

Modelling mental navigation in scenes with multiple objects

Patrick Byrne Suzanna Becker

February 1, 2004

Abstract

Various lines of evidence indicate that animals process spatial information regarding object locations differently from spatial information regarding environmental boundaries/landmarks. Following Wang and Spelke's (2002) observation that spatial updating of egocentric representations appears to lie at the heart of many navigational tasks in many species including humans, we postulate a neural circuit which can support this computation in parietal cortex, assuming egocentric representations of multiple objects can be maintained in prefrontal cortex in spatial working memory (not simulated here). Our method is a generalization of an earlier model by Droulez and Berthoz (1991), with extensions to support observer rotation. We can thereby simulate perspective transformation of working memory representations of object coordinates based upon an egomotion signal presumed to be generated via mental navigation. This biologically plausible transformation would allow a subject to recall the locations of previously viewed objects from novel viewpoints reached via imagined, discontinuous, or disoriented displacement. Finally, we discuss how this model can account for a wide range of experimental findings regarding memory for object locations, and we present several predictions made by the model.

1 Introduction

Spatial reasoning is of paramount importance in nearly all aspects of human behaviour, from planning and navigating a complex route through some environment in order to reach a relevant goal, to simply grasping a nearby object. The set of all entities which comprise our spatial surroundings may be divided into subsets in arbitrarily many ways. For the present purposes, we will partition these entities into environmental boundaries/landmarks and objects. There is substantial empirical evidence as to how the brain represents and processes the former; for a review, see Burgess, Becker, King & O’Keefe (2001). For example, O’Keefe & Dostrovsky (1971) found neurons in the hippocampus of the rat which respond to the rat’s location in space. O’Keefe & Nadel (1978) argue that this collection of “place cells” forms a cognitive map and is the rat’s internal allocentric map of the environment. Evidence of view-invariant hippocampal place cells has also been found in non-human primates (Ono et al., 1993) and in human hippocampus (Ekstrom et al., 2003).

For the case of short-term object location representation, the focus of this paper, empirical evidence strongly indicates involvement of the posterior parietal cortex. For example, Sabes, Breznen & Andersen (2002) perform single unit recordings which demonstrate that area LIP of monkey cortex encodes saccade targets in retinotopic coordinates. More generally, Colby & Goldberg (1999) review evidence showing that object locations are represented in a variety of reference frames in parietal cortex, while Andersen, Shenoy, Snyder, Bradley & Crowell (1999) review evidence suggesting that area 7a incorporates vestibular information to maintain localization of visual stimuli in a world-centered reference frame. Finally, Goodale & Milner (1992) review data suggesting that the dorsal stream from striate to posterior parietal cortex is responsible for the relatively short time-scale sensorimotor transformations used while performing visually guided actions directed at objects in the environment.

In this paper we first review empirical evidence from neuropsychological and electrophysiological studies as to the nature and locus of object representations in the brain. We then present

a biologically plausible computational model which allows for the transformation of object coordinates maintained in spatial working memory (WM). The function of this transformation is to update WM representations of object coordinates in egocentric space while the subject mentally navigates through their environment. This transformation is driven by the products of mental navigation (presumably some mentally generated equivalent to vestibular, proprioceptive and motor efference information). Next, we perform simulations which demonstrate the functioning of the model and how error is introduced into object coordinate representations by the model. Finally, we discuss how the model can account for various experimental findings.

The representation of object location information in the brain appears, at least under certain circumstances, to be quite different from the representation of environmental boundary information. A series of experiments performed by Wang & Spelke (2000) provides insight into how humans differentially process object location and environmental boundary information. For each experiment, the authors placed a few objects around a small room. Subjects were allowed to explore the room and studied objects' locations for as long as they pleased. They were then brought to the center of the room, blindfolded, rotated by a small amount and asked to point to the various objects as they were called out in a random order by the experimenters. Following this, subjects were required to sit in a swivel chair fixed at the center of the room and disorient themselves via a 1 minute self-induced rotation. They were again asked to point to the objects in a predetermined random order. In addition to objects, subjects were also required to point to room corners in some experiments. The main findings can be summarized as follows: After the initial rotation (without disorientation), subjects could point to objects and room corners with relatively high accuracy, implying that they had encoded correctly the information which they were asked to. After disorientation, subjects could no longer accurately point to the location of objects or room corners. However, analyses of the consistency of the pointing errors indicated that the relative configuration of the room corners could be accurately recalled (for non-rectangular as well as rectangular rooms), whereas the relative object configuration could not be. This supports a configural representation of geometric feature locations, and

non-configural or independent representations of object locations. Furthermore, when subjects were reoriented with a bright light (still blindfolded), they could accurately recall the absolute and relative locations of room corners, but neither the relative nor absolute locations of objects. On the other hand, when the bright light was left on throughout the disorientation procedure, subjects could recall the relative and absolute locations of both objects and room corners.

The results of the Wang & Spelke (2000) experiments indicate that subjects could rapidly form an accurate internal representation of environmental boundaries and object locations. However, it appears that during mental transformations of spatial information in trials where the subject is not continuously oriented, information about a given object location is updated independently of information regarding other objects and of the environment. In this way each object location is subject to independent transformation error. This is in contrast to environmental geometric cues such as room corners, which appear to be updated as a coherent whole.

Additional evidence that object location information is handled differently than environmental boundary information comes from an experiment by Shelton & McNamara (2001). Subjects were brought into a room with various objects placed at different locations inside, and allowed to observe the objects from various predetermined viewpoints. After leaving the room, they were asked to imagine standing at a given object facing a second object, and to point to where a third object is relative to themselves. Subjects performed best when their imagined viewpoint was aligned with the original viewpoint from which they observed the object configuration. This suggests the possibility that subjects are storing the object configuration in head-centered egocentric coordinates, and that they must transform it (introducing error) when they imagine observing it from other viewpoints. We say *head-centered* here because subjects were asked to imagine *facing* an object, which implies pointing their head towards it; subjects were not asked to constrain their gaze direction so neither retinal nor body-centered coordinates are implied. Another possible interpretation of Shelton & McNamara's results is that subjects form a view-based snapshot of the entire presentation scene which can be used for later matching. Wang & Spelke (2002) review evidence that humans do seem to make use of such snapshots, at least

under certain circumstances. However, we do not investigate this issue any further here.

In order to build a computational model of mental navigation which explains empirical results such as those of Wang & Spelke (2000) and Shelton & McNamara (2001), we first require a more complete understanding of how the brain represents objects. We next review empirical evidence as to the nature and locus of object representations in the brain.

2 Objects and Spatial Working Memory

A number of experiments suggest a transition from short-term spatial WM to long term memory representations of objects after several minutes of study time (Smith & Milner, 1989; Crane et al., 1995; Bohbot et al., 1998). The evidence provided by these studies suggests that medial temporal lobe structures are essential to location memory over periods of several minutes (≈ 4 minutes) or greater, but are less relevant over shorter time scales.

To investigate the nature of short-term object location memory, we consider the evidence from functional imaging and unit recording studies. In an fMRI experiment performed by Galati, Lobe, Vallar, Berthoz, Pizzamiglio & LeBihan (2000), subjects were required to report the location of a vertical bar flashed before them for 150 ms relative to their midsagittal plane. During this task, several frontal and parietal regions were more active on the location task, relative to a control color decision involving the same stimuli. Furthermore, Sala, Rämä & Courtney (2003) presented subjects with a sequence of flashes, and asked them to recall the location or identity of what was shown (picture of a house or a face) three flashes back. During location recall, fMRI scans revealed activation in the superior portion of the intra-parietal sulcus (IPS) and in the superior frontal sulcus, as well as other areas. During identity recall, activation was found in the inferior and medial frontal gyrus, as well as other areas. These observations indicate that areas of frontal and parietal cortices are of key importance in generating and maintaining internal representations of spatial locations, at least for short periods of time. The role of frontal cortical areas in this process is reviewed by Levy & Goldman-Rakic (2000) who

argue that the principle sulcus (area 46) plays a crucial role in spatial WM and that Walker’s areas 12 and 45 (the inferior convexity) play a crucial role in object identity WM. In particular, a unit recording study by Funahasi, Bruce & Goldman-Rakic (1989) showed that neurons in the principle sulcus of the monkey appear to code for egocentric spatial locations. Furthermore, a human study by Oliveri, Turriziani, Carlesimo, Koch, Tomaiuolo, Panella & Caltagirone (2001), requiring subjects to remember the position of a flash two steps back in a sequence, found that only when TMS was applied to the dorsolateral prefrontal cortex (DLPFC) was accuracy affected, although TMS applied to several different brain regions affected reaction times. These results suggest that egocentric representations of spatial locations are maintained in DLPFC. Unfortunately, the role of parietal cortical areas in spatial WM is somewhat less clear. It is known, however, that neurons in areas VIP and LIP of the IPS show receptive fields in head-centered coordinates (see Burgess, Jeffery & O’Keefe (1999) for an overview). Also, Chaffe & Goldman-Rakic (1998) showed that when monkeys are required to hold a target location in memory for a short delay before making a saccade to it, neurons in area 8a (near the principle sulcus of prefrontal cortex) and in area 7ip (near the IPS) become active and show temporally varying activation levels over the delay period. Neurons in both regions show spatial selectivity similar to that found in the principle sulcus of monkeys by Funahasi et al. (1989).

From the above evidence, it seems plausible that object locations are initially represented in parietal cortex and, if sufficient attention is allocated to them, then their locations and identities are maintained in WM in DLPFC. In particular, their locations are represented *egocentrically* (as Shelton & McNamara’s (2001) work might indicate) in one area of DLPFC, while maintenance of their identities involves a different area of DLPFC consistent with Goldman-Rakic’s hypothesis that the ventral/dorsal (what/where) distinction persists into the DLPFC. In order to mentally manipulate the positions of objects stored in WM, it seems reasonable to assume that a circuit involving areas of the parietal cortex (especially the IPS) would be involved, given this area’s involvement in spatial WM and its known ability to represent locations in multiple reference frames. In particular, we hypothesize that object coordinates are maintained in WM

egocentrically in DLPFC and manipulated and represented more transiently in the vicinity of the IPS.

3 Model

In this work, we develop a computational account of memory for object locations in the face of certain types viewer motion, specifically those in which the motion is imagined, or those in which the subject does not remain continuously oriented throughout. At a minimum, the model should provide an explanation of Shelton & McNamara’s (2001) finding that subjects more accurately recall object positions when asked to do so from a viewpoint in which they previously observed the object configuration, and of Wang & Spelke’s (2000) finding that disoriented subjects cannot accurately recall object configurations. One possible way in which subjects might perform either the Shelton & McNamara or the Wang & Spelke task is by mental navigation. In contrast to a simple mental rotation of the object array, mental navigation involves imagined egomotion. This entails making mental viewpoint transformations while simultaneously updating WM representations of egocentric object coordinates. More specifically, subjects could make a WM snapshot of object locations from a given viewpoint and, when asked to recall object locations from a new viewpoint, they could mentally navigate from the initial viewpoint to the new viewpoint and use the same motion signal driving mental navigation to simultaneously drive a transformation of egocentric object coordinates. In the case of the Wang & Spelke tasks, the new viewpoint would have to be estimated due to the disorientation procedure. Also, assuming a serial updating procedure in which the transformation just described must be repeated for each individual object allows for a possible explanation of the finding that object configurations could not be accurately recalled after disorientation and that recall errors were not systematic - indicating lack of a configurational representation. This will be discussed in more detail in the discussion section.

The present hypothesis requires neural circuits which can represent the environment allo-

centrically, to allow for mental navigation through the environment. Additional circuits are required to maintain egocentric object coordinates, and transform these coordinates, one object at a time, based upon the egomotion signal which drives mental navigation. To this end, models of real navigation based on internal allocentric cognitive maps have been developed for rats and humans (see Voicu (2003); Burgess, Donnett & O’Keefe (1997), for examples). Given the evidence that both real and imagined spatial tasks invoke nearly the same cortical circuitry (see Stippich, Ochmann & Sartor (2002); Ino, Inoue, Kage, Hirose, Kimura & Fukuyama (2002); Mellet, Bricogne, Tzourio-Mazoyer, Ghaëm, Petit, Zago, Etard, Berthoz, Mazoyer & Denis (2000); Kreiman, Koch & Fried (2000), for examples), these models of real navigation will be assumed to be applicable to mental navigation. We now require a neural circuit which performs the object location transformation.

To begin, we must decide how egocentric object coordinates are to be represented. Thelen, Schöner, Scheier & Smith (2001) and Compte, Brunel, Goldman-Rakic & Wang (2000) have created models of spatial WM which hold location as a bump of activity in a topographically organized neural circuit, in which each neuron represents a location or direction in egocentric space. Becker & Burgess (2001) model the parietal egocentric map in this way, and similarly the organism’s location in allocentric space is represented by a Gaussian bump of activity over an array of hippocampal place cells. Such bump attractor networks have been used by others in models of hippocampus, as well (Zhang, 1996; Samsonovich & McNaughton, 1997). Here also it will be assumed that the model contains a ‘main’ layer of neurons, each of which will be preassigned a unique location on a Cartesian grid covering head-centered egocentric space. We will represent object location in egocentric space as a Gaussian bump of activity in the main neuron layer. By definition, the person’s coordinates in this map are the origin and their orientation will be taken as facing along the positive y -axis. The existence of these ‘main’ layer neurons in parietal cortex is supported by the electrophysiological recordings of Chaffe & Goldman-Rakic (1998).

In addition to a representation of object locations, our model will be adapted from a model by

Droulez & Berthoz (1991), which could easily be implemented to handle translational movements of the observer, but which requires a non-trivial extension to handle rotational motion. To derive the model, first consider an observer standing at the origin, O , of some reference frame and facing along the positive y -axis (see figure 1). The position of an object at a point, P , will be denoted by the vector, \mathbf{r}_P . If the observer moves by an amount, \mathbf{r}_T , and rotates by an amount, θ , then we will call the new egocentric reference frame the primed frame. The position of the object w.r.t. this new frame is given by

$$\mathbf{r}'_p = \mathbf{r}_p - \mathbf{r}_T. \quad (1)$$

Reexpressed in terms of cartesian x, y components, this is

$$x'_p \mathbf{i}' + y'_p \mathbf{j}' = (x_p - x_T) \mathbf{i} + (y_p - y_T) \mathbf{j}, \quad (2)$$

where \mathbf{i} and \mathbf{j} are basis vectors oriented along the x and y axes, respectively. Making an appropriate change of basis on the right hand side yields

$$\begin{aligned} x'_p &= (x_p - x_T) \cos \theta + (y_p - y_T) \sin \theta \\ y'_p &= (y_p - y_T) \cos \theta - (x_p - x_T) \sin \theta. \end{aligned} \quad (3)$$

Next, we assume that the observer's viewpoint shift occurs over a time interval, $\Delta t \ll 1$, so that they are standing at O facing in the \mathbf{j} direction at time t , and at O' facing in the \mathbf{j}' direction at time $t + \Delta t$. Assuming that egocentric velocities vary smoothly enough so that for any time, $t' \in [t, t + \Delta t]$, we can write

$$\begin{aligned} v_x(t') &= v_x^0 + f_{v_x}(t'), \\ v_y(t') &= v_y^0 + f_{v_y}(t'), \\ \omega(t') &= \omega^0 + f_\omega(t'), \end{aligned} \quad (4)$$

where $v_x(t')$, $v_y(t')$ and $\omega(t')$ are the translational and rotational velocities of the observer as measured in the O frame, and the f functions vary by $O(\Delta t)$ over the interval and satisfy

$f(t' = t) = 0$, then we have

$$\begin{aligned} x'_p &= x_p - [v_x^0 - \omega^0 y_p] \Delta t + O(\Delta t^2) \\ y'_p &= y_p - [v_y^0 + \omega^0 x_p] \Delta t + O(\Delta t^2). \end{aligned} \quad (5)$$

Of course, the primed position variables are simply the egocentric object coordinates at time $t + \Delta t$, the unprimed position variables are simply the egocentric object coordinates at time t , and the velocity variables are the egocentric velocities at time t . Therefore, we can rewrite equations 5 as

$$\begin{aligned} x(t + \Delta t) &= x(t) - [v_x(t) - \omega(t)y(t)] \Delta t + O(\Delta t^2) \\ y(t + \Delta t) &= y(t) - [v_y(t) + \omega(t)x(t)] \Delta t + O(\Delta t^2), \end{aligned} \quad (6)$$

where $x(t)$ and $y(t)$ refer to the egocentric object coordinates at time t and $\{v_x(t), v_y(t), \omega(t)\}$ are the egocentrically measured velocities at time t .

We are interested in representing the egocentric location of an object by a bump of activity across a population of neurons moving around within the observer's egocentric reference frame over time. This activity can be expressed as a function, $A(x, y, t)$ over egocentric coordinates and time. Given the value of $A(x, y, t)$, the value of $A(x, y, t + \Delta t)$ can easily be found by applying the inverse of equations 6 to the arguments of A , as follows

$$\begin{aligned} A(x, y, t + \Delta t) &= \\ A(x + [v_x(t) - \omega(t)y] \Delta t + O(\Delta t^2), y + [v_y(t) + \omega(t)x] \Delta t + O(\Delta t^2), t). \end{aligned} \quad (7)$$

The RHS of this equation may be expanded using the second Mean Value Theorem. Such an expansion will be valid so long as the second derivatives of A and the coefficients of Δt remain small. Notice, however, that these coefficients actually depend on location, so for distant objects, more error will be introduced into the approximation. We are thus left with

$$\begin{aligned} A(x, y, t + \Delta t) &= \\ A(x, y, t) + [v_x(t) - \omega(t)y] \Delta t \frac{\partial A}{\partial x}(x, y, t) + [v_y(t) + \omega(t)x] \Delta t \frac{\partial A}{\partial y}(x, y, t) + O(\Delta t^2). \end{aligned} \quad (8)$$

Recall that space in this model is to be represented as a Cartesian grid of points, with a single main layer neuron allocated to each point. In an attempt to follow the notation of Droulez & Berthoz (1991), we note that each of these positions and, hence, the corresponding main layer neurons can be labelled with a positive integer, starting at 1 for the position (neuron) with the smallest x and y value and increasing with position in the direction of increasing x until one row is complete. The numbering continues on the row with the next lowest y value until the last position is reached (see figure 2). By doing this, equation 8 can be discretized as follows

$$A(x_i, y_i, t + \Delta t) \approx \sum_j (a_{ij} + b_{ij}v_x(t) + c_{ij}v_y(t) + d_{ij}\omega(t)) A(x_i + dx_{ij}, y_i + dy_{ij}, t), \quad (9)$$

where x_i and y_i are the x and y coordinates of the i th location and dx_{ij} and dy_{ij} are the distances from the i th location to the j th location along the x and y directions. The values of the a, b, c , and d coefficients are determined by the approximation used to calculate the gradients of A . For example, to approximate equation 8 using a centered difference rule for the partial derivatives, the coefficients should be selected as

$$\begin{aligned} a_{ij} &= \delta_{ij}, \\ b_{ij} &= \frac{\Delta t}{2dx_{ij}} (\delta_{i+1,j} - \delta_{i-1,j}) \quad , \quad c_{ij} = \frac{\Delta t}{2dy_{ij}} (\delta_{i+N_x,j} - \delta_{i-N_x,j}) \\ d_{ij} &= \frac{\Delta t}{2} \left[\frac{x_i}{dy_{ij}} (\delta_{i+N_x,j} - \delta_{i-N_x,j}) - \frac{y_i}{dx_{ij}} (\delta_{i+1,j} - \delta_{i-1,j}) \right] \end{aligned} \quad (10)$$

where δ_{ij} is the Kronecker delta function and N_x is the number of neurons spanning the x -direction.

Notice that equation 9 can be rewritten as

$$\begin{aligned} A^M(x_i, y_i, t + \Delta t) &= \sum_j a_{ij} A^M(x_i + dx_{ij}, y_i + dy_{ij}, t) + b_{ij} A^{V_x}(x_i + dx_{ij}, y_i + dy_{ij}, t) \\ &+ c_{ij} A^{V_y}(x_i + dx_{ij}, y_i + dy_{ij}, t) + d_{ij} A^\omega(x_i + dx_{ij}, y_i + dy_{ij}, t) \\ A^{V_x}(x_i, y_i, t) &= v_x(t) A^M(x_i, y_i, t) \\ A^{V_y}(x_i, y_i, t) &= v_y(t) A^M(x_i, y_i, t) \\ A^\omega(x_i, y_i, t) &= \omega(t) A^M(x_i, y_i, t), \end{aligned} \quad (11)$$

where we have relabelled $A \rightarrow A^M$. These equations can now be viewed as a neural network with four layers, and the parameters a, b, c , and d can be viewed as weights (see figure 2 for a schematic), each of which couples a spatial location at the input processing layer with another spatial location in the transformation layer. Notice, however, that these weights are modulated by a scalar velocity value (i.e. the network contains neurons which perform multiplicative operations on their inputs). Neurons with this ability have been postulated in models of visual cortex (Mel, 1993; Mel, 1994). It is also thought that neurons which show such gain modulation are important in performing coordinate transformations in parietal cortex (Salinas & Abbott, 1996; Andersen et al., 1985). Additionally, we are assuming that the signal used to drive the transformation is an egocentric velocity signal. This is reasonable because vestibular information is known to project to Brodman’s area 7 which borders portions of the IPS (Kawano et al., 1980; Kawano et al., 1984). Note that although we represent velocity as three scalar values, along the x, y and θ directions, the model could easily be extended to accommodate coarse coding of the velocity signal. Finally, we are assuming a “Cartesian” distribution of neurons over egocentric space. It seems unlikely that cortex is organized in such a way; however, this organization is convenient for the present purposes and we do not expect the behaviour of our model to be qualitatively changed by the specific choice of neural distribution so long as each region of egocentric space is represented by a sufficiently high neural density. Quantitatively, for a different neural-spatial map, we would expect more transformation error to be introduced in areas where space is more sparsely represented.

Satisfactory performance of our model can be attained using the weights defined by 10. However, further improvements may be made by applying a gradient descent algorithm to refine the connection strengths. To be specific, at each iteration of the algorithm a Gaussian bump of activity is injected into the main layer at a random location, random velocities are applied to the network, the activity is updated for one time step and the resulting main layer activation is compared to an exact calculation of what it should be. This comparison is used to calculate

the value of an error function given by

$$E(\{a_{ij}, b_{ij}, c_{ij}, d_{ij}\}) = \sum_i [A^M(x_i, y_i) - \mathcal{A}^M(x_i, y_i)]^2, \quad (12)$$

where \mathcal{A}^M is the target activation determined as described below. The gradient of this error function w.r.t. weights can be calculated analytically and the weight vector can thus be updated to reduce the value of the error.

To determine the target function, \mathcal{A}^M , we note that it too must be a Gaussian with the same width as the input Gaussian, but that it must be centered at a slightly shifted location. This location can be determined by dividing equations 6 by Δt and taking the limit as $\Delta t \rightarrow 0$. The result is a set of two exact differential equations for the egocentric coordinates of the peak of the Gaussian bump in time. The constant velocity solutions of these equations are

$$\begin{aligned} x(t) &= \left(x(0) + \frac{v_y}{\omega}\right) \cos(\omega t) + \left(y(0) - \frac{v_x}{\omega}\right) \sin(\omega t) - \frac{v_y}{\omega}, \\ y(t) &= \left(y(0) - \frac{v_x}{\omega}\right) \cos(\omega t) - \left(x(0) + \frac{v_y}{\omega}\right) \sin(\omega t) + \frac{v_x}{\omega}. \end{aligned} \quad (13)$$

As one final point regarding the training method, it should be noted that the gradient descent method used here can be made equivalent to a Hebbian-like mechanism by the addition of an extra layer of “input” neurons which compare main layer activity to an external target signal (see Droulez & Berthoz (1991)).

4 Simulations

In the simulations reported here, a network with 961 main layer neurons arranged in a 31×31 lattice in which each neuron represents a spatial area of one unit² was trained using the method described in the previous section. Each of the three transformation layers also contained 31×31 neurons. The width of each Gaussian activation bump injected into the main layer for training was chosen randomly from the interval $[2, 4]$ units (widths below 2 units result in large error due to the discrete nature of the lattice; large widths are not useful in representing localized objects), with the peak height always set to unity. Angular velocities were chosen randomly

from the interval $[-1, 1]$ radians/s, while translational velocities were chosen from the interval $[-3, 3]$ units/s. These ranges were chosen to be reasonable for a human subject in the case of 1 unit = 1 meter, but the exact values, which have little affect on training results, were chosen arbitrarily. A time step of 10 ms was used for training and for all simulations. With initial weights set to the values defined in equations 10, the simple gradient descent algorithm used here was found to converge after approximately 500,000 random infinitesimal transformations when the learning rate was set to 5×10^{-4} . Although error, as defined by equation 12, was reduced by a factor of about 4 during this procedure, all weights remained within 4% of their initial values after training was complete.

As a demonstration of the functioning of the model, a simulation was carried out in which the egocentric location of a stationary object was maintained while the subject performed a sequence of imagined motions in space (see figure 3). To begin, an object was placed 5 units away from the subject, directly to their right. The subject turned 90 degrees to their left, moved forward by 5 units and then turned 90 degrees to their right. The predicted trajectory of the stationary object through egocentric space, as computed by the model, can be seen in figure 4, where main layer activations, sampled every 20 time steps, have been superimposed. Initial and final activations are shown in figures 5 and 6.

From figure 6, we see that the bump of activity representing the object's egocentric coordinates is somewhat attenuated and deformed by the transformation procedure. In fact, error is introduced in the simulation via two distinct mechanisms. First, the truncated leading order terms in equation 6 contain factors of v_x^2, v_y^2 , and ω^2 , so simulation error should increase quadratically in velocity for large velocities (as long as $\text{velocity} \times \Delta t < 1$, after which point error will increase more rapidly with increasing velocity). Second, at low velocities, truncating the $O(\Delta t^2)$ terms will have little effect in terms of introducing error compared with the effect of approximating the first derivatives of activation using information from a discrete lattice of neurons. In this case, error should increase linearly with the number of time steps required for a given transformation, or equivalently, as the inverse of velocity. A number of simulations were

performed in order to investigate these effects. The results of two of these simulations will now be discussed. For the first, an object was placed 5 units to the egocentric right of an observer who then turned 90 degrees to their left. For the second, an object was placed 5 units to the egocentric left of an observer who then moved to their egocentric right a distance which is equal to the arc length traveled by the object in the first simulation. In both cases, the object was represented by a Gaussian activation bump with a width of 2 units. Squared error is calculated for the final activation using equation 12 and plotted against velocity in figures 7 and 8. In both cases, the optimum velocity is around 2 units/s. The optimum velocity actually remains in the [1.5, 3.5] units/s interval for a wide variety of paths. Finally, the solid curves in figures 7 and 8, which are of the form

$$f(x) = \frac{a_1}{x} + a_2 (x - a_3)^2 + a_4, \quad (14)$$

were fit to the data in order to demonstrate that transformation error depends on velocity in the way described above.

5 Discussion

In this paper, we have postulated a biologically plausible computation, implemented as a neural circuit, which explains how the contents of spatial WM can be updated to keep track of the egocentric coordinates of remembered locations during mental navigation over time scales shorter than 4 minutes. Error can be introduced into the representation of object location coordinates during transformation via a number of different mechanisms. We have demonstrated that the discrete nature of the representation of space and the exclusion of higher order velocity information in a rate-coded neural circuit of this kind leads to a non-zero error regardless of the velocity of the subject through space. This error appears to affect peak height predominantly, but the effect on final peak location is relatively small in the simulations performed here. There is, however, an optimum velocity range for performing transformations. Other potential sources of error are internal and external noise, which we have not studied here.

A final source of error in the kind of object location transformation postulated here is the potential error introduced by the mental navigation procedure itself. This process must depend on an internal allocentric map of the subject’s environment. Evidence indicates that this map contains inaccuracies and imprecisions (Hirtle & Jonides, 1985), perhaps moreso in novel environments (Hartley et al., 2000). Presumably, tasks requiring precise mental navigation would also suffer from these errors.

Shelton & McNamara’s (2001) finding that subjects more accurately remember object configurations when tested from viewpoints in which they have actually observed the object configuration is readily explained by our model. Subjects would simply store an egocentric view of the studied presentation in WM, and then mentally navigate to the test viewpoint while updating object locations using the proposed parietal circuit. Note that this circuit can manipulate and transiently represent information from spatial WM; however, we do not claim that the original representation of object location, maintained in DLPFC, is necessarily altered by this procedure. Therefore, the original WM snapshot information is still available for future transformation and/or retrieval. This implies that subjects are still more accurate when tested at a presentation viewpoint because no transformation is required and no error is introduced. Our model does not, however, address the question of why information from a particular presentation viewpoint is maintained in preference to others.

Our model also provides an explanation for the Wang & Spelke (2000) data. If we assume that subjects have an accurate WM snapshot of object locations before disorientation, then after disorientation they could mentally navigate from the original snapshot viewpoint to their new estimated viewpoint while simultaneously using the egomotion signal which drives mental navigation to transform egocentric object location coordinates. From the discussion above, we believe that the mental navigation/parietal update procedure would introduce error into an egocentric object location transformation. Since we have assumed serial update, i.e., the mental navigation procedure must be performed separately for each object, the independent error inflicted on each object coordinate transformation could generate an overall object configuration

error. This would also explain the object configuration error obtained from the experiment in which subjects were reoriented after disorientation by a bright light visible through their translucent blindfold. In this case a mental navigation still must be performed from the viewpoint of the subjects while facing the light directly to the new viewpoint.

Wang & Spelke’s (2000) results indicate that blindfolded subjects who are continuously oriented are able to maintain object configurations accurately even after rotation. Performance in these conditions, which include slow rotation, or rotation in the presence of an orienting light, suggest minimal transformation of internal object location representations, or transformation which induces a systematic error only. One possible explanation for how subjects could recall object locations under continuously oriented conditions is that they could generate a displacement/rotation vector via path integration. This vector could simply be added to the DLPFC maintained object location representations at each recall. In this way, all object locations would be subject to the same systematic error. Another possibility is that when the subject remains oriented, they make use of some allocentric system in which object locations are static. In this case they must simply orient themselves w.r.t. the configuration. Given the availability of path integration in these conditions, subjects would be able to accurately localize themselves within the environment and this procedure should only result in a small error which would be the same for all objects.

In addition to being consistent with a range of experimental findings regarding memory for object locations, our model makes two easily testable predictions. First, the serial updating we require for multiple objects suggests that a subject’s reaction time to determine if an object configuration, seen from a novel viewpoint, is the same as one seen from a different viewpoint, should monotonically increase with number of objects. Second, our hypothesis regarding how object configuration error is introduced into the Wang & Spelke (2000) task suggests that, as subjects become more familiar with a given environment, and hence develop a more precise cognitive map, their performance on these types of tasks should improve. These predictions are currently being investigated empirically.

Finally, another interesting question to investigate would be the nature of how objects become represented after being seen as stable for long periods of time, i.e., after several minutes. One way to investigate this would be to perform an experiment similar to that of Wang & Spelke (2000) in which the presentation phase was continued over a long period of time before testing began.

Acknowledgements

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to S.B. We thank Neil Burgess for helpful discussions.

References

- Andersen, R., Essick, G., & Siegel, R. (1985). The encoding of spatial location by posterior parietal neurons. *Science*, *230*:456–458.
- Andersen, R., Shenoy, K., Snyder, L., Bradley, D., & Crowell, J. (1999). The contributions of vestibular signals to the representations of space in the posterior parietal cortex. *Ann. N. Y. Acad. Sci.*, *871*:282–292.
- Becker, S. & Burges, N. (2001). A model of spatial recall, mental imagery and neglect. In T. Leen, T. Ditterich, & V. Tresp (Eds.), *Advances in neural information processing systems*, Vol. 13 (pp. 96–102). Cambridge, MA: MIT Press.
- Bohbot, V., Kalina, M., Stepankova, K., Spackova, N., Petrides, M., & Nadel, L. (1998). Spatial memory deficits in patients with lesions to the right hippocampus and to the right parahippocampal cortex. *Neuropsychologica*, *36*(11):1217–1238.

- Burgess, N., Becker, S., King, J., & O'Keefe, J. (2001). Memory for events and their spatial context: models and experiments. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 356(1413):1493–1503.
- Burgess, N., Donnett, J., & O'Keefe, J. (1997). The representation of space and the hippocampus in rats, robots and humans. *Z. Naturforsch [C]*, 53(7-8):504–509.
- Burgess, N., Jeffery, K., & O'Keefe, J. (Eds.) (1999). *The hippocampal and parietal foundations of spatial cognition.*, Chap. 1. Oxford: Oxford University Press.
- Chaffe, M. & Goldman-Rakic, P. (1998). Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *Journal of Neurophysiology*, 79(6):2919–2940.
- Colby, C. & Goldberg, M. (1999). Space and attention in parietal cortex. *Annual Rev. Neurosci.*, 22:319–349.
- Compte, A., Brunel, N., Goldman-Rakic, P., & Wang, X.-J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex*, 10:910–923.
- Crane, J., Milner, B., & Leonard, G. (1995). Spatial-array learning by patients with focal temporal-lobe excisions. *Society for Neuroscience Abstracts*, 21:1446.
- Droulez, J. & Berthoz, A. (1991). A neural network model of sensoritopic maps with predictive short-term memory properties. *Proc. Natl. Acad. Sci. USA*, 88:9635–9657.
- Ekstrom, A., Kahana, M., Caplan, J., Fields, T., Isham, E., Newman, E., & Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature*, 425:184–187.
- Funahashi, S., Bruce, C., & Goldman-Rakic, P. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, 61(2):331–348.

- Galati, G., Lobel, E., Vallar, G., Berthoz, A., Pizzamiglio, L., & LeBihan, D. (2000). The neural basis of egocentric and allocentric coding of space in humans: a functional magnetic resonance study. *Exp. Brain Res.*, *133*:156–164.
- Goodale, M. & Milner, A. (1992). Separate visual pathways for perception and action. *Trends Neurosci.*, *15*(1):20–25.
- Hartley, T., Burgess, N., Lever, C., Cacucci, F., & O’Keefe, J. (2000). Modelling place fields in terms of the cortical inputs to the hippocampus. *Hippocampus*, *10*(4):369–379.
- Hirtle, S. & Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Memory and Cognition*, *13*(3):208–217.
- Ino, T., Inoue, Y., Kage, M., Hirose, S., Kimura, T., & Fukuyama, H. (2002). Mental navigation in humans is processed in the anterior bank of the parieto-occipital sulcus. *Neuroscience Letters*, *322*:182–186.
- Kawano, K., Sasaki, M., & Yamashita, M. (1980). Vestibular input to visual tracking neurons in the posterior parietal association cortex of the monkey. *Neurosci Lett.*, *17*:55–60.
- Kawano, K., Sasaki, M., & Yamashita, M. (1984). Response properties of neurons in posterior parietal cortex of monkey during visual-vestibular stimulation. *J. Neurophysiol.*, *51*:340–351.
- Kreiman, G., Koch, C., & Fried, I. (2000). Imagery neurons in the human brain. *Nature*, *408*:357–361.
- Levy, R. & Goldman-Rakic, P. (2000). Segregation of working memory functions within the dorsolateral prefrontal cortex. *Exp. Brain Res.*, *133*:23–32.
- Mel, B. (1993). Synaptic integration in an excitable dendritic tree. *J. Neurophysiol.*, *70*(3):1086–1101.
- Mel, B. (1994). Information processing in dendritic trees. *Neural Computation*, *6*:1031–1085.

- Mellet, E., Bricogne, S., Tzourio-Mazoyer, N., Ghaëm, O., Petit, L., Zago, L., Etard, O., Berthoz, A., Mazoyer, B., & Denis, M. (2000). Neural correlates of topographic mental exploration: the impact of route versus survey perspective learning. *NeuroImage*, *12*:588–600.
- O’Keefe, J. & Dostrovsky, J. (1971). The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain Res.*, *34*(1):171–175.
- O’Keefe, J. & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Oxford University Press.
- Oliveri, M., Turriziani, P., Carlesimo, G., Koch, G., Tomaiuolo, F., Panella, M., & Caltagirone, C. (2001). Parieto-frontal interactions in visula-object and visual-spatial working memory: evidence from transcranial magnetic stimulation. *Cerebral Cortex*, *11*(7):606–618.
- Ono, T., Nakamura, K., Nishijo, H., & Eifuku, S. (1993). Monkey hippocampal neurons related to spatial and nonspatial functions. *Journal of Neurophysiology*, *70*(4):1516–1529.
- Sabes, P., Breznen, B., & Andersen, R. (2002). Parietal representation of object-based saccades. *J. Neurophysiol.*, *88*(4):1815–1829.
- Sala, J., Rämä, P., & Courtney, S. (2003). Functional topography of a distributed neural system for spatial and nonspatial information maintenance in working memory. *Neuropsychologica*, *41*:341–356.
- Salinas, E. & Abbott, L. (1996). A model of multiplicative neural responses in parietal cortex. *Proc. Natl. Acad. Sci. USA*, *93*:11956–11961.
- Samsonovich, A. & McNaughton, B. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. *J Neurosci.*, *17*(15):5900–5920.
- Shelton, A. & McNamara, T. (2001). Systems of spatial reference in human memory. *Cognitive Psychology*, *43*(4):274–310.

- Smith, M. & Milner, B. (1989). Right hippocampal impairment in the recall of spatial location: encoding deficit or rapid forgetting? *Neuropsychologica*, *27*(1):71–81.
- Stippich, C., Ochmann, H., & Sartor, K. (2002). Somatotopic mapping of the human primary sensorimotor cortex during motor imagery and motor execution by functional magnetic resonance imaging. *Neuroscience Letters*, *331*:50–54.
- Thelen, E., Schöner, G., Scheier, C., & Smith, L. (2001). The dynamics of embodiment: Adaptive field theory of infant perseverative reaching errors. *Behavioral and Brain Sciences*, *24*:1–86.
- Voicu, H. (2003). Hierarchical cognitive maps. *Neural Networks*, *16*:560–576.
- Wang, R. & Spelke, E. (2000). Updating egocentric representations in human navigation. *Cognition*, *77*:215–250.
- Wang, R. & Spelke, E. (2002). Human spatial representation: insights from animals. *Trends Cog. Neurosci.*, *6*(9):376–382.
- Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J Neurosci.*, *16*(6):2112–2126.

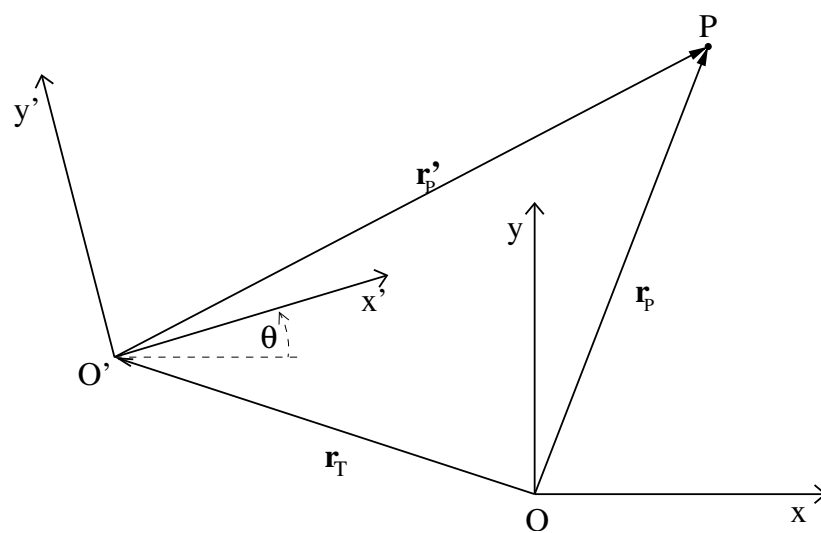


Figure 1: Byrne/ms2867

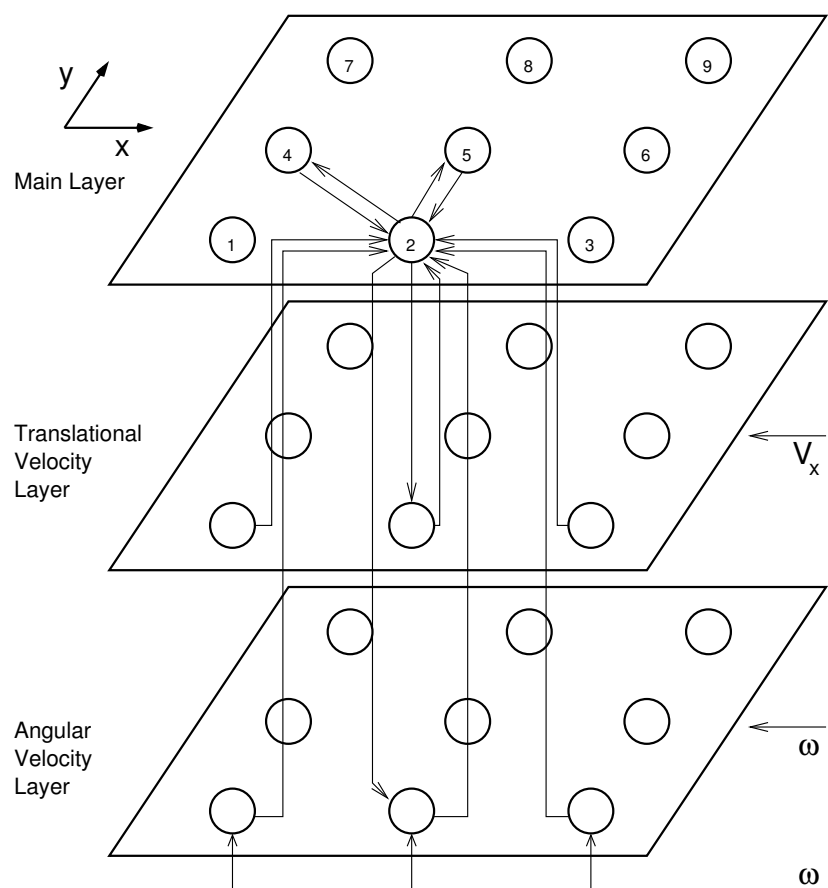


Figure 2: Byrne/ms2867

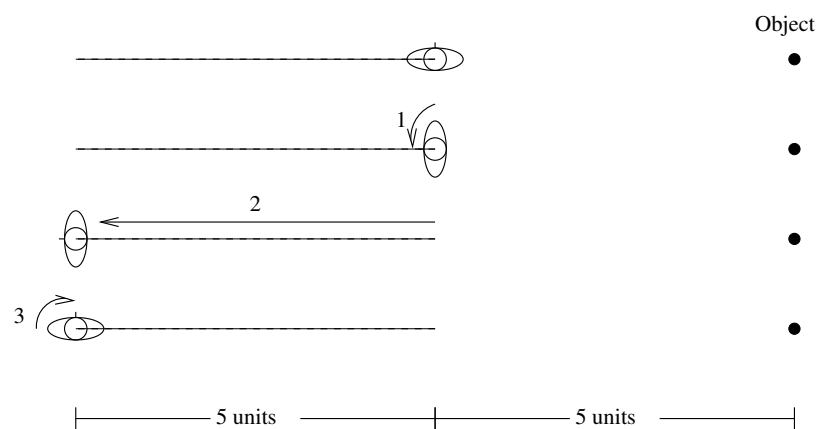


Figure 3: Byrne/ms2867

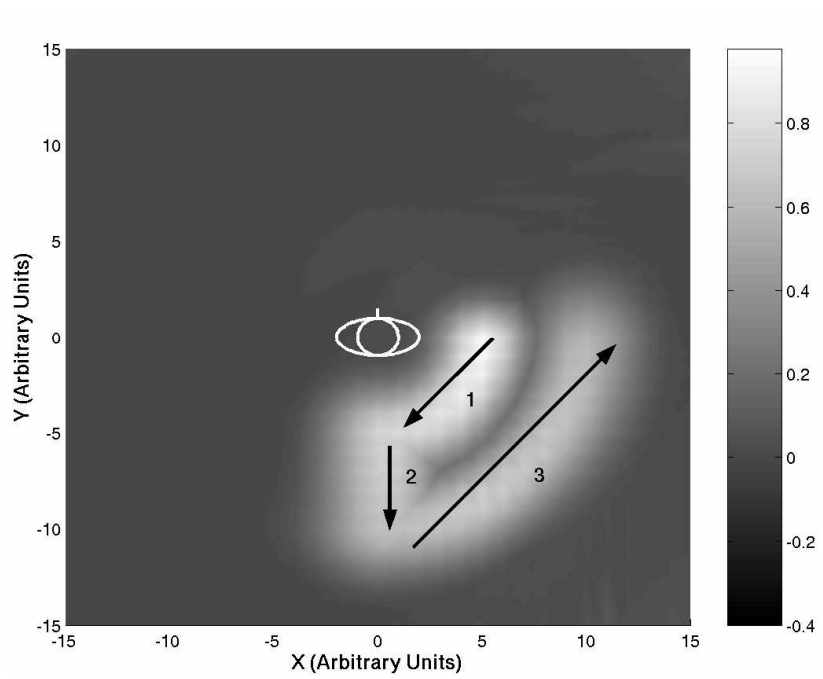


Figure 4: Byrne/ms2867

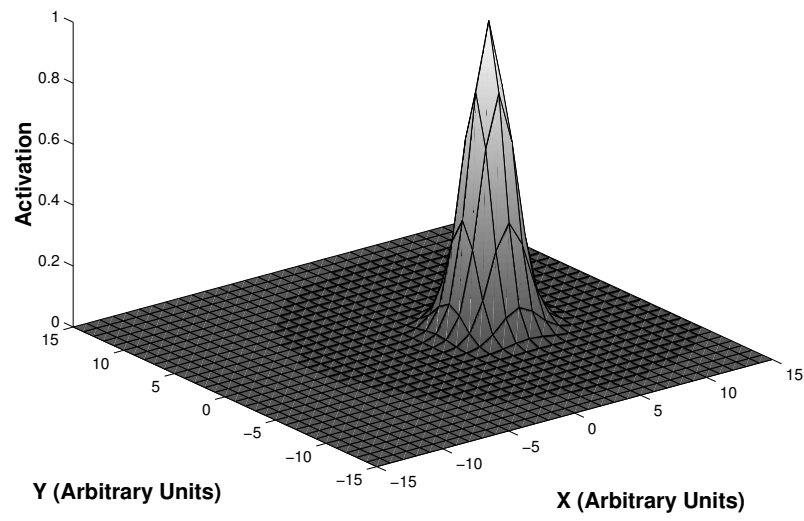


Figure 5: Byrne/ms2867

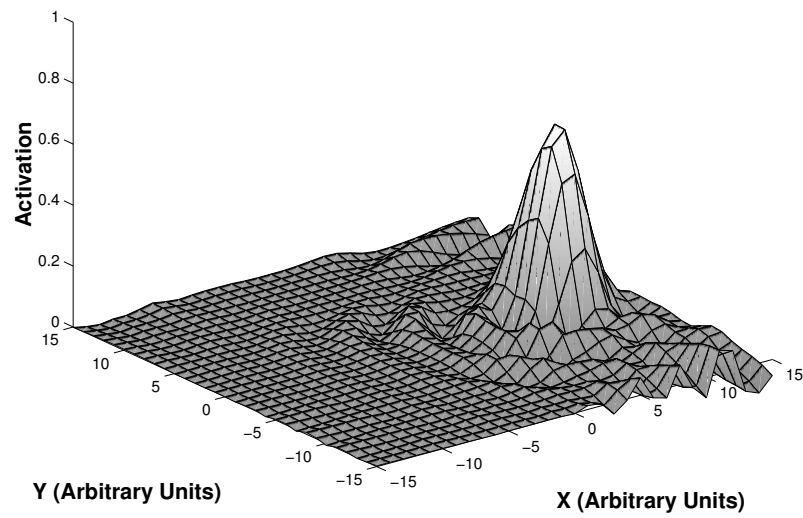


Figure 6: Byrne/ms2867

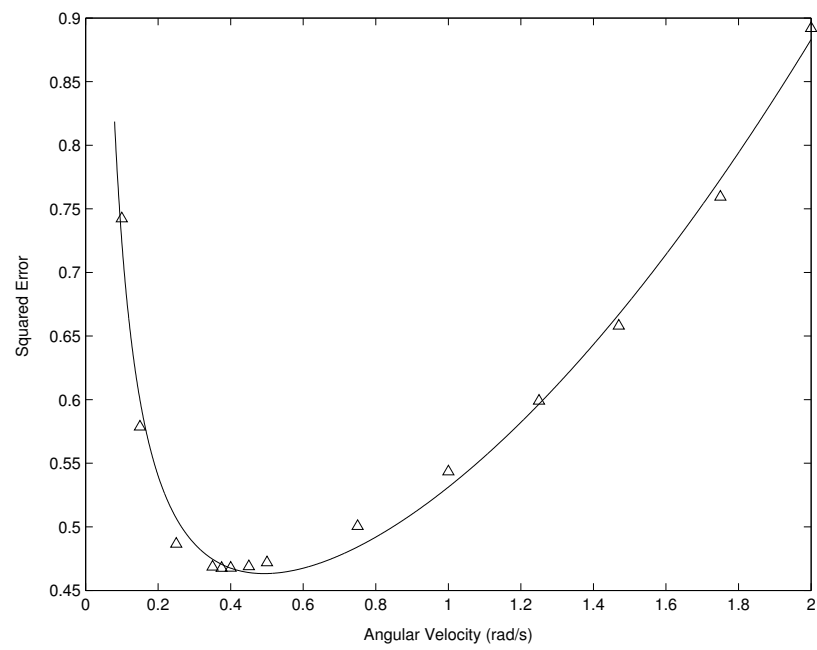


Figure 7: Byrne/ms2867

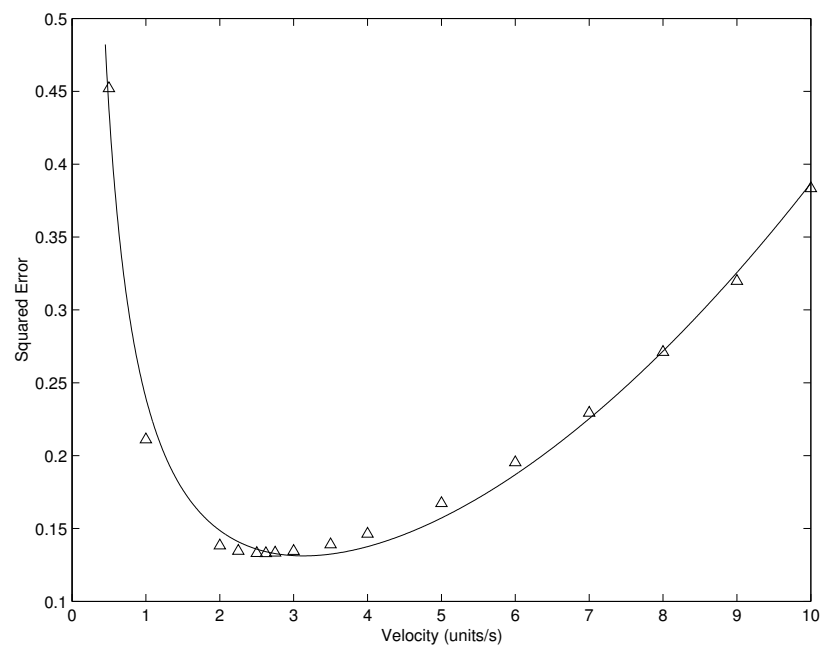


Figure 8: Byrne/ms2867

Figure 1: Translation and rotation of reference frame.

Figure 2: Schematic of proposed neural network. The main layer neurons are interconnected with weights, ‘a’. They are also each connected to the neurons with the same x and y coordinates in the other layers via unit weights. Neurons in the translational velocity layer multiply current egocentric velocity with the corresponding main neuron activation from the current time step and feedback with weights, ‘b’. V_y neurons have been left out of this diagram for clarity. Finally, neurons in the angular velocity layer multiply main layer activations by current head rotation speed and feedback with weights, ‘c’.

Figure 3: Trajectory of the person along the x -axis while maintaining the egocentric coordinates of a stationary object. Each of the four levels shows the position of the person relative to the object at a different stage of the navigation. The dashed line is the path traveled by the person.

Figure 4: Main layer activation representing the egocentric location of a stationary object, from the perspective of the moving observer depicted in figure 3. Main layer neurons are sampled every 20 time steps and superimposed. The numbered arrows correspond to the numbered observer motions in figure 3.

Figure 5: Initial main layer activation before navigation begins.

Figure 6: Final main layer activation when navigation is complete.

Figure 7: Transformation error produced by the network for an observer rotation of 90 degrees. Triangles are simulation data, solid curve is the fit to equation 14

Figure 8: Transformation error produced by the network for an observer translation of ≈ 7.8 units to their egocentric right. Triangles are simulation data, solid curve is the fit to equation