

## LOCALIZING OBJECTS IN SPACE AND TIME: A COMPUTATIONAL MODEL BASED ON A PREDICTIVE MECHANISM

Alessandro Carlini<sup>1,2</sup>, Rossana Actis-Grosso<sup>2</sup>, Thierry Pozzo<sup>1,3</sup>

<sup>1</sup> INSERM-U887 Motricité-Plasticité, Université de Bourgogne, Dijon, France

<sup>2</sup> Department of Psychology, University of Milano-Bicocca, Milano, Italy

<sup>3</sup> Istituto Italiano di Tecnologia, I.I.T., Genova, Italy

E-mail to: [alessandro.carlini@u-bourgogne.fr](mailto:alessandro.carlini@u-bourgogne.fr)

### Abstract

*A computational model is presented, which can account for spatial mislocalizations. In fact, there are small but consistent errors in position judgement for a moving object at both its starting point (both in- and opposite to- the direction of motion) and vanishing point (in the direction of motion). Our model is based on motion extrapolation, that we consider as a process, wherein each step of information processing contributes in minimizing the spatial error due to the neural delay. The model presented here mimics the first step of the extrapolation process, where an early mechanism is supposed to work at a very first level of the information processing, (i.e. at retinal and sub-cortical level). Our model can reproduce all the SP and VP mislocalizations, as a consequence of its described strategy, as well as their dependence on velocity.*

A wide body of research has focused the interest on the perceived position of a moving stimulus: there are small but consistent errors (spatial mislocations), when a position judgement is required for a moving object at both its starting point (SP) and vanishing point (VP) or with respect to a flash aligned with a moving object (Flash-Lag effect, FLE, for which a flashed object in the same location of a moving stimulus is perceived as lagging behind the moving stimulus, e.g. Nijhawan, 1994; Müsseler & Aschersleben, 1998; Tornton, 2002; Kerzel & Gegenfurtner, 2003; ; Actis-Grosso & Stucchi, 2003; Baldo, Caticha, 2005).

For VP, a position shifted forward is usually indicated (Representational Momentum, RepMo), whereas for SP two conflicting spatial mislocations are reported (i.e., one in the direction of motion and the other in the direction opposite to motion, Fröhlich effect – FE - and backward mislocation – BM – respectively). The magnitude of all these errors is depending on velocity.

Several accounts for these mislocations have been proposed (Actis-Grosso, Stucchi, 2003; Kerzel, Gegenfurtner, 2003; Müsseler, Aschersleben, 1998; Nijhawan, 1994). Among them, Nijhawan (1994) proposed an account for the FLE, hypothesising the existence of a mechanism in the visual system that extrapolates the trajectory of moving objects and compensates for the delays due to neural transmission, from the time the visual input reaches the retina until it is processed by the CNS and thus perceived. In other words, according with Nijhawan (1994), the visual cortex uses the delayed visual data from the eye to extrapolate the trajectory of a moving object, so that it is perceived at its actual location. A similar account has been proposed to be responsible for mislocations observed at SP and VP (Actis-Grosso and Stucchi, 2003).

Recently several neural and computational models have been put forward for the localization of moving objects, mainly based on feedforward and feedback connections (e.g Erlhagen 2003, Baldo e Caticha 2005, Eagleman and Sejnowsky 2007). However, all these models are capable to explain only the “forwards” errors (i.e. the errors in the direction of motion), since

they give outcomes which mimics only the displacement of FLE, RepMo and FE, as reported in literature (e.g. Hubbard & Bharucha, 1988; Müsseler & Aschersleben, 1998). At present, no model has been presented, which can account for “backwards” errors (i.e. errors in the direction opposite to motion), such as the BM observed at motion SP.

Here we present a computational model which can account for both forwards and backwards localization errors. In our view the compensation of the neural delay (i.e. the extrapolation) should be viewed as a process, wherein each step of information processing contributes in minimizing the spatial error due to the neural delay. The extrapolation process should thus be regarded as a hierarchy of systems – or mechanisms – whose aim is not only to “put on-line” the sensory data (i.e. to compensate for the neural delay), but also to predict the most likely “future” position of the moving object. Cavanagh (1997) dubbed motion extrapolation “predicting the present”: in our view motion extrapolation process is also aimed at “predicting the immediate future”. Berry, Brivanlou, Jordan, and Meister (1999) reported anticipatory responses to motion in the retinal ganglion cells of the rabbit and the salamander. When retinal ganglion cell responses to moving and flashed stimuli were measured, it was observed that cell responses to moving stimuli were spatially shifted in the direction of motion, analogous to the forward shift of moving stimuli in relation to flashed stimuli shown in psychophysical experiments in humans (Nijhawan, 1994). However, a possible retinal mechanism would appear to be inadequate to explain spatial mislocations such as the flash-lag (Nieman et al, 2006), which has been demonstrated under diverse stimuli conditions and manipulations, and also in other perceptual domains, such as acoustical (Alais & Burr, 2006) and “motor” (i.e. in the dark, Nijhawan & Kirschfeld, 2003).

In our model, a retinal mechanism would not account for spatial mislocations, but is responsible to start the extrapolation process, broadly compensating for the neural delay on the basis of motion velocity and kinematics. In this way sensory data are already “corrected” when they arrive at the cortical areas, where the prediction (i.e. both of the present and of the “future”) error could be minimized. In this perspective a retinal mechanism is just the first step of the extrapolation process.

In this paper we present the computational model of a supposed retinal – and possibly sub-cortical – mechanism that accounts for spatial mislocations as observed at SP and partially accounts for RepMo and FLE; we hypothesise that the extrapolation process would be complete at a cortical level, where other steps of the process contribute in a substantial way to minimize the error, thus fully explaining FLE and RepMo, which occur in a later stage of motion extrapolation.

### **The model**

The model we present mimics the functioning of a supposed mechanism, which starts the visual extrapolation process. This mechanism is supposed to work at a very first level of the information processing, i.e. at retinal and sub-cortical level. At this stage, the aim of the mechanism is to detect the presence of a moving stimulus and to broadly forecast its present/future position. A further step is also implemented, in which the precision of the forecasted position is verified. We hypothesize that these two steps could be performed at the retinal level and at a sub-cortical level, possibly in the SC (Superior Colliculus).

The model has been developed on Matlab R2008a, using the Image Processing Toolbox ver. 6.1, and implemented on an Intel dual-core PC P9400 computer, 2.4 GHz CPU and 4 GB RAM. The model works on an “animation” constituted by a 100Hz snapshot sequence of 989x796 pixel images. Each animation presents a white dot stimulus on dark background, moving with different laws of motion and along different trajectories.

At present we limit our model to foveal vision (and for this reason both visual attention and eye-movements are not included) and to high luminance contrast (i.e. a white dot on a black background, corresponding to a lightness contrast, calculated according to the Weber function, of approximately 7,66).

For the detection of a moving stimulus, the mechanism uses a strategy based on a saliency-map approach (Itti, Koch, Niebur, 1998). This implies that when a supra-threshold luminance level is detected at a certain position in the visual field, the mechanism virtually starts its “prediction process”. What the mechanism does is to detect the presence of a stimulus *in* a certain position *with* a certain velocity. According to both Uchiyama, Goto, Matsunobu (2001) and Pearlman, Hughes (1976) it is likely that the human eye concurrently codifies both position and velocity of a target already at a retinal level, in which ON-OFF cells encode more specific stimuli than simply general movement: ON-OFF sequence of light intensity change are encoded with a spike pair with an interval of approximately 20 ms, “indicating that temporal coding is utilized in the vertebrate visual system as early as the retina” (Uchiyama, et al., p. 611). The *ad hoc* strategy we implemented to reproduce the physiological encoding of velocity in the model is to compute the velocity of the target by comparing the positions occupied in two successive frames (i.e. separated by an interval of 10 ms) of a given animation. In this way the mechanism has information on both the position and the velocity of the target. According to Berry, Brivanlou, Jordan and Meister (1999), “rather than lagging behind the visual image, the population activity (of retinal ganglion cells) travels near the leading edge of the moving bar” (p. 334).

In our model this anticipation is made by means of a code reproducing the supposed subcortical process, based on position and velocity information coming from the retina. The resulting value (an “internal variable” of the system) represents the “knowledge of the actual/future position” of the moving target. The value of this variable is crucial for our model: we obtained it by reviewing several papers reporting spatial displacements on both SP and VP (e.g. Kerzel & Gegenfurtner, 2003; Thornton, 2002; Hubbard & Bharucha, 1988; etc).

Target velocity is one of the main differences between the studies that found a BM or FE at SP: BM is typically observed with lower velocities. Given that at this stage of the process velocity and size are the only information available, it is likely that the computation of the internal variable, which triggers the extrapolation, is based on velocity. We look for a rule describing the relation between the perceptual displacement at the starting point and the target velocity.

This relation is depicted by Equation 1 and draw on the xy plane (Figure 1) the same behaviour described for the SP: depending on target velocity (x-axis), positive y-values for the function denote a backward dislocation, and negative values a forward dislocation (Fröhlich effect). As evident in the chart of Figure 1, the relation between error and velocity described by Equation 1 is similar to that found in Berry, Brivanlou, Jordan and Meister (1999, figure 5b, p. 337), but here velocity has opposite sign.

To start the extrapolation process, the mechanism applies to the first moving object position a transmission time delay of ~ 40 ms, under open-loop condition; a second time delay (~ 40 ms) is necessary for the mechanism to have a feedback on its “prediction”. We choose this value (i.e. 40 ms) on the basis of the value of the human’s eye capture rate, as reported in the literature Schlag and Schlag, 2002; Meister and Berry (1999); Spering and Gegenfurtner (2008), Krauzlis and Lisberger (1994). However, there is no a general agreement on this value, due to its possible relation with, for example, the environmental interaction (Demb, 2002) or ambient luminance, (Schlag and Schlag, 2002); furthermore the complexity of the retinal signal processing makes difficult to precisely calculate this value.

After 40 ms the model gather a new frame-image and, if the target is still detected at a supra-

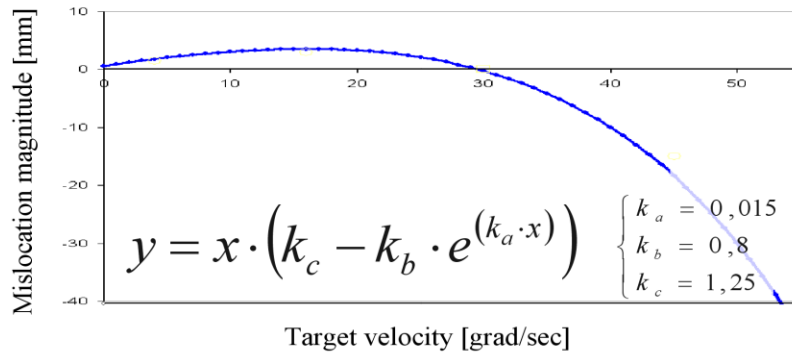


Figure 1 & Equation 1 – Relation between target velocity and its perceived position at the Starting Point, and coefficient value for the best fitting. Positive y-values are anticipatory mislocation at SP; x-axis intercept at 30 grad/s.

threshold luminance value and in a different position, the new information on position and velocity are compared with the previous ones and the internal variable is generated. From this moment on the sensory information is corrected on the basis of this internal variable. After further 40 ms the “corrected” position is compared with the new information: only at this stage the mechanism has a feedback on its forecast, thus is at this stage that the mechanism could correct a possible position error by modifying the internal variable. Thus, the generation of the internal variable (and the consequent verification procedure) is possible only when two successive positions have been detected.

This implies that the first perceived position is affected by the misperception and not corrected. The resulting mislocalization is variable in magnitude and direction (BM or FE) in relation to the target velocity, in accord to the Equation 1.

Things are different for VP: the mechanism has a feedback regarding the end of motion only after 40 ms, which always entails a mislocalization forward.

This “prediction – verification – correction – prediction” process constitutes a –fast– functional loop effectively supporting the tracking task.

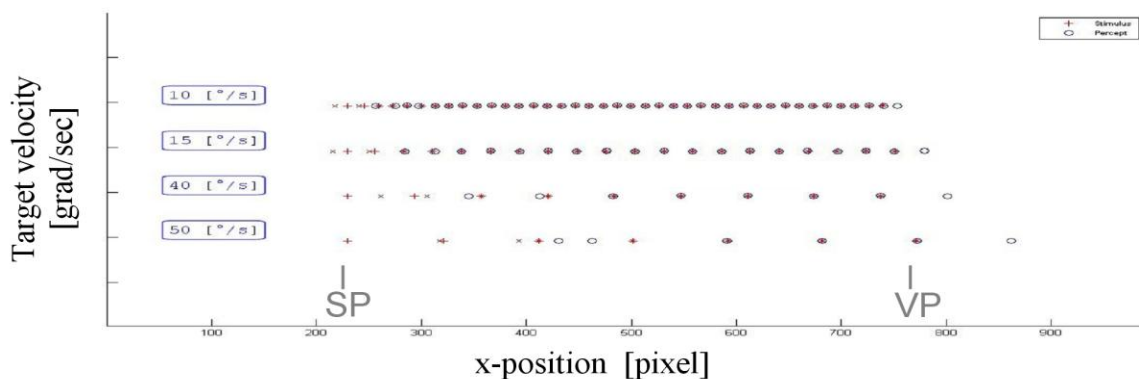


Figure 2 - Typical outcome of the model, reproducing the internal reference to the target position (blu circle). The trace of the stimulus (red cross) allows us to appreciate when a dislocation is perceived. Tracking from the left (SP) to the right (VP); traces recorded each 40 ms. X-axis present displacement in mm, the four line presents results for four different stimulus velocity (10, 15, 40 and 50 grad/sec).

## Results

In our model the internal value of the target position is updated in consequence of every forecast verification (i.e. every 40 ms).

Figure 2 shows the time-history of this variable, for four target velocity condition.

The gap between the internal reference (blu circle) and the actual position (red cross) of the stimulus allow appreciating the presence of a mislocalization both at the SP and the VP.

According to the experimental evidences on the human's visual system, in our model the outcome is depending on stimulus velocity: target moving slower than 30 grad/sec produce backward mislocalization, faster than 30 grad/sec produce forward mislocalization and the magnitude of the mislocalization is directly related with velocity.

The magnitude of the outcome mislocalizations at both SP and VP has been compared with that of several psychophysical studies: for SP this value fits the range-distribution of "human" values, whereas for VP generates mislocalization equal or a little greater than the experimental ones.

## Future developments

At present the model can reproduce SP and VP mislocalization as a consequence of its forecasting strategy. Next challenges are the extension of the "tracking" capacity to the partially occluded trajectories, to reproduce the "inference" task in humans and the consequent different strategies for biological or not-biological law of motions.

Furthermore, the extension of the model to multiple targets will permit to mimic also the third mislocalization: the Flash Lag Effect.

## Acknowledgements

This research was supported by the French ANR, projet Tecsan, Simaction.

## References

- Actis-Grosso, R., Stucchi, N. (2003). Shifting the start: Backward mislocation of the initial position of a motion. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 675–691
- Alais, D., Burr, D. (2006). The "Flash-Lag" Effect Occurs in Audition and Cross-Modally *Current Biology*, Vol. 13, 59–63
- Baldo, M., Caticha, N. (2005). Computational neurobiology of the flash-lag effect. *Vision Research* 45, 2620–2630
- Berry, M.J., Brivanlou, I.H., Jordan, T.A., Meister, M. (1999). Anticipation of moving stimuli by the retina. *Nature*, 398, 334–338
- Berry, M.J., Meister, M. (1998). Refractoriness and neural precision. *J. Neurosci.* 18, 2200–2211.
- Berry, M.J., Warland, D.K., and Meister, M. (1997). The structure and precision of retinal spike trains. *Proc. Nat. Acad. Sci. USA* 94, 5411–5416.
- Cavanagh, P. (1997). Predicting the present. *Nature (News and Views)*, 386, 19–21

- Demb, J.B. (2002). Multiple Mechanisms for Contrast Adaptation in the Retina. *Neuron*, Vol. 36, 781–786
- Eagleman, D.M., Sejnowski, T.J. (2007). Motion signals bias localization judgments: A unified explanation for the flash-lag, flash-drag, flash-jump, and Fröhlich illusions. *Journal of Vision*, 7(4):3, 1–12.
- Erlhagen, W. (2003). Internal models for visual perception. *Biol. Cybern.* 88, 409–417
- Hubbard, T. L., Bharucha, J. J. (1988). Judged displacement in apparent vertical and horizontal motion. *Perception & Psychophysics*, 44, 211–221
- Hubbard, T.L., Müsseler, L., Aschersleben, G. (1998). Judged displacement in apparent vertical and horizontal motion. *Perception & Psychophysics*, 44 (3), 211-221
- Itti, L., Koch, C., Niebur, E. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on pattern analysis and Machine intelligence* Vol. 20, No. 11, 1254-1259
- Kerzel, D. (2000). Eye movements and visible persistence explain the mislocalization of the final position of a moving target. *Vision Research*, 40, 3703–3715
- Kerzel, D., Gegenfurtner, R. (2003). Spatial distortions and processing latencies in the onset repulsion and Fröhlich effects. *Current Biology*, 13, 1975–1978
- Krauzlis, R.J., Lisberger, S.G. (1994). Temporal properties of visual motion signals for the initiation of smooth pursuit eye movements in monkeys. *Journal of Neurophysiology*, Vol. 72, No. 1, 150-162.
- Müsseler, J., Aschersleben, G. (1998). Localizing the first position of a moving stimulus: The Fröhlich effect and an attention-shifting explanation. *Perception & Psychophysics*, 60, 683–695.
- Müsseler, J., Kerzel, D. (2004). The trial context determines adjusted localization of stimuli: reconciling the Fröhlich and onset repulsion effects. *Vision Research*, 44, 2201–2206
- Nieman, D., Nijhawan, R., Khurana, B., Shimojo, S. (2006). Cyclopean Flash-lag illusion. *Vision Research* 46 , 3909–3914
- Nijhawan, R. (1994). Motion extrapolation in catching. *Nature*, vol 370, 1994.
- Nijhawan, R., Kirschfeld, K. (2003). Analogous Mechanisms Compensate for Neural Delays in the Sensory and the Motor Pathways: Evidence from motor Flash-Lag. *Current Biology*, Vol. 13, 749–753
- Pearlman, A.L., Hughes, C.P. (1976). Functional role of efferents to the avian retina. Analysis of ganglion cell receptive fields. *J. Comp. Neurol.* 166, 111-112
- Runeson, S. (1974). Constant velocity - not perceived as such. *Psychological Research*, 37, 3–23
- Spering, M., Gegenfurtner, K.R. (2008). Contextual effects on motion perception and smooth pursuit eye movements. *Brain Research* 1225, 76-85
- Schlag, J., Schlag-Rey, M. (2002). Through the eye, slowly: delays and localization errors in the visual system. *Nature Reviews*, vol 3, 191-200
- Thornton, I. (2002). The onset repulsion effect. *Spatial Vision*, Vol. 15, No. 2, 219–243
- Uchiyama, H., Goto, K., Matsunobu, H. (2001). ON-OFF retinal ganglion cells temporally encode OFF/ON sequence. *Neural Networks* 14, 611-615