends on the changing relative distance between environmental elements and the observer: nearby objects move across the field more rapidly than distant ones. Such an inverted metric between viewing distance and optical velocity creates motion parallax between distant objects whose retinal images are close together. Moreover, the actual *retinal* image motion depends on the observer's motion and on the location of gaze relative to the direction of displacement. As the observer stares in the vicinity of the direction of

¹Most of the studies concerned with the effects of cognitive or attentional factors on eye movements have focused on saccadic eye movements. However, for a review of previous work on pursuit eye movements, see Kowler (1990)

self-motion, visual motion streams past the observer on all sides, creating a radial flow which emanates from the focus of expansion, lying in the direction of self-motion (Figure 1C,b). Such *radial* flows also signal a change in the distance of all elements located in front of the walking observer. The visual velocity of a given element increases as the observer approaches it. However, when the observer stares in a direction perpendicular to his/her displacement, there are no changes in the relative distance between the observer and environmental elements. All visual elements move in the same direction, but with different visual velocities, depending of their actual distance to the observer. Such *lamellar* optic flow field is illustrated by the situation of an observer looking through a window of a moving train (Figure 1C,d).

Such complex flow fields are constantly generated in our daily life, and biological systems have developed exquisite solutions to deal with them. For instance, when a moving animal directs its gaze towards a prey or an obstacle, it usually does it by a combination of eye and head movements without stopping its forward motion. The resulting complex distortion of the visual image must be analyzed and some stabilization mechanism must hold the corresponding images of the object of interest on the fovea of both retinas. Such stabilization can be solved by an appropriate combination of conjugate (or version) and disconjugate (or vergence) eye movements. There is now some evidence, in human and non-human primates, that such version and vergence eye movements can be triggered by complex flow fields (e.g. Kawano *et al.*, 1986; Gellman *et al.*, 1990; Busetini *et al.*, 1996; Mestre & Masson, 1997a). For instance, pure radial flows can generate appropriate vergence eye movements at short latencies (≈ 80 ms) in humans (Busettini *et al.*, 1997).

The observation of such short latency, "machine-like" eye movements is particularly interesting, since it points out that there are some fast, "build-in" optic flow processors within the primate visual system, which can quickly deliver a drive signal to the oculomotor system for visual stabilization of gaze. The importance of studying the initiation of eye movements has been underscored for a long time (see Lisberger *et al.* 1987; Miles, 1993). Reflexive or voluntary pursuit eye movements may be understood as the outcome of a negative feedback loop which operates to minimize the continuous slippage of the visual scene on the retina. Thus, the actual *retinal* optic flow depends both on the displacement of environmental elements relative to the vantage points and on the movements of the eyes themselves. In order to start understanding how ocular responses to optic flow are initiated, we have to open the loop in some way. We might then be able to decipher the relationships between visual motion and behavioral responses. Such relationships might provide a read-out of the complex underlying sensorimotor transformations. A natural way to open the feedback loop is to take advantage of its physiological latency. Several authors have suggested that the initiation of eye movements is the best time to describe the automatic, early visual processes underlying behavioral control. We will describe thereafter one instance of this approach.

As previously mentioned, when a moving observer looks out at the passing scene and fixates the far horizon, distant objects appear stationary whereas nearby objects seem to pass rapidly as though the world was pivoting around the most distant point (Figure 1C,c). If the observer wishes to scrutinize objects in the middle-ground, then he must track them with his eyes and ignore the competing motion of objects that are nearer or farther away. Busettini *et al.* (1996) investigated the suggestion that low-level motion detectors provide the initial drive to the primate tracking system and achieve this selectivity through a binocular stereomechanism (Howard & Simpson, 1989; Miles, 1993). This idea relies on the fact that the object on which the eyes are aligned is imaged at corresponding positions on the two retinas, whereas objects that are nearer or farther away have images that fall onto non-corresponding parts of the two retinas : they are said to have "binocular disparity" (Figure 2A). Thus, the visual motion processing may act as a binocular stereoscopic filtering where motion signals from the plane of fixation are singled out and motion signals from outside the plane of fixation are actively eliminated

from the input to the oculomotor tracking mechanism.

Figure 2 illustrates such an active filtering process. Busetini, Masson and Miles (1996) recorded binocular eye movements using the electromagnetic search coil technique in both macaque monkeys and human subjects. Large textured images were dichoptically presented with or without disparity. One traditional problem is that when the images seen by the two eyes are separated, compelling vergence eye movements are elicited that operate to realign the two eyes on the two images, eliminating the disparity. To avoid this, Busetini and his colleagues have used step-ramp motions of the two images and analyzed only the earliest ocular following responses that were generated "open-loop", that is before eye movements had any chance to affect the visual stimuli projected onto the retinas. The steps were disconjugate, applied immediately after a saccade had landed at the center of the screen, and served to position the binocular image of the random dot pattern nearer ("crossed disparities") or farther ("uncrossed disparities") than the screen. The ramps started synchronously with the steps and were conjugate for 100-200 ms, to elicit ocular following responses with motion in the new depth plane. Such motion elicited a machine-like tracking response at ultra-short latencies in both species (humans : ≈ 85 ms; monkeys : ≈ 55 ms, see Figure 2B). Results plotted in Figure 2C indicate that the magnitude of the tracking response strongly depended on the disparity of the visual scene. The largest responses were observed with a null disparity, that is when the motion signal was located in the fixation depth-plane, whereas responses were weaker and weaker as binocular disparity increased. Quantitative estimates of this effect are given by the change in version position over a brief period of time, during the initial open-loop part of the response (humans: 85-118 ms; monkeys: 60-77 ms). It is evident from Figure 2B that the largest responses are observed for 0 deg. disparity and that the initial response magnitude decreases as disparity increases. A minimum is reached for disparities of 3-5 deg., regardless of the sign of the disparity (uncrossed or crossed) as well as the direction of motion (leftward or rightward). The fact that such filtering is an active process is further demonstrated by the fact that, when compared with monocularly-evoked responses, version magnitude was larger in the plane of fixation and smaller for large disparities.

These results indicate that low-level motion detectors, triggering ocular following responses at short latencies (Miles *et al.*, 1986; Gellman *et al.*, 1990) are, in both species, binocular and disparity-selective. Such selectivity for zero disparity, should make these detectors particularly sensitive to motion located in the vicinity of the plane of fixation in depth. Such neurons have been found in the monkey primary visual cortex (Poggio & Talbot, 1981). They are direction- as well as disparity-selective and a population of such neurons might implement the binocular behavioral filter illustrated in Figure 2C. However, the range of disparities detected by such early visual neurons is much more restricted than the width of the behavioral filter. This is coherent with the small receptive field size described for V1 neurons. This discrepancy between neuronal and behavioral levels argue for a population coding over a large range of binocular disparity selectivity. At what level of the visual motion pathways such neural code is elaborated is still unclear, although recent recording experiments conducted by Kawano and his colleagues suggest that area MST might be the site for such implementation (Kawano *et al.*, 1992; Takemura *et al.*, 1997).

3 Selection of motion: role of attentional and perceptual factors

Integrating different visual cues can thus be an efficient solution for the parsing of complex optic flows, for the selection of a motion signal corresponding to the object of interest and for eliminating competing motion signals from other parts of the visual scene. Such an ac-

tive mechanism is particularly efficient for initiating (Busetini *et al.*, 1996) and maintaining (Howard & Simpson, 1989) reflexive tracking in a crowded visual world. Meanwhile, we have also indicated the usefulness of focusing on the initiation of oculomotor responses in order to isolate early visual processes. However, on a larger time-scale, we also have to understand how perception and action interact in order to adapt behavioral responses to the perceptual and cognitive states of the observer. The case of a motion parallax field offers an experimental framework to tackle this question. In such an optic flow field, all the motion information is controlled and available to the organism after the stimulus onset. However, a complex processing must take place, in order to segment local motion signals and to integrate them into a coherent global percept attributed to moving rigid objects (see Hildreth & Royden, 1995). The problem is then to understand how a representation of a coherent 3D visual environment emerges from the processing of velocity gradients and whether such representation can be used by both the perceptual and the motor system.

Mestre and Masson (1997a) have studied the characteristics of the steady-state optokinetic nystagmus (OKN) elicited in humans by motion parallax flow fields, in which motion direction was constant and a number of visual velocities were randomly distributed in a large random dot kinematogram (RDK). They carefully avoided the presence of other depth cues, such as size, luminance and binocular disparity variations. The available information to depth perception was thus restricted to only one dimension: a velocity gradient. Because vectors are randomly distributed, different motion signals can be found in close vicinity within the image. Such a display defines a motion transparency perception where rigid, transparent surfaces, defined solely by their relative speed, can be perceived moving through each other (see Figure 3). When all the dots moved at the same velocity, the resulting optic flow specified a translation of the observer parallel to a flat vertical surface. When the RDK was divided into three interleaved sub-groups, each one attributed a single velocity (e.g. V_{max} , $V_{max}/2$ and $V_{max}/4$), subjects perceived three transparent surfaces located at three different distances from the vantage point, the slowest group of dots being perceived as the farthest surface away from them. However, while local motion information is available as soon as the stimulus is set into motion, the build-up of a coherent perception is characterized by its sluggish temporal dynamics (Treue *et al.*, 1991). Such slow build-up of a coherent percept contrasts with the fast initiation of reflexive pursuit eye movements. We could then dissociate the automatic initiation of eye movements by low level motion detectors from the long-term stabilization of the behavior, in which perceptual and attentional factors come into play. Concerning oculomotor behavior, the basic hypothesis was that, if a given motion vector is selected to be pursued, such selection must be related at least in part to the same processing of the velocity gradient than that involved in the perception of a three dimensional structure from motion gradient ("depth-from-motion perception, DFM).

With "single-velocity" stimuli, Mestre and Masson (1997a) observed a steady-state OKN where, over large periods of time (up to 30 seconds), slow phase tracking eye movements and saccadic resetting eye movements alternate in a regular pattern. Gain of the slow phases was almost 1 for velocities up to 40 deg/sec. The distributions of slow-phases velocities are illustrated in Figure 3A. For single-velocity flows, the distribution was centered around the stimulus speed and its variability increased as the stimulus speed increased. When subjects were presented with "triple-velocity" flows (Figure 3B), the distribution was broader and centered around the slowest speed present in the flow field. In such conditions, different vectors were tracked one after another with a clear dominance of the slowest speed presented in the motion parallax flow field. Clearly, visual stabilization mechanisms were not controlled by "*en masse*" global motion of the RDK. If, before the stimulus onset, subjects were asked to pay attention to one of the specified surfaces-from-motion, the distribution of slow phases velocity was centered around the visual velocity of the attended surface with a smaller standard deviation, indicating a regular smooth tracking behavior, very similar to that observed in the single-velocity

condition. In a multi-velocity flow (Figure 3C), where more than 10 different velocities were simultaneously displayed, we observed a largely different behavior. Now, the distribution of slow phases velocity was centered around the average velocity of the RDK, with a large standard deviation, biased toward velocities lower than the average velocity. This is the only instance where, in a complex motion parallax field without any depth cues but relative motion, the steady-state OKN appears to be controlled by "*en masse*" motion of the visual scene.

On the perceptual side, subjects always spontaneously perceived "triple-velocity" flows as three, transparent, surfaces sliding in the same direction with different speeds. Moreover, subjects were able to pay attention to individual surfaces defined by their relative depth. By contrast, in the "multi-velocity" case, subjects perceived a cloud of dots, extending in depth and moving in the same direction. No distinct surfaces were reported. This result is consistent with that of psychophysical experiments conducted by Andersen (1989), indicating a similar limit of 3-4 individual planes in the perception of multiple surfaces defined by velocity gradients.

The correlation between perception (i.e. the segregation of 3 surfaces extending in depth *vs.* the perception of a rigid moving object) and oculomotor behavior (single surface *vs.* "*en masse*" tracking) might suggest that both behaviors rely at least partly on a similar representation of the visual motion information². We further tested this hypothesis by investigating the temporal aspects of both responses. We previously mentioned that perception of transparent moving surfaces from optic flow is a rather slow process. Previous studies using structure-from-motion stimuli where subjects had to detect a rigid three dimensional rotating cylinder from a RDK, indicated that the percentage of correct responses increased as stimulus duration increased and that asymptotic performance was never reached before 500 to 600 msec of stimulus duration (Treue *et al.*, 1991). To get a direct comparison between the temporal time-course of perception and oculomotor behavior, we recorded pursuit eye movements while subjects had to decide, quickly after stimulus onset, whether this latter specified a single-surface in motion, two moving transparent surfaces or a clouds of dots moving coherently.

As illustrated in Figure 4B, the rather long reaction times observed with all the stimuli suggest a slow build-up of the perception of surfaces from motion parallax. While for many other psychophysical tests with motion stimuli, reaction times are found typically between 250 and 350 ms. (e.g. Mateeff *et al.* 1995 for the detection of a change in velocity), the discrimination between stimuli defined by, and only by, velocity gradients takes more than 1200 ms. Moreover, it takes subjects longer to detect a three-dimensional moving structure in a double- or multiple-velocity stimulus than it takes to detect a two-dimensional rigid moving surface. These long reaction times largely exceed the latency for pursuit eye movements. Figure 4A illustrates a sample of initial velocity rises of the optokinetic responses triggered by the stimulus onset. Latencies were typically about 150 ms. More interesting is the comparison between dotted and continuous lines, which indicates that initial velocity of eye movements driven by a double-velocity flow is the average of the initial velocities for responses driven by the component vectors presented independently. Such average response is also observed with a multiple-velocity flow. Quantitative analysis of the velocity reached by the eye 400 ms after stimulus onset (but well *before* the psychophysical reaction time) demonstrates that when a flow with more than one velocity is presented, the initiation of the optokinetic is driven by the average velocity of the flow field. For this experiment, stimuli lasted 5 seconds. Because subjects were not instructed to follow

²Representations are taken here as dynamical population coding, as introduced in the motor domain (e.g. Lee *et al.*, 1988). In the sensory domain, population coding can be regarded as the projection of many single cells' responses to a common stimulus into a space representing the stimulus parameter of interest, here the velocity space. In such a representation, a single-velocity stimulus leads to a unique, broadly tuned, activation, while motion transparency leads to multiple broadly tuned activations, simultaneously present in the distributed representation (see Simoncelli, 1993 for example)

a particular plane, we may wonder how the eye velocity changed over time. In the case of double-velocity flow, we found a significant trend in the eye velocity profile after the first saccadic eye movement which occurred at the end of the response initiation. Over the 4 remaining seconds, the instantaneous eye velocity significantly decreased from the initial eye peak velocity to a steady-state eye velocity equal to the slowest velocity in the optic flow. That is to say, the steady-state optokinetic response was driven by the slowest velocity component, as observed in the previous study, when subjects were instructed to pay attention to the pattern as a whole in a triple-velocity flow (Figure 3B). By contrast, eye velocity did not significantly vary over the same time window for the single- or the multiple-velocity flows, as expected for a motor response driven by either the single or the average motion present in the optic flows, respectively.

These results indicate that, in the absence of segregation cues or voluntary selection, the initiation of optokinetic eye movements is driven by the average velocity when several motions are simultaneously, and transparently, presented. Similar velocity averaging has been demonstrated in monkeys for the initiation of voluntary smooth pursuit eye movements, when two spot targets are set into motion simultaneously at the same speed in different directions (Lisberger & Ferrera, 1997). Motion averaging suggests that motor commands for pursuit eye movements rely on a distributed representation of image motion where neurons are tuned for both the direction and speed of target motion. This distributed activity might occur in extra-striate visual area MT (V5) as demonstrated by the effects of micro-stimulations on smooth and saccadic eye movements evoked by moving visual targets (Groh *et al.*, 1997). These two types of results point out the importance of velocity averaging but might lead to largely different interpretations in term of the nature of the averaging location. Lisberger and Ferrera (1997) suggested that the use of velocity averaging to compute motor responses reflects a general computation used by the brain to read a distributed representation of sensory inputs. Such process might be very useful for the integration of local motion measurements, in order to define a reliable velocity signal linked to one particular object. In that sense, velocity averaging is the way the motor system reads out a stable, distributed sensory representation of the velocity of an individual moving stimulus. Moreover, velocity averaging can also provide an accurate estimate of the aggregate velocity of non-rigid objects or objects that are simultaneously rotating and translating (Groh *et al.*, 1997). For a sensorimotor transformation, this mechanism could be equivalent to integration in visual motion processing, i.e. a process that integrates visual motion signals from neighboring locations in the visual field, smoothing out spatial variations in velocity (see Braddick, 1993). For instance, this process might enable the initial smooth eye movements evoked by a dynamic RDK to reflect a vectorial combination of the motion of all the dots with a precision equivalent to the precision of the perceptual decisions based on the same stimulus (Watamaniuk & Heinen, 1994). In the same vein, velocity averaging allows the optokinetic system to perform a steady-state tracking of a rigid moving cloud of dots when more than 3 to 4 visual velocities are simultaneously coded in the distributed representation. Thus, as suggested by Lisberger and Ferrera (1997) such vector averaging is done downstream from the representation of visual motion in area MT.

The second example of velocity averaging that we described for the initiation of optokinetic eye movements triggered by complex flows might result from another process located within the distributed sensory representation, presumably as early as in area MT. Our hypothesis is that the averaging initial ocular response as well as the sluggish temporal dynamics of motion transparency perception might be due to a slow segmentation process. In that view, interactions between multiple motions might lead first to a crude representation of motion centered on the average velocity in the flow field before segmentation of global motion occurs and a simultaneous representation of the two velocities in the same distributed representation occurs. Therefore, the initial averaging motor response might reflect the state of the distributed representation at a given time. Such interactions between multiple motions have been demonstrated in area MT in several studies (Qian &

Andersen, 1994, Movshon *et al.*, 1985; Recanzone and Wurtz, 1994) and can be used to implement some vector averaging at the level of the distributed representation of inputs. Interestingly, such interactions are strongly modulated by attention directed toward one or the other targets (Treue & Maunsell, 1996).

4 Towards a behavioral approach of active vision

We have illustrated several examples of experimental manipulations where a great deal of information about how the visual system works can be dragged from behavioral responses. In this short overview, our goal was to demonstrate that we can probe early motion filtering as well as higher-level motion representation within the visual system by simultaneously recording eye movements and psychophysical performance. We can then address the nature of automatic and attentional processes that are involved in the selection of a moving target, prior to motor programming. We can probe the state of a distributed motion representation at any given time. One future issue is certainly to understand how the transition between an initial "vector averaging" read-out process of such a representation can be followed by a "winner-take-all" read-out process, in which only one among the multiple distinct motion signals present in the distributed representation gains access to the motor system. Such an approach differs largely from the previous cybernetic approach to oculomotor control (Schöner *et al.*, 1996). Results discussed in the present review emphasize the need for a representation of the sensory input, the existence of dynamical changes in such a representation, and different, task-dependent, read-out mechanisms. Physiological (Salzman & Newsome, 1994) and behavioral (Ferrera & Lisberger, 1995; Ferrera & Lisberger, 1997) evidence suggest that, because pursuit eye movements must be made quickly (within 200 ms), the oculomotor system might use a vector averaging of motion information in local regions of space to obtain a rapid guess concerning the direction to be tracked. Another advantage of such smoothing procedure might be to activate simple templates, indicating the direction of tracking in depth and then the coordinated initiation of vergence and version eye movements (Busettoni *et al.*, 1997). However, in complex perceptual contexts, as camouflage breaking and prey tracking, a competitive segregation of distinct motion signals is needed, in order to elaborate an adapted oculomotor behavior. At this stage, an active attentional selection, based on the perceptual state, is involved, as well as decision-making processes. A winner-take-all mechanism for reading the distributed code might enable such selection, both at the level of motion representation (i.e. motion segregation, see Salzman & Newsome, 1994) and downstream from it (i.e. pre-motor target selection, see Ferrera & Lisberger, 1995). A challenge is then to dissociate the respective contribution of these different mechanisms and their relative weight, depending on the task, the time allowed for the response to reach a steady-state or the amount of information available. To tackle these questions is the main goal of behavior-oriented approaches to brain function and, *in fine* cognition (see Mallot, 1997).

Vision is in many respects an active process (see Blake and Yuille, 1992). Most notably, eyes and head movements simplify a number of otherwise complicated problems in the computational theory of vision. Visual tracking of a moving object simplifies recognition and inspection of this object since it is now at rest on the retina. Moreover, pursuit eye movements will also profoundly transform the visual flow in a rigid world: the visual motion of objects located in the depth-vicinity of a motion plane containing the pursued target will be small, while image motion of objects located far from the plane of interest will be large. An additional transformation of the visual flow by eye movements concerns the relative direction of visual motion. For instance, with optic flows such as the "double-velocity" stimulus presented above, a pursuit eye movement driven by the mean stimulus velocity will create a shearing motion pattern: the dots which were previously

moving slower than the average velocity will reverse their direction of motion, while the dots which were moving faster than the average velocity will simply reduce their speed. Are these transformation of the retinal flow, which are generated by eye movements, useful for the perceptual system ? Preliminary results from our laboratory suggest that occurrence of pursuit eye movements in motion-parallax flows trends to lower speed difference thresholds for motion segmentation in a transparent display. More interestingly, the time course of the perceptual build-up was faster for tracking condition, as compared to fixation condition (Mestre & Masson, 1997b). Within this experimental framework, we can next tackle the difficult question of whether such oculomotor responses are involved in motion segmentation and three-dimensional layout analysis by the visual system. There are two ways in which eye movements can be involved in motion perception. We have just mentioned the former, that is modifying the retinal flow pattern. However, many studies have pointed out that perceptual systems must take into account any movements of the receptor. Although initial, averaging eye movement responses greatly modified the retinal flow field in stimuli as illustrated in Figure 3, subjects always perceived surfaces moving in the optical, not retinal, directions. Such results indicate that, together with other experimental evidences, conscious perception is made of both visual and non-visual information regarding the on-going eye movements. Taking into account the role of such extra-retinal information (either efference copy or proprioceptive inflow, see Jeannerod, 1995) is the next crucial step for the approach proposed in the present article. Integrating the contribution of non-visual sources of information might explain why active observers are more accurate in processing structure-from-motion for instance (Rogers & Rogers, 1992; van Damme & van de Grind, 1996). It should also convince us that visual stabilization and visual stability are two different but not independant problems that must be solved by oculomotor and perceptual systems, respectively.

5 Conclusion

In this review, we have emphasized the importance of eye movements studies to discover functional aspects of the various stages of motion processing in human and non-human primates. Since oculomotor behaviors are not independent from perceptual and attentional processes, we have indicated the need to understand the interplay between the two systems. We suggested that some shared representation of the moving three-dimensional world might offer such an interface by which perception controls action and action enables perception. We pointed out that studying the relative needs and temporal dynamics of both system is then crucial within this framework. From that perspective, the scope for future work is clearly enormous and call for some major paradigm-breaking in both oculomotor and psychophysical research fields.

6 Résumé

Les organismes vivants se distinguent par leur capacité remarquable à détecter et à analyser l'information visuelle de mouvement afin de contrôler leur comportements. Chez les primates, plusieurs aires visuelles corticales sont organisées de façon hiérarchique pour mesurer localement le mouvement dans l'image et extraire les informations de niveaux supérieurs sur les mouvements de soi ou des objets. Ces traitements nous permettent notamment d'exploiter les flux visuels complexes qui résultent de nos déplacements dans un environnement tridimensionnel. Notamment, lors de ces déplacements nous pouvons re-

diriger notre regard vers un objet. La stabilisation visuelle des yeux sur cet objet nécessite un traitement complexe des composante de translation du flux visuel global dans lequel l'image de l'objet est immergée. En enregistrant les mouvements oculaires chez un sujet dont la tête est immobile, nous pouvons disséquer les différentes étapes du traitement visuel du mouvement et distinguer la part fonctionnelle des traitements automatiques de bas niveaux de celle des selections attentive et cognitive du mouvement à poursuivre. En enregistrant les phases précoces de l'initiation du mouvement de poursuite, nous pouvons isoler fonctionnellement les processus précoces de bas niveaux. En étudiant les changements graduels du comportement oculomoteur au cours du temps, en fonction du niveau d'attention et de l'état perceptif du sujet, nous pouvons comprendre la nature des relations dynamiques entre la perception et l'action. Dans ce cadre expérimental, nous pouvons commencer à ouvrir le boîte noire dans laquelle l'information visuelle a été maintenue pendant longtemps par tous les modèles sensorimoteurs existants. Ce peut être aussi un point de départ pour étudier les inter-relations complexes existantes entre la perception et l'action.

Acknowledgments. This work was supported by the Centre National de la Recherche Scientifique. GSM was supported by the Fondation pour la Recherche Médicale and the Société de Secours des Amis des Sciences. Part of this work was conducted in collaboration with Frederick A. Miles and Claudio Busetini at the Laboratory of Sensorimotor Research, National Eye Institute, NIH, Bethesda, USA).

Address correspondance to Dr. G.S. Masson, Centre de Recherche en Neurosciences Cognitives, CNRS UPR 9012, 31 Chemin Joseph Aiguier, F-13402 Marseille, France. Ph: (33) 491.164.314, Fax: (33) 491.774.969, email: masson@lnf.cnrs-mrs.fr, <http://vision2.cnrs-mrs.fr/>

7 References

- Andersen, G. (1989) Perception of three-dimensional structure from optic flow without locally smooth velocity. *Journal of Experimental Psychology. Human Perception & Performance* 15, 49-55
- Albright, T.D & Stoner, G.R. (1995) Visual motion perception. *Proceedings of the National Academy of Sciences* 92, 2344-2440
- Blake, A. & Yuille, A. (1992) *Active vision*. The MIT Press, Cambridge, MA, USA
- Braddick, O. (1993) Segmentation versus integration in visual motion processing. *Trends in Neuroscience* 16, 262-268
- Busetini, C., Masson, G.S. & Miles, F.A. (1996) A role for stereoscopic depth cues in the rapid visual stabilization of the eyes. *Nature* 380, 342-345
- Busetini, C., Masson, G.S. & Miles, F.A. (1997) Radial optic flow induces vergence eye movements at ultra-short latencies. *Nature* 390, 512-515
- Cornilleau-Perez, V. & Gielen, C.C.A.M. (1996) Interactions between self-motion and depth perception in the processing of optic flow. *Trends in Neuroscience* 19(5), 196-202

Ferrera, V.P. & Lisberger, S.G. (1995) Attention and target selection for smooth pursuit eye movements. *Journal of Neuroscience* 15, 7472-7484

Ferrera, V.P. & Lisberger, S.G. (1997) Neuronal responses in visual areas MT and MST during smooth pursuit target selection. *Journal of Neurophysiology* 78, 1433-1445

Gellman, R.S., Carl, J.R. & Miles, F.A. (1990) Short-latency ocular following responses in humans. *Visual Neuroscience* 5, 107-122

Gibson, J.J. (1950) *The Perception of the Visual World*. Houghton Mifflin

Groh, J.M., Born, R.T. & Newsome, W.T. (1997) How is a sensory map read-out ? Effects of microstimulations in visual area MT on saccades and smooth pursuit eye movements. *Journal of Neuroscience* 17, 4312-4330

Hildreth, E.C. & Royden, C.S. (1995) Motion perception. In Arbib, M.A. (Ed) *The handbook of brain theory and neural networks* The MIT Press, Cambridge MA. pp 585-588

Howard, I.P. & Rogers, B.J. (1996) *Binocular vision and stereopsis* Oxford Psychology Series N 29. Oxford University Press. London.

Howard, I.P. & Simpson, W.S. (1989) Human optokinetic nystagmus is linked to the stereoscopic system. *Experimental Brain Research* 78, 309-314

Jeannerod, M. (1995) Corollary discharge in visuomotor coordination. *The handbook of brain theory and neural networks* The MIT Press, Cambridge MA. pp 266-269

Kawano, K., Shidara, M., Watanabe, Y & Yamane, S. (1994) Neural activity in cortical area MST of alert monkey during ocular following responses. *Journal of Neurophysiology* 71, 2305-2324

Koenderink, J.J. (1986) Optic flow *Vision Research* 26, 161-180

Kowler, E.H. (1990) The role of visual and cognitive processes in the control of eye movement. In Kowler, E.H. (Eds) *Eye Movements and Their Role in Visual and Cognitive Processes*. Reviews of Oculomotor Research Vol.4, Elsevier, Amsterdam

Krauzlis, R.J. (1994) The visual drive for smooth eye movements. In Smith, A.T. & Snowden, R.J. (Eds) *Visual detection of motion* Academic Press, New York

Lee, C., Rohrer, W.H. & Sparks, D.L. (1988). Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature* 332, 357-360

Lisberger, S.G., Morris, E.J. & Tychsen, L. (1987) Visual motion processing and sensorimotor integration. *Annual Review of Neuroscience* 10, 97-129

Lisberger, S.G. & Ferrera, V.P. (1997) Vector averaging for smooth pursuit eye movements initiated by two moving targets in monkeys. *Journal of Neuroscience*, 17, 7490-7502

Mallot, H.A. (1997) Behavior-oriented approaches to cognition: theoretical perspectives. *Technical Report NM-050*. Max-Planck Institut für biologische Kybernetik.

Masson, G.S., Busetini, C. & Miles, F.A. (1997) Vergence eye movements in response to binocular disparity without the perception of depth. *Nature* 390, 512-515

Mateeff, S., Dimitrov, G. & Hohnsbein, J. (1995) Temporal thresholds and reaction time to changes in velocity of visual motion. *Vision Research* 35, 355-363

Mestre, D.R. & Masson, G.S. (1997a) Ocular responses to motion parallax stimuli: the role of perceptual and attentional factors. *Vision Research* 37, 1627-1641

Mestre, D.R. & Masson, G.S. (1997b) Differential velocity thresholds in transparent motion displays *Investigative Ophthalmology and Visual Sciences (ARVO abstracts)* 38, 75

Miles, F.A. (1993) The sensing of rotational and translational optic flow by the primate optokinetic system. In *Visual motion and its role in the stabilization of gaze* F.A. Miles & J. Wallman (Eds) pp. 393-403, New-York, Elsevier

Miles, F.A., Kawano, K. & Optican, L.M. (1986) Short-latency ocular following responses of monkey: I. Dependence on temporospatial properties of the visual input *Journal of Neurophysiology* 56, 1321-1354

Movshon, J.A. & Newsome, W.T. (1996) Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *Journal of Neuroscience* 16(23), 7733-7741

Movshon, J.A., Adelson, E.H., Gizzi, M.S. and Newsome, W.T. (1985) The analysis of moving visual patterns. In: *Pattern recognition mechanisms*. Chagas, C., Gattass, R. & Gross, C. (Eds) pp 117-151. New-York: Springer

Poggio, G.F. & Talbot, W.H. (1981) Mechanisms of static and dynamic stereopsis in foveal cortex. *Journal of Physiology (London)* 315, 469-492

Qian, N. & Andersen, R.A. (1994) Transparent motion perception as detection of unbalanced motion signals. *Journal of Neuroscience* 14, 7367-7380

Recanzone, G.H. & Wurtz, R.H. (1994) Responses of MT and MST neurons in macaque monkeys to objects moving in two directions. *Society for Neuroscience Abstracts* 20, 773

Rodman, H.R. & Albright, T.D. (1989) Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Experimental Brain Research* 75:53-64

Rogers, S. & Rogers, B.J. (1992) Visual and non-visual information disambiguate surfaces specified by motion parallax. *Perception and Psychophysics* 52, 446-452

Salzman, C.D. & Newsome, W.T. (1994) Neural mechanism for forming a perceptual decision. *Science* 264, 321-237

Schöner, G., Kopecz, K. & Erlagen, W. (1997) The dynamic neural field theory of motor programming: arm and eye movements. In Morasso, P.G. & Sanguineti, V. (Eds) *Self-organization, computational maps and motor control*. Psychology Series, Vol. 119, pp. 271-310, Elsevier. Amsterdam.

Simoncelli, E.P. (1993) Distributed representation and analysis of visual motion. PhD Thesis. M.I.T. Media Laboratory. Published as Technical Report 209, Media Lab, MIT

Smith, A.T. & Snowden, R.J. (1994) *Visual detection of motion* Academic Press, New York

Takemura, A., Inoue, Y., Kawano, K. & F.A. Miles (1997). Short-latency discharges in medial superior temporal area of alert monkey to sudden changes in horizontal disparity. *Society for Neuroscience Abstracts* 23, 608.4

Treue, S., Husain, M. & Andersen, R.A. (1991) Human perception of structure from motion. *Vision Research* 31, 59-75

Treue, S. & Maunsell, J.H.R. (1996) Attentional modulation of visual motion processing in areas MT and MST *Nature* 382, 539-541

van Damme, W.J.M. & van de Grind, W.A. (1996) Non-visual information in structure-from-motion. *Vision Research* 36, 3119-3127

Waespe, W. & Henn, V. (1987) Gaze stabilization in the primate. The interaction of the vestibulo-ocular reflex, optokinetic nystagmus and smooth pursuit. *Reviews in Physiology, Biochemistry and Pharmacology* 106, 37-125

Watamaniuk, S.N.J. & Heinen, S.J. (1994) Smooth pursuit eye movements to dynamic random-dot stimuli. *Society for Neuroscience Abstracts* 20, 317

Wurtz, R.H. & Duffy, C.J. (1992) Neuronal correlates of optic flow stimulation. *Annals of the New York Academy of Sciences* 656, 205-219

Zeki, S. (1994) *A vision of the brain* Blackwell Scientific Publications. London

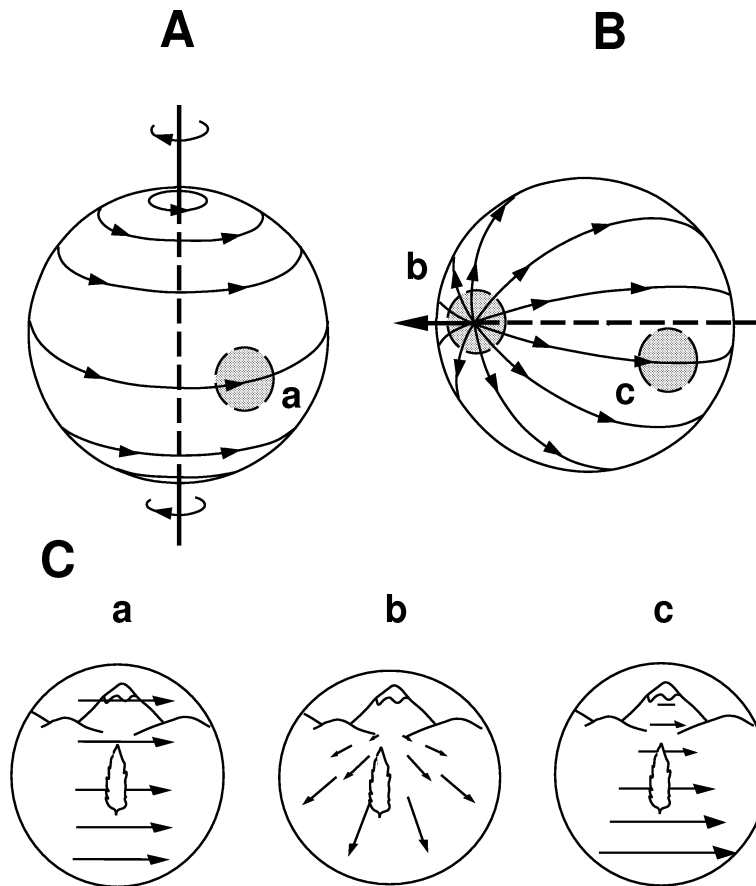


Figure 1: *Types of optic flow. [A] A rotational optic flow generated by rotation of the vantage point around the vertical axis. [B] A translational optic flow generated by translation of the vantage point along the horizontal axis. [C] Schematic illustration of the retinal flows when the moving observer maintains his/her fixation on the portion of the optic flow indicated by the grey patches in the top figures. For a pure rotation, the retinal flow is an homogeneous displacement of all the visual points, in the same direction and at the same speed. When the observer looks straight ahead while walking, the retinal flow resembles the velocity vector field illustrated in b. The retinal flow is a radial flow where dots move from the focus of expansion toward the limit of the field of view with an angular speed inversely related to their distance to the vantage point. In c, the moving observer maintains his/her gaze in a direction orthogonal to his own motion. The retinal flow consists in a lamellar flow with all vectors in the same direction but with velocity inversely related to the viewing distance. Such a lamellar flow is also called a motion parallax flow field.*

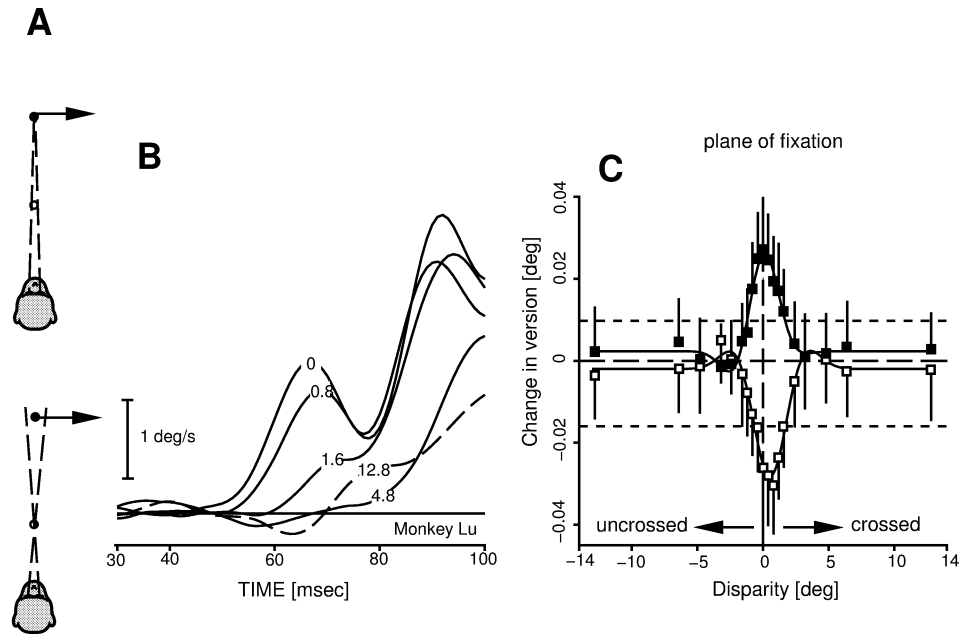


Figure 2: *Dependence of ocular following on the horizontal disparity of the tracked images. [A] When the monkey binocularly foveates the moving black dot, the two images of this object fall on similar points of the two retinæ: 0 disparity. The two images of the open dot, located in front of the plane of fixation, have a given disparity. In the bottom panel, the monkey fixates the open dot which now defines the plane of fixation and the motion stimulus is presented with a given (uncrossed) disparity. [B] Initial velocity profiles of ocular following responses in one monkey when the rightward moving textured visual scene is presented in the plane of fixation or in front of the plane of fixation. Numbers indicate the crossed disparity value. Responses are maximal for motion stimuli positioned in the depth-vicinity of the plane of fixation and get smaller as disparity increases. [C] Relationships between the change in version position over the 60-77 msec time window and the disparity of the textured scene. Closed symbols and open symbols indicate rightward and leftward motion, respectively. Horizontal dotted lines indicate the change in version evoked by a monocular motion of same velocity.*

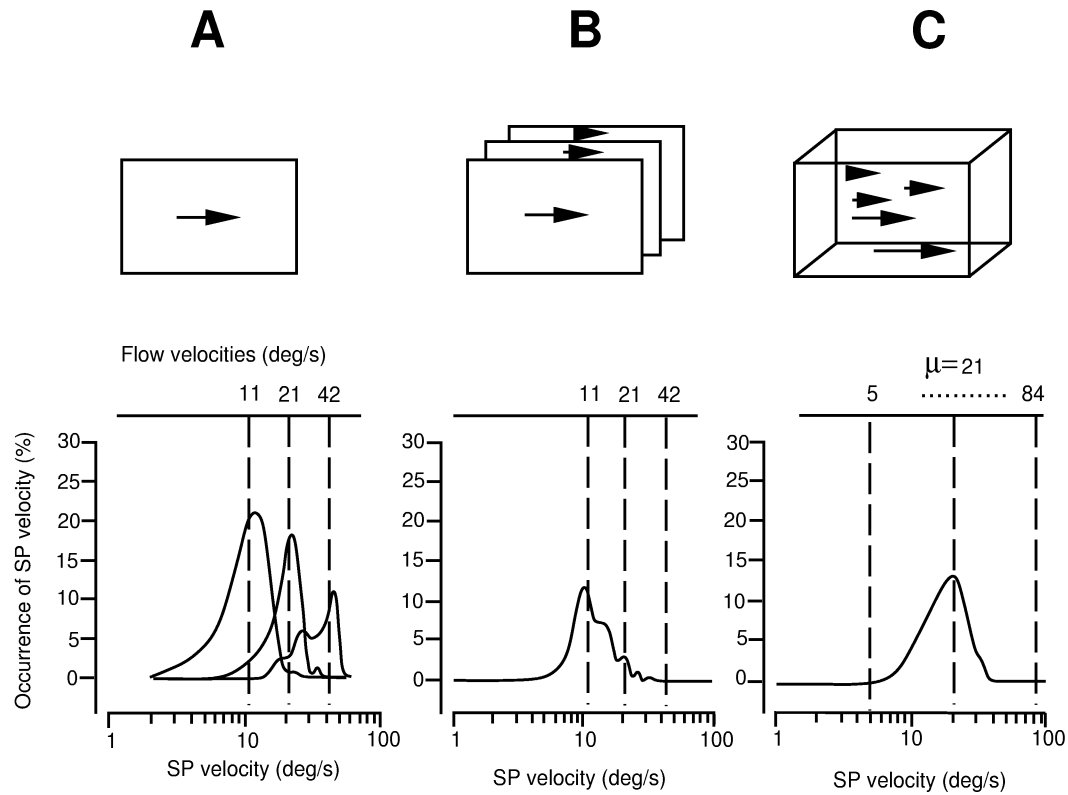


Figure 3: *Dependence of the optokinetic nystagmus on the complex flow field structure. Plots illustrate distributions of the occurrence of slow phase (SP) velocity (expressed as a percentage of the total number of slow phases, averaged across subjects), as a function of SP velocity. In [A], subjects faced three examples of single-velocity flows. In [B], subjects face a triple-velocity flow, with velocity components similar to the three conditions illustrated in A. In [C], subjects stare at a complex flow field where more than 10 different velocities are randomly distributed over the moving dots.*