

The Role of Internal Models and Prediction in Catching Balls

Mary Hayhoe, Neil Mennie, Brian Sullivan, & Keith Gorgos

Center for Visual Science
University of Rochester
Rochester, NY 14627, USA
mary@cvs.rochester.edu

Abstract

There is considerable evidence for the role of internal models of the body's dynamics in the control of movement. However, the existence of internal models of the environment is less well established. The present work provides further evidence of the existence of sophisticated internal models of the structure of the environment. We suggest that such models are used to predict upcoming events and plan movements in anticipation of those events. We recorded eye, head, and hand movements while subjects caught balls thrown with a bounce. Subjects initially fixate the hands of the thrower, then saccade to the anticipated bounce point, and then pursue the ball until it is close to the hands. However, ability to pursue the ball depends on experience with the ball's dynamic properties. When the ball was unexpectedly replaced with a more elastic ball, subjects were unable to track the ball, and instead made a series of saccades. Within 2 or 3 trials, subjects were once again able to accurately pursue the ball. Subjects displayed a different pattern of movements when throwing or watching other players. The observer's head movements from thrower towards the catcher often begin as much as half a sec before the ball leaves the thrower's hands. All these observations suggest that observers position their bodies in anticipation of expected events, in order to gather critical information. In addition, they suggest that observers maintain an internal model of the dynamic properties of the world, and rapidly update this model when errors occur.

Background

There is considerable evidence for the role of internal models of the body's dynamics in the control of movement (eg Wolpert et al, 1998). Such models predict the internal state of the body as a consequence of a planned movement, and help mitigate the problem of delays in sensory feedback about body posture. Subjects also appear to use internal models of the physical properties of objects in order to plan and control grasping (eg Johansson, 1996; Flanagan & Wing, 1998). The need for internal models of the environment is less well established.

In natural behavior, the retinal image is constantly changing because of movements of the eye and body. To build internal models of the visual environment, observers must be able accumulate visual information over the time-varying sequence of visual images resulting from these movements. However, the nature of the visual information preserved across different gaze positions is still poorly understood. Visual representations that span fixations are typically thought to be very impoverished. It is generally agreed that, following a change in gaze position, observers retain in memory only a small number of items, consistent with the capacity limits of visual working memory, together with information about scene "gist," and other higher level semantic information (Irwin & Gordon, 1998; Hollingworth & Henderson, 2002).

However, to understand just what information is retained across gaze positions, it seems necessary to consider the functional demands of vision in ordinary behavior. Some kind of internal model of the environment, such as memory for spatial structure, seems necessary to ensure coordinated movement (Loomis & Beall, 2004; Chun & Nakayama, 2000). Eye, head, and hand all need to act with respect to a common coordinate system and remain synchronized in time, across multiple actions. The reduction in temporal and spatial uncertainty afforded by the continuous presence of stimuli in ordinary behavior allows for the use of visual information acquired in fixations prior to the current one, to plan both eye and hand movements. We recently found evidence in support of this in a model-copying task in a virtual environment. Examination of eye movement targeting indicated that observers build a detailed representation of the spatial structure of the environment, that is precise enough to guide saccade programming (Aivar et al, 2005). Memory for spatial structure also influenced coordination patterns for eye, head, and hand movements, by allowing for earlier initiation of head and hand, which are much slower than the eye (Hayhoe et al, 2003). Evidence for the existence of models of the dynamic structure of the environment, in addition to the static structure, has been provided by the work of Zago et al (2004), who showed that observers use models of gravitational acceleration in the timing of manual interceptions with balls. The present work provides further evidence of the existence of sophisticated internal models of the structure of the environment. We suggest that such

models are used to predict upcoming events and plan eye movements in anticipation of those events.

The recent development of eye trackers mounted on the head has allowed the study of eye movements over extended tasks in natural settings, where a much wider variety of natural coordinated behaviors is possible (Land, 2004; Hayhoe & Ballard, 2005). In the context of such natural behaviors, eye movements are often made to a location in a scene in advance of an expected event. For example, Land and MacLeod (2000) measured eye movements in cricket, and found that batsmen anticipated the bounce point of the ball by a few hundred ms, and more skilled batsmen arrived at the bounce point about 100 ms earlier than less skilled players. These saccades were always preceded by a fixation on the ball as it left the bowler's hand, showing that batsmen use current sensory data in combination with learnt models of the ball's motion to predict the location of the bounce. This suggests that eye movement patterns are shaped by learnt internal models of the dynamic properties of the world.

In the present study, we followed up the observations of cricket to see if similar anticipatory eye movements are observed in conditions requiring less skilled performance.

Methods

We recorded eye, head, and hand movements while five subjects caught balls thrown with a bounce. In one condition, three participants stood in a triangular formation, separated by about 3m, and threw a ball around the circle. Monocular eye position of one of the participants was recorded using an ASL501 infra-red video-based eye tracker, mounted on the head. Calibration was performed on 9 points on the plane of the floor. Accuracy of the tracker is about 1 deg of visual angle. In addition to the data stream giving eye position in the head at 60 Hz, the eye-tracker also provides a 30 Hz video record of the scene from the observer's viewpoint, from a camera mounted on the head-band of the tracker. The video record is useful for analyzing the timing of the eye movements with respect to events in the scene, such as the point in time at which the ball bounces. Head and hand position were recorded at 60 Hz using three Polhemus Fastrack magnetic 6 degree-of-freedom position sensors, one attached to the head-band of the eye-tracker, and the others mounted on the back of each hand.

Initially, subjects threw a tennis ball ten times, in the same order, around the circle of three participants. Each throw was performed with a single bounce approximately mid-way between the participants. One of the throwers then changed the ball without warning, to one with greater elasticity (bounciness). This ball was then thrown for another 10 times.

The location and timing of the eye movements were analyzed frame-by-frame from the video records. (This is laborious, but more reliable than automated coding in these real-world settings.)

Results

Catching

Similar to batsmen in cricket, when catching a ball, subjects initially fixate the hands of the thrower, then saccade to the anticipated bounce point, and then pursue the ball until it is close to the hands. Average departure time of gaze from the hands of the thrower was 61 ms after the ball left the hands. (Between subjects standard error of the mean was 15 ms.) Gaze then arrives at a point a little above the anticipated bounce location an average of 53 ms before the bounce (30 ms SEM). Subjects maintain gaze at this location until the ball comes into the fovea, and then make a smooth pursuit movement, maintaining gaze on the ball until the catch. Since the minimum time to program a saccadic eye movement is 200-250 ms, the saccade from the hands to the bounce point must anticipate the release of the ball. We plotted the landing points of the saccades relative to the actual bounce point and found that they clustered within about 5 deg laterally, and about 15 deg vertically above the bounce point. This is shown in Figure 1. Thus subjects appear to be targeting a region just above the bounce point, rather than the bounce point itself. This

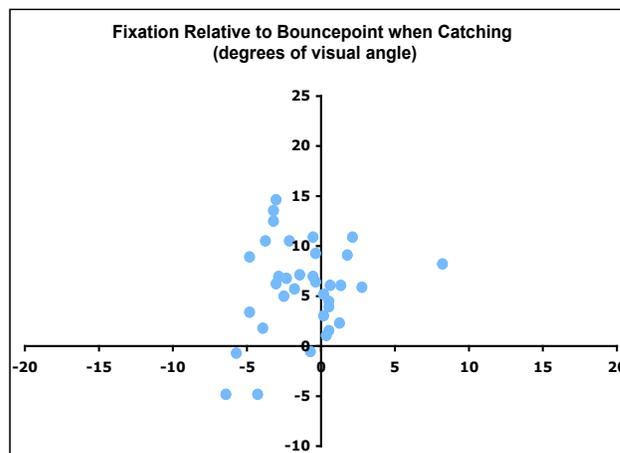


Figure 1. Scatter plot showing landing point of saccades anticipating the bounce for all subjects, plotted relative to the bounce point.

presumably facilitates the subsequent tracking movement by allowing time to capture the ball's trajectory after the bounce. The tight lateral clustering of the saccade landing points relative to the bounce point suggests that subjects are using information from the early part of the throw to target the likely location of the bounce.

Adjusting to the Ball's Dynamic Properties

Ability to pursue the ball depends on experience with the ball's dynamic properties. When the tennis ball was unexpectedly replaced with a bouncier ball, subjects were unable to track the ball, and instead made a series of saccades. Within a few trials, subjects were once again able to accurately pursue the ball. A crude evaluation of pursuit accuracy was made by measuring the proportion of time

gaze was less than two ball diameters away from the ball, in the period between bounce and catch. Improvement in pursuit performance over 6 trials is shown in Figure 2, which shows the proportion of time close to the ball improving rapidly over the first three trial, close to the performance level with the tennis ball. The ability to make accurate pursuit movements in this context therefore

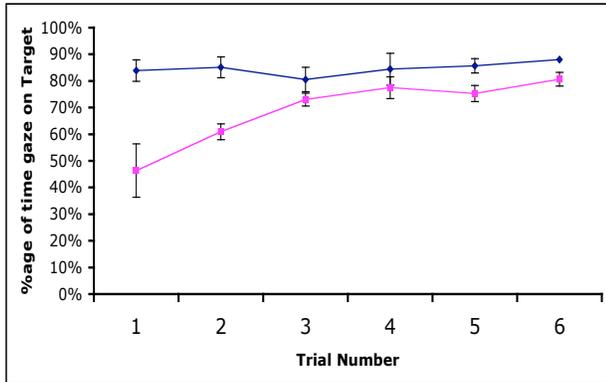


Figure 2. Pursuit performance as a function of trial number for the tennis ball (top curve) and more elastic ball (bottom curve). Error bars are +/- 1 SEM between subjects.

depends on knowledge of the dynamic properties of the new ball. The adjustment in performance is quite rapid, and uniform across subjects, suggesting that adjusting to such changes in the environment is an important feature of natural behavior. Note that two-dimensional retinal velocities were quite variable, and ranged between 50 and 80 deg per sec, so it is unlikely that subjects improved performance by simply increasing pursuit gain.

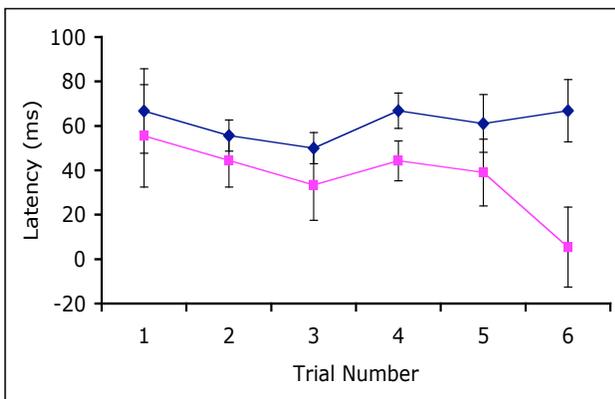


Figure 3 Arrival time of gaze at the bounce point, relative to the time of the bounce, as a function of trial number, for the tennis ball (top curve) and bouncier ball (bottom curve). Error bars are +/- 1 SEM between subjects.

The latency of the first saccade from hands to bounce point also changed over the course of a few trials. Figure 3 shows arrival time at the bounce point advancing by about

100 msec over the first 6 trials following the change from tennis to bouncy ball. The earlier arrival of the eye at the bounce point is accompanied by earlier departure from the hands at the point of release. This is shown in Figure 4. Thus subjects rapidly adjust timing of anticipatory movements to the new ball dynamics.

Watching

When subjects watched one player throwing the ball to the other player, they displayed a different pattern of

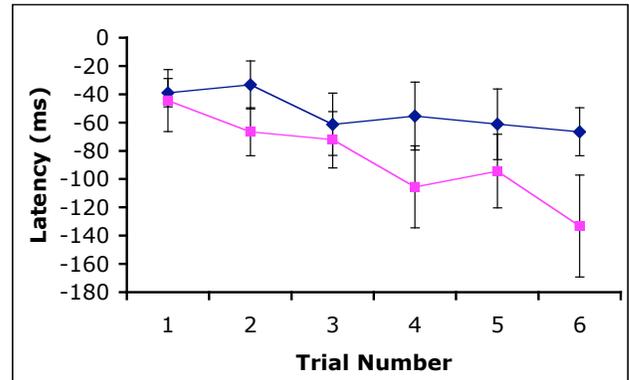


Figure 4. Departure time of gaze relative to release, as a function of trial number, for the tennis ball (top curve) and bouncier ball (bottom curve). Error bars are +/- 1 SEM between subjects.

movements, first making a saccade to a point near the bounce point, and then making a second saccade to the catcher's hands in anticipation of the catch. The timing of these movements was well in advance of the ball. Gaze departed the thrower's hands 51 ms before the release of the ball, and arrived at a point above the bounce point, 167 ms before the ball. Subjects then immediately made a saccade to the catcher's hands, arriving over 500 ms ahead of the ball. Head movements from thrower towards the catcher often begin as much as 500 ms before the ball leaves the thrower's hands. The location of the saccade in anticipation of the bounce was scattered widely in the region between thrower and catcher, as shown in Figure 5. Presumably, the exact location of the bounce is not critical to the subject as he/she is not catching the ball. Similarly, there is typically no smooth pursuit between bounce and catch. This suggests that the pursuit movement has some function in providing information about the ball's trajectory when the subject must catch the ball, possibly from the location of the eyes in the orbit during pursuit. On some trials, watchers pursued the ball part of the way between the bounce and the catch. This happened much more frequently with the bouncy ball: 83% vs 22% in first 3 trials. This suggests that subjects use watching as well as catching trials to learn the dynamic properties of bouncier ball.

When subjects are themselves throwing the ball, yet another pattern of fixations is observed. The subject fixates

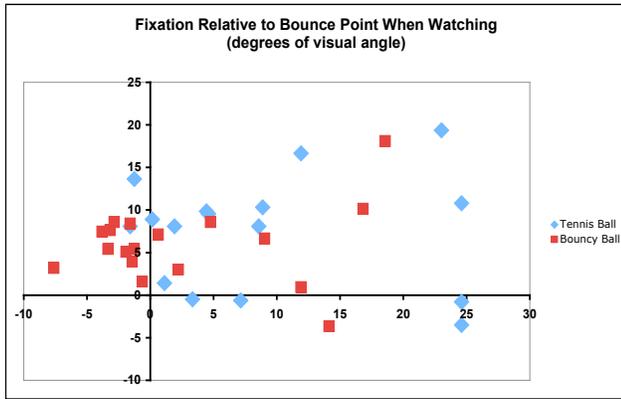


Figure 5. Scatterplot of saccade landing points in anticipation of the bounce, relative to the bounce point, for all subjects while watching a ball thrown from one player to another.

the ground mid-way between the bounce point and the catcher, in advance of the throw. Gaze is transferred to the catcher's hands after the bounce, arriving 330 ms before the ball.

Hand Movements

Examination of the hand movements also revealed anticipation. The hands start to move together, up and outwards from the body, in preparation for the catch, just as the ball bounces or about 100ms before. This is illustrated in Figure 6. The decrease in separation between the hands in preparation for the catch, shown in the red curve, begins a little later, about 100-200 msec after the bounce. It is possible that this movement is also programmed before the bounce, but the timing is consistent with the fastest manual reaction times, so a reactive response to the bounce cannot be ruled out. It appears catching movements are synchronized to the timing of the ball's path between bounce and catch.

Conclusions

Retinal motion, stereo, and extra-retinal information from pursuit eye movements have all been implicated in catching balls (Oudejans et al, 1999; Rushton & Wann, 1999; Tresilian, 1999). We have demonstrated here that prediction is also important. This is consistent with Land & MacLeod's (2000) observations that prediction of the bounce point is important for intercepting the ball with the bat in cricket. In the present study, anticipatory saccades, head movements, and pursuit movements all reveal that acquisition of visual information is planned for a predicted state of the world. Such predictions must be based on a stored memory representation of some kind. The precision of the predictions reveals the quality of the information in the stored memory. The spatial and temporal precision of the anticipatory saccades, and the fine-tuning of these movements following a change in the ball's dynamic properties indicate that subjects have an accurate internal

model of the ball's spatio-temporal path. The different patterns observed for catching as opposed to throwing or watching reveal that precise timing of the eye movements was important for controlling the hands for catching.

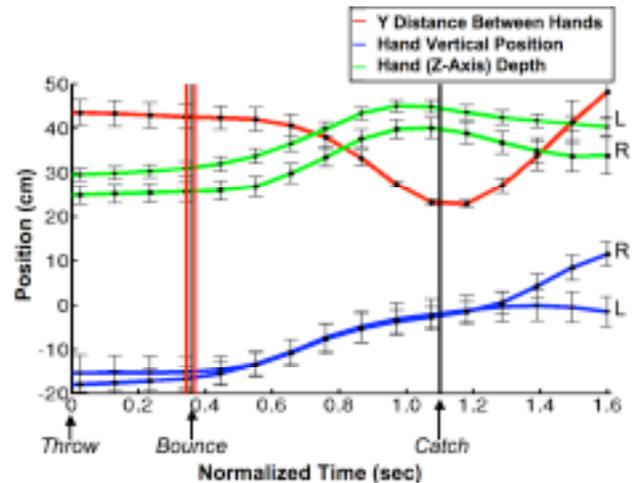


Figure 6. Hand position while catching. Zero indicates release of the ball, Trials were normalized to the mean duration between release and catch, The times of the bounce and catch are indicated by the vertical lines. The red curve shows the horizontal separation between the hands, the green curves show distance of the hands from the body, and the blue curve shows vertical hand position.

The present claim that subjects use internal models of the ball's motion to guide anticipatory movements is consistent with the findings of Zago et al (2004) who found that observers had an accurate internal model of the effects of gravity on the motion of a target that they had to manually intercept. Interestingly, when the (virtual) target failed to conform to gravitational acceleration, subjects adjusted their interception movements within five trials, comparable to the time course of learning observed in the present experiment. Zago et al argued that in their experiment subjects were not adjusting the internal model, but simply delaying the initiation of the hand movement. The ability of subjects to accurately pursue the new ball in the present experiment, despite variation in the 2D retinal velocity, suggests that in the present experiment, subjects are indeed learning an updated model, but we cannot argue this case strongly on the basis of the current data.

Since subjects were unable to track the ball on the first trial with the more elastic ball, prediction is necessary for smooth pursuit. Smooth pursuit eye movements are thought to be driven primarily by retinal velocity, although it has also been recognized that pursuit performance improves when the motion of the tracked object is predictable (Pola & Wyatt, 1991). The present results reveal that prediction is a critical component of naturally occurring pursuit movements. The trial-to-trial variability in retinal velocity, together with the overlap in the distributions of velocities for the two different balls, suggest that the adjustment of pursuit may be based on information about the expected

three-dimensional trajectory, rather than a simple increase in pursuit gain.

In summary, the observations in the present experiment suggest that observers maintain an internal model of the dynamic properties of the world, and rapidly update this model when errors occur. Rapid adjustment of performance suggests that such prediction is a ubiquitous feature of visually guided behavior.

Acknowledgements

Supported by NIH grants EY05729 and RR09283. Thanks to Jennifer Semrau for assistance with the experiments.

References

- Aivar, P., Hayhoe, M., & Mruczek, R. 2005 Role of spatial memory in saccadic targeting in natural tasks. *Journal of Vision*, 5, 177-193.
- Chun, M., & Nakayama, K. 2000 On the functional role of implicit visual memory for the adaptive deployment of attention across scenes. *Visual Cognition*, 7, 65-82.
- Flanagan, J. & Wing, A. 1997 The role of internal models in motion planning and control: evidence from grip force adjustments during movements of hand-held loads. *J. Neurosci* 17, 1519-1528,
- Hayhoe, M & Ballard, D. 2005 Eye Movements and Cognition. *Trends in Cognitive Science*, 9, 188-194.
- Hayhoe, M. et al. (2003) Visual memory and motor planning in a natural task. *J. Vis.* 3, 49–63.
- Hollingworth, A., & Henderson, J. M. 2002. Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 113-136.
- Irwin, D. & Gordon, R. 1998 Eye movements, attention, and trans-saccadic memory. *Visual Cognition*, 5, 127-155.
- Johansson, R. S. 1996 Sensory control of dextrous manipulation in humans. In: *Hand and Brain: the neurophysiology and psychology of hand movements.* (Wing, A, Haggard, P, & Flanagan, J eds), pp381-414. San Diego, Academic Press.
- Land, M. (2004) Eye movements in daily life. In Chalupa, L. and Werner, J., (eds) *The Visual Neurosciences*, 2, pp. 1357– 1368, MIT Press.
- Land, M.F. and McLeod, P. 2000 From eye movements to actions: how batsmen hit the ball. *Nat. Neurosci.* 3, 1340–1345.
- Loomis, J. & Beall, A. 2004 Model-based control of perception/action. In Vaina, L. et al (Eds) *Optic Flow and Beyond* Netherlands: Kluwer 421-441.
- Oudejans, R, Michaels, C, Davids, K, & Bakker, F. 1999 Shedding some light on catching in the dark: perceptual mechanisms for catching fly balls. *J Exp Psychol: Human Perception & Performance* 25, 531-542.
- Pola, J. & Wyatt, H. 1991 Smooth Pursuit: response characteristics, stimuli, and mechanisms. In Carpenter, R. (ed) *Eye Movements. Vision & Visual Dysfunction* 8, Boca Raton, CRC Press, 138-156.
- Rushton, S & Wann, J. 1999 Weighted combination of size and disparity: a computational model for timing a ball catch. *Nat Neurosci* 2, 186-190.
- Tresilian, J 1999 Visually timed action: time-out for “tau”? *Trends Cog Sci* 3, 301-310.
- Wolpert, D, Miall, C, & Kawato, M. 1998 Internal models in the cerebellum. *Trends in Cognitive Science*, 2, 338-347.
- Zago, M, Bosco, G, Maffei, V, Iosa, M, Ivanenko, Y, & Lacquaniti, F 2004 Internal models of target motion: expected dynamics overrides measured kinematics in timing manual interceptions. *J. Neurophysiol* 91, 1620-1634.