Computations underlying the visuomotor transformation for smooth pursuit eye movements

T. Scott Murdison¹,²,³, Guillaume Leclercq⁴, Philippe Lefèvre⁴ and Gunnar Blohm¹,²,³

¹Centre for Neuroscience Studies, Queen’s University, Kingston, Ontario, Canada
²Canadian Action and Perception Network (CAPnet)
³Association for Canadian Neuroinformatics and Computational Neuroscience (CNCN)
⁴ICTEAM Institute and Institute of Neuroscience (IoNS), Université catholique de Louvain, Louvain-La-Neuve, Belgium

Contributions:
T. Scott Murdison: A, B, C, D, E, F, G, H
Guillaume Leclercq: A, B, D, G, H
Phillippe Lefèvre: A, D, G, H
Gunnar Blohm: A, D, E, G, H

KEY: (A) Research conception and design, (B) Performed experiments, (C) Data analysis, (D) Results interpretation, (E) Prepared figures, (F) Drafted manuscript, (G) Edited and revised manuscript and (H) Approved final version of manuscript

91 total pages (including figures)
17 figures (17 color figures)
235 words in abstract
0 tables

RUNNING TITLE: Computations underlying the 3D pursuit transformation

Corresponding author: Gunnar Blohm
Centre for Neuroscience Studies, Queen’s University
Botterell Hall, Room 229
18 Stuart Street
Kingston, Ontario, K7L 3N6, Canada
Email: gunnar.blohm@queensu.ca

Copyright © 2014 by the American Physiological Society.
Abstract

Smooth pursuit eye movements are driven by retinal motion and enable us to view moving targets with high acuity. Complicating the generation of these movements is the fact that different eye and head rotations can produce different retinal stimuli but giving rise to identical smooth pursuit trajectories. However, because our eyes accurately pursue targets regardless of eye and head orientation (Blohm and Lefèvre 2010), the brain must somehow take these signals into account. To learn about the neural mechanisms potentially underlying this visual-to-motor transformation, we trained a physiologically-inspired neural network model to combine 2D retinal motion signals with 3D eye and head orientation and velocity signals to generate a spatially correct 3D pursuit command. We then simulated conditions of (1) head roll-induced ocular counter-roll, (2) oblique gaze-induced retinal rotations, (3) eccentric gazes (invoking the half-angle rule) and (4) optokinetic nystagmus to investigate how units in the intermediate layers of the network accounted for different 3D constraints. Simultaneously, we simulated electrophysiological recordings (visual and motor tunings) and microstimulation experiments to quantify the reference frames of signals at each processing stage. We found a gradual retinal-to-intermediate-to-spatial feedforward transformation through the hidden layers. Our model is the first to describe the general 3D transformation for smooth pursuit mediated by eye- and head-dependent gain modulation. Based on several testable experimental predictions, our model provides a mechanism by which the brain could perform the 3D visuomotor transformation for smooth pursuit.

KEYWORDS: smooth pursuit, visuomotor transformation, reference frames, Listing’s law, retinal motion, artificial neural network
Introduction

Every day, we perform smooth pursuit eye movements to foveate objects moving across our visual field, allowing our brain to process characteristics of those objects with high acuity. Smooth pursuit initiation is predominantly driven by the two-dimensional (2D) velocity of the target across the retina, i.e. retinal slip (e.g. Orban de Xivry and Lefèvre 2007, Ilg 2008, Ilg and Thier 2008, Krauzlis 2004 and Lisberger 2010). However, as it is usually the case that eye and head orientations are not perfectly aligned in space, the brain must account for three-dimensional (3D) eye-head geometry to produce spatially correct pursuit from 2D retinal slip (Blohm and Lefèvre, 2010). For example, during head roll towards the shoulders, the eyes counter-rotate (in the opposite direction) by a small amount – a phenomenon known as ocular counter-roll (OCR).

As a result, the target’s retinal projection is rotated relative to the direction that the extraocular muscles (attached to the skull) need to move the eyes to minimize retinal slip, therefore requiring that the brain accounts for 3D eye-in-head geometry for spatially correct pursuit, as depicted in Figure 1A (see Methods section for details). Blohm and Lefèvre (2010) recently showed that we initiate spatially correct smooth pursuit movements taking the 3D head and eye-in-head orientations into account (see Methods section for details). Thus the central nervous system (CNS) interprets each of these signals in a geometrically correct way when transforming 2D retinal velocity into a 3D eye velocity command, while also obeying the behavioral constraints of Listing’s law (e.g. Crawford et al. 2003). This means that the CNS generates different spatially correct smooth pursuit motor commands from the exact same retinal input, depending on the 3D eye and head orientations. However, exactly where and how populations of neurons in the brain perform this sensory-to-motor transformation is unclear. Furthermore, no model predictions exist
regarding what neural properties electrophysiologists might expect to find when recording from areas involved in the 3D visuomotor velocity transformation for smooth pursuit.

In general, visual inputs to the pursuit system arise in the primary visual cortex (V1) and are then projected through the middle temporal area (MT) and medial superior temporal area (MST) to parietal and frontal regions (Hubel and Wiesel 1968; Hawken et al. 1988; Movshon and Newsome 1996; Maunsell and van Essen 1983; Illg 2008, Krauzlis 2004 and Lisberger 2010) as well as subcortical structures such as the superior colliculus (SC) and the cerebellum (e.g. Lisberger et al. 1987, Keller and Heinen 1991, Illg 1997, Illg 2008, Krauzlis 2004 and Lisberger 2010). Thus, the visuomotor transformation for pursuit could theoretically be carried out at any point throughout the pursuit circuitry. However, it has been hypothesized that areas MT and MST possess all the properties required to perform the visuomotor velocity transformation for smooth pursuit (Blohm and Lefèvre 2010). For this reason, we modeled the inputs to our network after the inputs to areas MT and MST.

MT and MST provide the primary visual input to the rest of the pursuit circuitry, but their firing characteristics are distinct from one another. In particular, neurons in MT are selective to retinal velocity and position – that is, neurons in MT have defined retinal receptive field locations and at each receptive field location there are neurons tuned for retinal velocity (Albright 1984; Mikami et al. 1986; Newsome et al. 1988; Gattass and Gross 1991; Perrone and Thiele 2001; Inaba et al. 2007; Illg 2008; Inaba et al. 2011; Richert et al. 2013). Area MST receives direct input from MT (Maunsell and van Essen 1983; Ungerleider and Desimone 1986), but also receives input from frontal areas which likely provide eye and head orientation information (e.g. Illg 2008). In contrast to MT, it is thought that MST neurons code visual motion stimuli in non-retinal (i.e. head-centered, spatial or intermediate) coordinates (Fujiwara et al.
2011; Inaba et al. 2007; Chukoskie and Movshon 2009; Inaba et al. 2011), which might help generate a correct interpretation of visual information for spatially accurate pursuit. It has therefore been hypothesized that MT and MST are involved in the visuomotor transformation of retinal signals (Bremmer et al. 1997; Lisberger and Movshon 1999; Blohm and Lefèvre 2010).

While the visual tuning properties of areas MT and MST are relatively well understood, there is a dearth of empirical evidence and model predictions regarding what electrophysiologists might find regarding neural properties related to motor tuning in MT and MST. Only a few studies have investigated the motor tunings of neurons in area MT (Groh et al. 1997; Born et al. 2000), and those of either the lateral (Thier and Erickson 1992) or dorsal portions of area MST (Komatsu and Wurtz 1989; Fujiwara et al. 2011). This lack of both empirical and theoretical evidence might have prevented neuroscientists from fully understanding the neural mechanisms for the visuomotor velocity transformation underlying smooth pursuit.

Previous neural network studies of MT and MST have provided theoretical predictions about how neurons in areas MT and MST can be used to perform several 1D and 2D aspects of the transformation, including detecting spatial heading direction from optic flow (Cameron et al. 1998), reconstructing head- and world-centered target motion in 2D (Dicke and Thier 1999), and how MT and MST cells can interact to use target, eye and background motion signals to control smooth pursuit and suppress the optokinetic nystagmus (OKN) in 2D (Pack et al. 2001), but none of these studies modeled the general, 3D visuomotor transformation for the initiation of smooth pursuit (Blohm and Lefèvre 2010), as we do here.

In this study, we used a neural network modeling approach to decipher the neural properties underlying the 3D visuomotor velocity transformation for the initiation of visually guided smooth pursuit. To do this, we used a 3D geometrical smooth pursuit model to train a
simple rate-based, feed-forward network model. We then probed the emergent network properties by implementing several realistic experimental simulations, each with different 3D requirements for the visuomotor transformation, including (1) head roll-induced OCR, (2) oblique gaze-induced retinal rotations, (3) eccentric gaze orientations (invoking the half-angle rule) and (4) pursuit following optokinetic nystagmus. Meanwhile, we simulated electrophysiological recording techniques (visual tuning and motor field tuning) and microstimulation techniques. Using these simulations to assay the firing properties of our artificial neurons, we then made testable predictions about how neurons involved in the visuomotor velocity transformation should respond to changes in retinal and extraretinal inputs, as well as how their contributions to pursuit output (i.e. their motor tunings) should change depending on the 3D requirements of the visuomotor transformation. The implementation of these experimental simulations using our network model therefore provides specific, testable predictions about what neural properties electrophysiologists might expect to find when recording areas thought to be involved in the visuomotor velocity transformation for smooth pursuit, such as in areas MT and MST.

Methods

It has been shown that the brain performs a spatially correct visuomotor reference frame transformation of retinal signals to head-centered smooth pursuit commands using the kinematics of the eyes in the head and the head relative to the shoulders (Blohm and Lefèvre 2010). The goal of this study was to reveal how populations of neurons in the brain could combine retinal and extraretinal signals to produce these motor commands. We did this by first training a physiologically inspired artificial neural network model to produce spatially accurate smooth
pursuit from various eye and head orientations and velocities, then probing the network’s activation properties while it performed the transformation under different eye-head geometries. By examining the activation properties of hidden layer units (HLUs) in these simulations, we could reveal the coding mechanisms that lead to the transformation, ultimately enabling us to compare our findings to those found in pursuit areas of the brain.

**Neural network model architecture**

We modeled the brain’s 3D visuomotor velocity transformation for smooth pursuit (Blohm and Lefèvre 2010) using a physiologically inspired, four layer feed-forward neural network. Figure 2A illustrates one example of the simulated pursuit task that we used to generate the geometrically correct training set used to train the network. Figure 2B shows a schematic of the network model’s architecture. Retinal location and retinal slip information was provided to the network using an area MT-like combined retinal target position and velocity input signal (Gattass and Gross 1991; Richert et al. 2013). The eye and head orientations and velocities required for a geometrically correct pursuit command (Blohm and Lefèvre 2010) were also provided to the network using neural activities coded in 3D push-pull coordinates (King et al. 1981; Fukushima et al. 1990; Fukushima et al. 1992; Xing and Andersen 2000; Blohm et al. 2009; Blohm 2012), representing eye and head efference copy signals (Crawford 1994; Crawford et al. 2003; Klier et al. 2007), which are not subject to proprioceptive delays that might preclude these signals from being used in online processing of visual signals (Wang et al. 2007). Thus, a total of five input population activities (Figure 2B, left: (1) combined retinal position and velocity, (2) head orientation and (3) velocity and (4) eye orientation and (5) velocity) were then passed through two consecutive hidden layers (HLs), each consisting of between 9 and 100
hidden layer units (HLUs), depending on the desired network size. The number of HLUs in the
two hidden layers was equivalent for each network size. The output layer consisted of a
rotational eye velocity command coded in 3D brainstem coordinates (i.e. with a mixed vertical-
torsional coding of eye orientation; Crawford et al. 1991; Crawford and Vilis 1992; Crawford
1994; Suzuki et al. 1995; Blohm et al. 2009; Blohm 2012), similarly to the input eye velocity
signal. Each layer was fully connected by weight matrices (\(w_{\text{in}}\), \(w_{\text{HL}}\), and \(w_{\text{out}}\)) whose values
were adjusted during training to minimize output error. The input-output relationship for each
HLU was sigmoidal, mimicking the nonlinear transfer function of actual neurons (Naka and
Rushton 1966a, 1966b, 1966c):

\[
a(x) = \frac{1}{1+e^{-x}} \quad [1]
\]

Input activations were not put through this sigmoidal function and the transfer functions for
output units were purely linear.

**Inputs**

*Combined retinal position and velocity*

We defined a population of neurons to code for the retinal target position and velocity based on a
cyclopean eye representation (Ono and Barbeito 1982; Ono et al. 2002; Khokhotva et al. 2005;
Blohm et al. 2009; Blohm 2012). This was done because we were not interested in distance
effects driving vergence eye movements. Neuron receptive fields (RFs) were placed at various
eccentricities about a circle and each neuron had a preferred retinal direction and speed
(velocity). The neurons were distributed according to these parameters (thus the neurons were
distributed across four dimensions, i.e. horizontal/vertical position and velocity). Represented by
grey dots in the retinal input panel of Figure 2B, visual input unit RFs were distributed across
four eccentricities (0°, 5°, 10° and 25°) and across eight linearly spaced polar angles (0° through 315° in 45° steps). At each unit’s RF location, the units’ motion response functions had preferred velocities distributed across four speeds (5°/s, 20°/s, 45°/s and 80°/s) and across eight linearly spaced directions (0° through 315° in 45° steps). These eccentricities and speeds allowed for reliable coding of our visual inputs, which had a maximum eccentricity of 20° and maximum speed of 84°/s, since units tuning functions had enough overlap so that for every possible retinal position and velocity there was at least one input unit with an activation of at least 0.7 (maximum activation = 1; minimum = 0). The activation function for these neurons was determined such that the activity of each neuron depended on the similarity of the retinal input to the RF location and preferred velocity (i.e. the greater the difference between the two, the lower the activation), mimicking the visual motion responses of neurons in area MT/V5 (Albright 1984; Mikami et al. 1986; Perrone and Thiele 2001):

$$a_i = \exp \left[ -\frac{(p_x-x_i)^2 + (p_y-y_i)^2}{2\sigma_{RF}^2} \right] \exp \left[ -\frac{(\theta_T-\theta_i)^2}{2\sigma_\theta^2} \right] \exp \left[ -\frac{(\log_2(v_T/v_i))^2}{2\sigma_v^2} \right]$$  \[2\]

Where:

- $a_i$ represents the normalized activation of unit $i$ (value between 0 and 1, a.u.),
- $p$ represents the horizontal (subscript $x$) and vertical (subscript $y$) retinal target eccentricities (°),
- $x_i$ and $y_i$ represent the horizontal and vertical eccentricities (°) of unit $i$’s receptive field (RF) center, respectively,
- $\sigma_{RF}$ represents the width of the unit $i$’s RF (°)
- $\theta_T$ represents the retinal direction of target $T$ (°),
- $\theta_i$ represents the preferred retinal direction of unit $i$ (°),
- $\sigma_\theta$ represents the width of unit $i$’s preferred direction tuning (°),
$v_T$ represents the retinal speed of target $T$ (°/s),
$v_i$ represents the preferred retinal speed of unit $i$ (°/s),
and $\sigma_v$ represents the width of unit $i$'s preferred speed tuning (a.u.).

The space between units coding for retinal position and the width of the Gaussian RFs ($\sigma_{RF}$) increased linearly with eccentricity to account for decreased visual acuity away from the fovea, given by the piecewise equation [3]:

$$\sigma_{RF} = \min(1.2 \times \sqrt{x^2 + y^2}, 20)°$$  [3]

Similarly, the space between units coding for retinal velocity increased with speed, however the response widths remained constant for both the lognormal speed tuning ($\sigma_v$) and for Gaussian directional tuning ($\sigma_\theta$), which were set to 1.25 (a.u.) and 45°, respectively. Note that because retinal velocity drives smooth pursuit initiation (Rashbass 1961) neurons were not sensitive to zero retinal speed, leaving a point of discontinuity at zero in our model. Accordingly, our training set contained no trials with zero retinal speed. A total of 1024 neurons coded for the combined retinal position and velocity input space.

Eye-in-head and head-on-shoulders orientations and velocities
We also coded for eye-in-head and head-on-shoulder orientations and velocities required for the 3D visuomotor velocity transformation for smooth pursuit (Blohm and Lefèvre 2010). For each of these extraretinal inputs we converted the 3D angle vector orientation, which describes rotations as angles about the vertical, torsional and horizontal axes ($r_x$, $r_y$, $r_z$), into a 6-dimensional array consisting of input unit activities (Smith and Crawford 2005; Keith et al. 2007; Blohm et al. 2009; Blohm 2012), in an antagonistic push-pull arrangement (King et al.
1981; Fukushima et al. 1990; Fukushima et al. 1992; Xing and Andersen 2000; Crawford 1994; Crawford et al. 2003; Klier et al. 2007). This resulted in 6 units coding for each 3D eye orientation, head orientation, eye velocity and head velocity. The activations of each unit \( a_{\pm,i} \) were computed in the following way (Smith and Crawford 2005; Keith et al. 2007; Blohm et al. 2009; Blohm 2012):

\[
a_{\pm,i} = 0.5 \pm \frac{r_i}{2 \times r_0} \quad [4]
\]

where the maximum angle \( r_0 \) of eye-in-head orientation was 50°, maximum angle of head-on-shoulders orientation was 75° and maximum velocity of both eye and head was 100°/s. The orientation axes for both eye orientation and eye velocity were rotated by 45° about the vertical axis to account for the mixed vertical-torsional coding of eye orientations seen in the brainstem neural integrator (Crawford et al. 1991; Crawford and Vilis 1992; Crawford 1994; Suzuki et al. 1995; Blohm et al. 2009; Blohm 2012).

**Output**

*Eye-in-head velocity motor output*

We had six output units representing the push-pull eye velocity motor commands (King et al. 1981; Fukushima et al. 1990; Fukushima et al. 1992; Xing and Andersen 2000; Crawford 1994; Crawford et al. 2003; Klier et al. 2007) and the network was trained according to the activities of these six units (Smith and Crawford 2005; Keith et al. 2007; Blohm et al. 2009; Blohm 2012). At each training epoch, using equation [4] we computed the activities associated with the required 3D angular velocity vector \( (r_x, r_y, r_z) \) from the training set, and compared those activities to the current output layer activities. The maximum speed of output eye-in-head velocity \( r_0 \) was 100°/s and, like the input coding for eye-in-head orientation and velocity, the orientation axis
was rotated by 45° about the vertical axis to account for the mixed vertical-torsional coding of
eye orientations seen in the brainstem neural integrator (Crawford et al. 1991; Crawford and

Training set and training method

Because the goal of our network model was to describe the neural mechanisms underlying the
visuomotor velocity transformation for smooth pursuit, using a physiologically plausible set of
retinal and extraretinal signals to train our network was paramount. With this in mind, we
generated our training set using a 3D geometrical model which has been described previously
elsewhere (Blohm and Crawford 2007; Blohm and Lefèvre 2010; Leclercq et al. 2012; Leclercq
et al. 2013a; Leclercq et al. 2013b) that adhered to the known physiological and kinematic
constraints (Blohm and Lefèvre 2010) of head-free smooth pursuit eye movements. Figure 2A
shows the schematic of this geometrical model. We simulated a pursuit task in which a subject
pursued a point stimulus displayed at various eccentricities and velocities on a frontoparallel
screen, under various head and eye orientations and velocities. Using this simulated setup, we
randomly generated over 800,000 training points such that the full range of plausible eye, head
and retinal position and velocity inputs and outputs was covered approximately uniformly within
the training set. For any given training point, our model assumed that: (1) we describe the
transformation at a single moment in time; (2) we only consider pursuit motor commands for the
eyes (i.e. there was no head contribution to the output gaze shifts); (3) the stimulus is moving on
a screen which is frontoparallel to the subject; (4) the desired gaze path is also within that
frontoparallel plane; and (5) we only consider version pursuit movements and do not model
vergence movements. The training set therefore adhered to several physiological constraints,
including Donders’ law (Blohm and Crawford 2007; Glenn and Vilis 1992) and Listing’s law (Blohm and Crawford 2007), while accounting for the natural variability of head and eye orientations (Glenn and Vilis 1992; Aw et al. 1996; Goossens and Van Opstal 1997; Murdison et al. 2013) and movements (Aw et al. 1996; Tweed and Vilis 1987; Tweed et al. 1992; Tweed and Vilis 1990; Blohm and Crawford 2007). Details of model implementation along with a dual quaternion Matlab toolbox can be found in Leclercq et al. (2013a,b).

We used this 3D geometrical model to generate a training set consisting of the retinal and extraretinal inputs and required eye velocity output to train eight different network sizes (9, 16, 25, 36, 49, 64, 81 and 100 HLUs in each hidden layer). The inputs to the first layer consisted of retinal, eye and head orientation and velocity signals while the required output eye velocity signals were compared to the fourth layer activations. We batch trained each network using a pseudo-Newton method with pre-conditioned gradient descent, which encouraged faster training convergence by multiplying the weight adjustments at each training step by a term which corresponded to the sign of the mean squared error (MSE) gradient, but was independent of its value. The connections whose weights were adjusted during training included those between layers 1-2, layers 2-3 and layers 3-4 (i.e. connections adjacent to the hidden layers). The number of training points used depended on the total number of adjustable connection weights in each network to avoid over-fitting. To implement and train the neural networks we utilized the Neural Network Toolbox (v6.0.4) within Matlab 7.10.0 (R2010a) (The Mathworks, Natick, MA) running on a Dell T7500 computer, equipped with a 64 bit Intel Xeon X550 CPU (8 Mb cache, 2.66 GHz, 6.4 GT/s dual channel QuickPath Interconnect, 24 GB RAM) and running a Windows 7 (Professional Edition) operating system. Training lasted from under an hour (9 HLU network) to up to a few days (100 HLU network). We stopped training when the maximum number of
epochs (500,000) was reached. Note that good network performance based on test points not used for training (see RESULTS) verified that our network was not over-specified, given the number of training points we used in batch training.

**Neural network analysis**

We performed several analyses to assess the extent to which and mechanisms by which our network model (and specifically our hidden layer units) performed the visuomotor velocity transformation. These analyses were similar to those used in previous work (Zipser and Andersen 1988; Buneo et al. 2002; Pesaran et al. 2006; Blohm and Crawford 2007; Blohm et al. 2009; Blohm 2012; Bremner and Andersen 2012).

**Network performance**

First, we assessed the performance of the network models after training by ensuring that our network both incorporated extraretinal signals and accounted for the physiological constraints of Listing’s law when computing smooth pursuit movement commands (Tweed et al. 1992; Tweed and Vilis 1990; Blohm and Crawford 2007; Blohm and Lefèvre 2010). We computed the observed 3D compensation index, which indicates the extent to which extraretinal signals were accounted for to transform the retinal velocity vector into the head-centered motor command (Blohm and Crawford 2007; Blohm et al. 2009). Briefly, the compensation index was defined by the dot product between the actual, network-generated 3D eye velocity and the geometrically required 3D eye velocity (each with reference to the retinally-predicted movement). As such, a compensation index of 0 represented network output movements that were generated as if eye-head orientations were at 0, i.e. that none of the geometry was accounted for. To ensure that our
network obeyed Listing’s law, we compared the torsional component of the output rotational velocity with that predicted by the half-angle rule and also computed the Listing’s law error by finding the magnitude of the difference between the predicted and actual 3D velocity output vectors (Tweed and Vilis 1987; Blohm and Lefèvre 2010). Together, these methods ensured that our network both incorporated extraretinal signals and accounted for the physiological constraints of Listing’s law when computing smooth pursuit movement commands (Blohm and Lefèvre 2010).

Visual receptive fields and velocity tuning curves (VTCs)
We visualized the activation properties of our HLUs by computing their receptive field center-of-mass (COM) locations (similar to a hot spot of unit activation at which stimuli were presented for our simulations) and, for that location, determining the retinal tuning across the entire velocity space (which combines direction and speed) for each HLU, which we call the retinal velocity tuning curves (VTCs). Using these VTCs, we determined the preferred retinal velocity direction (visual PD) of each HLU by computing the activity-weighted circular mean across all retinal velocities (up to 90º/s).

Visual input properties: gain modulation and tuning shifts
We then probed the emergent properties used by our network HLUs by observing how VTCs changed in response to changes to extraretinal inputs during experimental simulations. The two main properties we observed were gain fields and velocity tuning shifts. Gain fields were characterized by the up- and down-modulation of each unit’s VTC with non-visual inputs. The
gains of each HLU were quantified using regression analysis between the normalized extraretinal input and the average activity in the VTC, as described by equation [5]:

\[
\bar{a}_{i,j} = G_j \times \frac{X_i}{\max(|X|)} + \text{intercept} \quad [5]
\]

where \( \bar{a} \) represents the average activity across the entire retinal velocity space (a.u.),

the subscript \( i \) represents the dimension of the current extraretinal signal (horizontal, vertical or torsional),

the subscript \( j \) represents the current HLU number,

\( X \) represents the magnitude of the extraretinal signal (° or °/s),

and \( G_j \) represents the gain computed via regression analysis for unit “j” (a.u.).

We computed the gain \( G_j \) using equation [5] for each unit regardless of if there was also a shift in tuning resulting from extraretinal changes, meaning that the term \( G \) captured any modulatory behaviors, not only those that were gain field-like. In contrast, tuning shifts were characterized by modulations of VTCs in response to changes in extraretinal inputs resulting in shifts of the HLU’s overall tuning properties. We quantified these shifts using regression analysis between the change in extraretinal inputs and the shift of overall tuning directions (compared to the tuning with zero extraretinal inputs).

Output properties: motor fields and simulated microstimulation

In contrast with VTCs, which represented each unit’s response across retinal velocity inputs, we also examined how each HLU’s activity correlated with the network eye velocity output by computing the 3D motor field of each unit (Blohm et al. 2009). The 3D motor field was determined by the activity of each HLU across all possible 3D eye velocity outputs, based on our complete geometrical model, such that for each 3D eye velocity output (and given extra-retinal
signals) we computed the corresponding retinal inputs. For each 3D motor field, we computed
the activity-weighted 3D motor field center-of-mass (COM$_{MF}$) and 2D preferred motor output
direction (PD$_{MF}$). The COM$_{MF}$ was computed as the activity-weighted mean across each velocity
output axis. The PD$_{MF}$ was the activity-weighted circular mean in the frontoparallel output plane,
and thus ignored the effects of torsional velocity output. We used the PD$_{MF}$ in reference frame
analyses of OCR and oblique gaze simulations because only rotations of the 2D retinal
information were required for spatially correct compensation.

Additionally, simulated microstimulation of each unit allowed us to examine how each
unit contributed to the motor output while accounting for downstream connectivity. We
simulated microstimulation-induced evoked eye movements by creating circumstances in which
there would normally be no required network output (by fixating on a foveated target and
locating the probe at each unit’s retinal RF center-of-mass location), then setting the activity of
each HLU to an artificially high value (activation = 5) to ensure significant microstimulation-
induced movement vectors. We then computed the network 3D output velocity command and
repeated this procedure for every unit under various simulated experimental conditions.

**Experimental simulations**

To see how the network model carried out the transformation for smooth pursuit, we simulated
several pursuit experiments (illustrated in Figure 1), including: (1) head roll-induced ocular
counter-roll (Fig. 1A), (2) retinal rotations due to oblique gaze positions (Fig. 1B), (3) adherence
to the half-angle rule (Fig. 1C) and (4) pursuit initiation during optokinetic nystagmus (Fig. 1D).
Overall, we used ten different experimental simulations to describe the input (i.e. visual tuning)
and output (i.e. motor fields and microstimulation) coding reference frames of HLUs in order to
fully characterize the progression of the visuomotor transformation through the network layers.

This framework ultimately enabled us to make testable predictions about the neurophysiology underlying smooth pursuit eye movements.

First (see Figure 1A), to determine the visual input reference frame of HLUs, we examined the influence of simulated head roll-induced ocular counter-roll (OCR) on the VTCs of each HLU. The transformation for smooth pursuit is complicated by the addition of head roll, resulting in OCR. The counter-rotation of the eyes results in a misalignment between the spatial, head-centered and retinal target direction, as revealed by the direction of the target (black arrow) when projected into each reference frame’s coordinates (spatial, screen-centered frame as black dotted axis; head-centered frame in blue; retinal frame in red). This misalignment therefore requires that the brain accounts for head orientation and OCR when transforming retinal slip into a pursuit command. Although typically small in magnitude (e.g. static OCR between 6-16% of head roll on average; Murdison et al. 2013), with the dynamic vestibulo-ocular reflex, ocular torsion can be as large as 70% of head roll (Aw et al. 1996). Ocular torsion-related pursuit errors (Murdison et al. 2013) and perceptual errors (Wade and Curthoys 1997) are significant, and their (presence) absence indicates the (non) existence of 3D eye orientation signals in the generation of motor commands (Murdison et al. 2013) and in visual perception (Wade and Curthoys 1997).

We investigated how the network accounted for these signals by simulating head roll-induced OCR and mapping the directional velocity tuning of hidden layer units, a method similar to electrophysiological recordings from area MSTd under conditions of whole-body static roll-tilt (Fujiwara et al. 2011). We made predictions about how a unit should change in response to extraretinal changes if that unit coded information in different reference frames. For example, if an HLU was coding according to a retinal reference frame (i.e. based solely on retinal input), the
tuning would be independent of extraretinal changes and the retinal prediction (red on-screen vector) would be rotated by the angle of OCR ($\theta_{\text{OCR}}$; exaggerated here for illustration purposes), resulting in an idealized regression gain of 0 between $\theta_{\text{OCR}}$ and VTC shifts. However, if OCR is accounted for (head-centered hypothesis), then the predicted gaze vector (blue on-screen dotted vector) is equivalent to the spatial prediction (black on-screen vector) since the extraocular muscles are attached to the skull (VTC shift regression gain equal to 1). Using a reference frame approach similar to the visual tuning analyses, we determined the motor field reference frames of HLU s by observing how motor fields varied under several simulated experimental conditions identical to those used in the input analyses. Specifically, we used regression analysis to compare between changes to the PD$_{\text{MF}}$ (representing each unit’s contribution to the change in only the horizontal and vertical output velocity components) and changes in head roll and head roll-induced OCR. Using these same reference frame predictions, we performed multiple regression analysis between head roll, OCR and microstimulation-evoked network output to determine the HLU output reference frames.

Second (see Figure 1B), we simulated the effects of oblique gaze-induced retinal rotations (i.e. with no actual eye-in-head torsional component; Blohm and Lefèvre 2010). Oblique eye orientations result in rotations of retinal input relative to space, without any accumulation of ocular torsion (because eye orientations are within Listing’s plane, i.e. the plane containing the rotation axes of all possible eye orientations; Tweed and Vilis 1990; Crawford and Vilis 1991; Blohm and Lefèvre 2010). This retino-spatial misalignment is demonstrated in Figure 1B; as the gaze becomes more eccentric in an oblique (45°) direction, the retinal velocity vector becomes increasingly rotated ($\theta$). Since this is an effect of projecting fronto-parallel planes onto a spherical retina without any actual ocular torsion, the brain must have an internal
model of this effect induced by oblique eye orientations in order to compensate for these
distortions when planning pursuit movements. Conversely, if the brain were to carry out the
transformation based solely on retinal information (and did not use eye orientation signals), this
would result in eye movements (on-screen red arrows) that are rotated by \( \theta \) relative to the spatial
target direction (on-screen grayscale arrows). Note that for demonstration purposes the predicted
pursuit trajectory errors in Figure 1B are not drawn to scale. Here, we compared the shifts of the
VTCs with the retinal rotation angle at various oblique gaze positions using regression analysis
(Blohm et al. 2009; Blohm and Lefèvre 2010; Blohm 2012). We also examined how HLUs
compensated for the rotations induced on the retina during oblique gaze positions by comparing
this rotation angle with the rotation of the PD\(_{MF}\) (motor fields), as well as with the rotation across
evoked network output (microstimulation).

Third (see Figure 1C), we used the predictions made by Listing’s law and the resulting
half-angle rule to investigate the input coding reference frame of units. Listing’s law necessitates
that the rotational eye velocity axis contains a torsional “tilt” proportional to half the angle of
gaze eccentricity \( \theta/2 \) in order to keep the eye orientation axis within Listing’s plane, known as
the half-angle rule (Tweed and Vilis 1990; Crawford and Vilis 1991; Blohm and Lefèvre 2010).
As eccentricity increases, although the retinal input remains the same, the required eye velocity
axis tilts according to the half-angle rule. In Figure 1C, these axes are drawn to scale such that
red- and gray-scale axes correspond to the pursuit velocity axes for matching vertical
eccentricities predicted either by the retinal hypothesis (red-scale) or by the half-angle rule (gray-
scale). However, if the network does not account for the half-angle rule and simply uses retinal
stimulation to drive eye movements, it might be expected that the eye velocity axis does not tilt
at all, as would be the case for movements initiated from primary position (lightest gray on-
screen arrow). In this analysis, we computed the predicted torsional component of the output  
velocity, given by the half-angle rule with initial eye orientations at various vertical eccentricities  
and compared the torsional shift of VTCs with the required value using regression analysis  
(Blohm et al. 2009; Blohm and Lefèvre 2010; Blohm 2012). We used a similar approach under  
different eye orientation conditions in which the half-angle rule specifies that the 3D output  
velocity axis have a torsional “tilt” proportional to half of the gaze angle. Here, we used  
regression analysis to compare this required torsional tilt both to the torsional change in the  
location of each COM
( representing each unit’s motor field contribution to the change in 3D  
motor velocity output) and to the required tilt to the change in evoked network output  
(microstimulation).

Fourth (see Figure 1D), we investigated the effects of retinal stimulus speed under  
various eye velocities using a simulated optokinetic nystagmus (OKN) task with a large visual  
field of motion. To initiate smooth pursuit under these initial visual and gaze motion conditions,  
the brain must perform a vector addition of the retinal target velocity \( (s_{tr}) \) and gaze velocity \( (s_{g,0}, \)  
green) in order to estimate the speed of a screen-centered target stimulus \( (\hat{s}_{t,s}, \) gray), and finally  
execute a pursuit movement in response \( (s_{g,f}, \) black). In this analysis we computed each HLU’s  
activity at every combination of retinal target velocity and eye velocity greater than 10º/s to  
examine the extent to which the activity of each HLU was modulated by either target or eye  
velocity. Target and eye velocities were either moving in the visual preferred direction (positive  
speeds) or in the null direction (visual PD+180º; negative speeds), and we simulated every  
possible combination of target and eye velocities by systematically varying both the on-screen  
target speed and the contribution of the eyes to the OKN. As such, the direction of the summed  
activity gradients across all combinations of target and eye velocities indicated the extent to
which target and eye signals were coded separately (separable) or in a combined fashion (inseparable). Twice this gradient sum is equal to the separability index used in electrophysiological and network studies (Buneo et al. 2002; Pesaran et al. 2006; Blohm et al. 2009; Blohm 2012; Bremner and Andersen 2012) and we used the same index here (separability indices of 0° and 180° represented separable encodings of target and eye signals while indices of 90° and 270° represented inseparable, combined encodings of target and eye signals). However, we often found units that displayed multiple inseparable coding schemes that offset one another, resulting in an incorrect finding of a “separable” target-centered coding scheme. To overcome this limitation, we first selected units that exhibited this behavior by locating the target velocities of minimal and maximal activity. If the absolute velocities were less than 30°/s apart (presumably indicating a non-monotonic gradient direction), we split the velocity field from which we sampled the characteristic gradient direction into two fields at the maximum or minimum target velocity that was closer to the center of the velocity range (i.e. at the target velocity that was not near the edge of the field). The rest of the gradient analysis proceeded as previously described, but with each of these particular units essentially counting as two units with two gradient directions. In this way, we could characterize the true separability of target and eye motion signals in network units.

Results

Network performance

We first ensured that the network adequately performed the 3D visuomotor transformation. To do so, we first computed the observed 3D compensation, which indicates the extent to which the network used extraretinal eye and head signals to transform the 2D retinal velocity into a 3D
motor output for the eyes (Blohm and Crawford 2007; Blohm et al. 2009; Blohm and Lefèvre 2010). We then performed a regression analysis between the observed 3D compensation and the predicted 3D compensation, based on an ideal spatially accurate output for a set of 10,000 test points previously unseen by the network, sub-selected from the full training set (see Methods). Results are shown in Figure 3 for a 100 HLU network (black solid line and grey dots), as well as the other network sizes (regression fits only, dashed, various colors). Figure 3A shows that the 100 HLU network compensated for most of the extraretinal signals, as the regression fit had a slope of 0.86 (and ≥ 0.78 for all networks) and the $R^2$ value of 0.81 also indicated a strong fit ($R^2 ≥ 0.69$ for all networks). Figure 3B shows the distributions of the 3D compensation errors (or the components orthogonal to the predicted 3D compensation vectors), and reveals that across all network sizes the mean errors were relatively small (mean errors < 4.02º/s, SD of error < 3.09º/s). While these compensation parameters indicate that the networks accounted for the extraretinal signals giving rise to a spatially correct transformation, they do not necessarily address the physiological plausibility of the network outputs, and, specifically, their adherence to Listing’s law and the resulting half-angle rule. We computed this torsional component for each network eye velocity output and compared it to the torsional component required by the half-angle rule using regression analysis, revealing a slope of 0.998 and an $R^2$ value of 1.000 for the 100 HLU network (slopes ranging from 0.998 to 1.001 and $R^2 > 0.998$ for all networks), corresponding to a narrow distribution of torsional error with absolute means < 0.03° and SDs < 0.94° for each network size, shown as a histogram in Figure 3C. Together, these analyses provide evidence that the performance of the visuomotor transformation for smooth pursuit, as observed in Blohm and Lefèvre (2010), can theoretically be computed by the brain in a simple, distributed, feed-forward way.
Network analysis: General HLU properties

To see how the network carried out the visuomotor transformation, we needed to determine the relationship between the inputs and outputs of the network’s HLUs. We used a reference frame approach to essentially trace the gradual transformation of information from a retinal reference frame to a spatially correct, head-centered reference frame (Blohm and Lefèvre 2010). We characterized the reference frames of information input to each HLU as well as those of information output from each HLU using experimental simulations. For example, if information input to a HLU were coded retinally, the activity of the HLU should be independent of changes to eye orientation input, and, contrastingly, if the input were coded in a spatially correct way, the HLU’s activity should be modulated by eye orientation in a way that compensates for these changes. Using this methodology, we performed experimental simulations in which we altered the head and eye inputs in order to deduce the input reference frames of each HLU and the reference frames in which each HLU coded motor outputs. Importantly, this simulation-based framework allowed us to make testable predictions about the neurophysiological properties of areas involved in the visuomotor transformation for smooth pursuit.

To do this, we simulated electrophysiological recording studies; that is, we computed the activity of HLUs in our network for different simulated experimental conditions, such that we could create a map of the HLU activity dependent on either retinal input or motor output. First, we mapped the visual receptive fields of HLUs (i.e. the HLU response across all retinal positions). Then, at each RF location, we found the retinal velocity tuning of each HLU from the activities across all possible retinal velocities, leaving us with a retinal velocity tuning curve (VTC) for each HLU, as presented in the centers of the left and right panels in Figure 4A for
example units 85 and 72 from the first and second hidden layers of the 100 HLU network, respectively. Thus, to determine the input reference frame of a unit, we examined how VTCs were modulated by eye and head signals. We investigated the output reference frame of units in two ways: we examined both (1) how a unit’s activity alignment with motor output (motor field) was modulated by eye and head signals and (2) how each unit contributed differently to network output when modulated by eye and head signals (simulated microstimulation).

**Input properties: Retinal velocity tuning curves (VTCs)**

We first analyzed the emergent input reference frame of HLUs in the second layer of the network (or the first hidden layer). To do this, we determined each unit’s activity in response to retinal velocity inputs and varied retinal velocity across a range of retinal speeds from > 0º/s to 45º/s and all directions, while keeping all other inputs constant. For the units in the centers of Figure 4A, left and right sides, the retinal position was centered on each unit’s visual RF, and all eye and head orientation and velocity inputs were kept at zero (representing the target at the center of the visual RF, the eye at primary position and the head in an upright position). Within each VTC, we found the preferred retinal velocity tuning direction (PD; white bars in Figure 4A), the center-of-mass velocity (white circles in Figure 4A), maximum activity (green, upward triangles in Figure 4A) and the minimum activity (magenta, downward triangles in Figure 4A), which we used as topographical “landmarks” for tracking modulations of the VTCs with eye and head input changes. In both layers, VTCs were typically complex in shape, often with more than one hill of activity and more than one trough of activity (e.g. layer 1 unit #85).

The next step in determining the reference frame of the inputs to each unit was to observe how it was modulated by changes in extraretinal signals, such as eye velocity. More precisely,
we asked if the VTCs were gain-modulated by eye velocity, i.e. whether the eye velocity had an approximately multiplicative effect on the activity of the unit, regardless of the retinal velocity. For units in both hidden layers of the 100 HLU network, we found evidence of such gain modulation. For instance, the activity of units 85 (panel A, left) and 72 (panel A, right) is gain-modulated by the eye velocity. For those two units, we observed a constant PD and constant locations of the activity COM, minimum and maximum regardless of eye velocity, but, in general, units can have both a shift of their PD as well as a gain-modulation of their activity.

The modulatory behaviors shown in the left and right panels of Figure 4A were typical of units in each layer, as illustrated in the left and right panels of Figure 4B. In Figure 4B, each box represents the overall gain range of the activity modulation exhibited by the units in each trained network size (the left and right panels represent the first and second hidden layers, respectively) with horizontal and vertical eye velocities. The width of these boxes represents the maxima and minima of the gain values, as determined by regression analyses (see Methods section). Also shown in Figure 4B are the horizontal and vertical gains associated with the units in Figure 4A (units 85 and 72, respectively; white disks with black outlines). Figure 4 reveals that although there was modulation of VTCs in both layers across network sizes, this modulation was often of greater magnitude (i.e. larger gains) in the first hidden layer, though this was not necessarily the case for all networks.

Input properties

In the first experiment, we asked how network units accounted for the rotational misalignment that occurs between the retina and the extraocular muscles under conditions of head roll-induced OCR. Figure 1A illustrates this misalignment. Because the eyes rotate in the opposite direction
of the head, and the extraocular muscles are head-fixed, the direction of the target velocity on the
er retina and the direction of eye movement required to minimize that retinal slip are not spatially
equivalent (see Methods). We thus tested if the units in our network coded for this OCR
compensation by investigating how visual PDs change across head roll and OCR. We show an
example of this analysis in Figure 5A for typical first and second hidden layer units (#s 1 and 7,
respectively) from the 100 HLU network. To show the isolated effects of varying either head roll
or ocular torsion on the tuning properties of these units, we show (inset polar-coordinate plots)
the directional tuning of these example units for each head roll/ocular torsion amount across the
zero-head/ocular torsional cross-sections of the full head roll-OCR space (i.e. we computed the
PD at each head roll/ocular torsion angle while keeping the other signal equal to 0, and the shift
was the difference of these PDs from the head and eyes at 0 torsion; colored to black tuning
curves and corresponding colored to black radii; for a constant retinal target speed of 20º/s). We
then compared the magnitude of this PD shift with that required for a spatially correct output (i.e.
OCR angle) and with the head roll angle using regression analysis. For the first hidden layer unit
(#1), the tuning curve was gain-modulated but there was no significant shift of the tuning
preference, resulting in a regression gain close to zero, thus indicating that this HLU coded in a
reference frame that was approximately retinal. Second hidden layer units’ PDs shifted in a way
that was proportional, though not perfectly compensatory, for the torsional head and eye signals.
Depicted for a typical second hidden layer unit (#7) in the lower panel of Figure 5A, this shifting
behavior is indicative of a mixed coding scheme for target motion, intermediate to both retinal
and spatial coordinate frames.

In the second experiment, we investigated how the HLUs compensated for retinal
rotations induced by oblique eye orientations. When the eyes change orientation in the head to
direct gaze to an oblique (e.g. up and to the right) position, the retinal projection undergoes a slight rotation as illustrated in Figure 1B. As such, the rotations of retinal input are nonlinear and result from 3D properties of rotation (Blohm and Crawford 2007; Blohm and Lefèvre 2010) and also depend on head movements (because of head-movement-related changes to Listing’s law). Thus, the brain must account for the full 3D geometry of the retina, eyes and head to generate a spatially correct pursuit movement from oblique gaze locations.

Using a methodology similar to that in the OCR simulation, we compared the shifts of PDs for each HLU to the magnitude of the retinal rotation induced by oblique eye orientations ranging from eccentricities of 0º to 45º. We found the retinal rotation by computing the angle between a horizontal unit vector rotated from a 0º eye orientation (i.e. with no rotational component) to each oblique position. We then used regression analysis to compare the PD shifts to these values. Thus a retinally-coding unit would exhibit no PD shift for changes in retinal rotation (regression gain = 0), whereas a spatially-coding unit would exhibit a fully compensatory PD shift (regression gain = 1).

In addition to accounting for OCR and oblique eye orientations in the transformation, the network HLUs should also account for half-angle rule (Tweed and Vilis 1990; Crawford and Vilis 1991; Blohm and Lefèvre 2010), as we investigate in the third experiment. As presented in Figure 1C, for the pursuit of a horizontal stimulus starting from a vertical eye orientation, the axis of rotation should be tilted (in the torsional direction) by half of the angle of vertical eccentricity. Although the retinal input is identical for each of these eye orientations (see inset retinal projection, Figure 1C), the motor requirements differ. Therefore, in order for the pursuit command to obey Listing’s law, this torsional tilt must be accounted for in the visuomotor transformation.
Similar to the OCR and oblique gaze simulations, we used regression analysis to compare the PD shifts for each HLU to the torsion predicted by the half-angle rule at various vertical eye orientation eccentricities (equivalent to half the angles of eccentricity, accordingly). Accordingly, retinally coding units should exhibit a regression gain of 0. On the other hand, the preferred tunings of spatially accurate coding units should shift by the required torsional velocity. However, because the half angle tilt directly influences the motor command output (rather than influencing the retinal input), spatially compensatory shifts of visual PDs were represented by regression gains of -1.

**Input reference frame analyses for experiments 1, 2 and 3**

Because head roll and OCR are correlated in our training set ($R^2 = 0.83$), both signals can provide information about OCR, despite the fact that head roll itself is irrelevant to the visuomotor transformation for smooth pursuit. Indeed, technically only knowledge about OCR is needed for spatially accurate pursuit, as illustrated by Figure 1A. However, because of this correlation, we performed a multiple regression analysis to identify head roll and OCR gains. As shown in the pseudo-color plots in Figure 5A for units 1 and 7, the shifting of units’ visual PDs remained constant for most units within the first hidden layer (e.g. unit 1) or varied monotonically across the full head roll-ocular torsion space for units within the second hidden layer (e.g. unit 7).

We determined the head roll- and OCR-related compensatory gains for HLUs in each layer and plotted their distributions with the retinal prediction (dashed lines) in Figure 5B. The gains of multiple regression terms corresponding to head roll and ocular torsion indicated that when the spatial transformation required the network to use head roll and/or OCR signals, the
first hidden layer units (HL1, grey) were coding almost exclusively according to a retinal or nearly retinal input reference frame, whereas the second hidden layer units (HL2, black outline) were coding according to both a retinal frame and an intermediate input reference frame. Group level t-tests on median gains agreed with this observation, as median ocular torsion gains were not significantly different from 0 (first hidden layer: \( t(7) = 1.12, p = 0.30, 95\%CI [-0.002, 0.006] \); second hidden layer: \( t(7) = 1.62, p = 0.15, 95\%CI [-0.02, 0.09] \)). Qualitatively, for all networks the interquartile gain distances were larger in the second hidden layers than in the first hidden layers for both head roll and ocular torsion, suggesting that there were also units coding visual inputs according to an intermediate reference frame in the second hidden layers of each network; however, this finding could not be confirmed using group-level statistical tests.

We found that units within the second hidden layer often coded for a compensation for head roll in addition to OCR \( (r = 0.70, p < 0.001) \), even though OCR was the only signal for which the network had to explicitly compensate. This finding suggests that these units were able to (at least partially) learn an internal model of the head roll-OCR interaction. Furthermore, the combined use of head roll and ocular torsion signals might explain why neither signal independently produced statistically detectable intermediate coding of visual inputs (i.e. strong compensatory shifts) in our second hidden layers. Therefore, when retinal signals are rotated relative to the head due to OCR (Blohm and Lefèvre 2010), this analysis reveals that first hidden layer units coded visual inputs according to a retinal frame and suggests that second hidden layer units coded visual inputs according to both retinal and mixed reference frames, intermediate to retinal and spatial.

In each of the other two experiments (oblique eye orientations and half-angle rule), HLUs exhibited input coding frames consistent with and behaviors qualitatively similar to those found
in the head roll-OCR simulation, as shown in the histograms in Figures 5C and 5D (following the same conventions as Figure 5B). However, when comparing the first and second hidden layer interquartile gain distances, for all networks and for both experiments we found significantly larger distances, strongly suggesting an intermediate coding of visual signals in the second hidden layer (oblique eye orientations: t(7) = -10.0, p < 0.01, 95% CI [-1.35, -0.83]; half-angle rule: t(7) = -7.47, p < 0.01, 95% CI [-0.76, -0.39]).

Experiment 4: visual tuning effects of target and eye speeds during optokinetic nystagmus

For the pursuit system to correctly interpret the motion of a target in space and subsequently pursue it (Blohm and Lefèvre 2010), it must ultimately perform a vector addition of retinal target signals and eye movement signals. However, it is unclear how 2D retinal signals and 3D eye movement signals might be combined in the brain. Here, we simulated an optokinetic nystagmus (OKN) experiment (Fox et al. 1978; Wolfe et al. 1981; Archer 1987) in which a large, textured visual stimulus was displayed moving at various speeds in a direction either parallel to the PD of the unit or in a direction exactly opposite (PD+180º) while the participant underwent OKN, with gaze near the center of the background stimulus at various speeds in the same direction, as illustrated for a positive speed (parallel to PD) in Figure 1D. In this imagined task, we specify that participants would then be asked to pursue a new target, interrupting OKN, though this final part of the simulated task is unimportant for investigating how the brain might add together retinal target motion and 3D eye motion in order to reconstruct spatial target motion.

We computed the separability index of our network HLUs, which captures the extent to which inputs are coded by units either in a combined (inseparable) or separable (independent) fashion, based on previous electrophysiological and network studies (see Methods; Buneo et al.
The pseudo-color plots presented Figure 6A and B show this analysis for 4 example HLUs from the 100 HLU network (#s 74 and 23 from the first hidden layer; 38 and 72 from the second hidden layer). In these plots, we show the normalized, color-coded activation of each unit across all combinations of target (abscissa) and eye velocities (ordinate) for target velocities in the PD+180° (left) and in the PD (right). Examining the change in activation of each unit across each axis reveals the dependence on either target or eye speed. For example, units 74 and 23 in panel A show a strong dependence on target speed, appearing as gradients in the horizontal (0°) direction, while unit 72 in panel B showed a distinct separability for each target velocity direction. In the null (-) direction this unit showed strong target speed dependence while in the preferred (+) direction it showed strong eye speed dependence (180°) direction. Granted, these gradient directions were not perfectly separable for each unit, indicating some eye speed dependence. Additionally, unit 23 displayed modulation gradients resulting from combined target-eye coding but in opposite directions, resulting in a gradient direction of 0° that appeared purely separable when in truth was not (see Methods for details). To overcome this limitation in these particular units, at the target velocity of minimal (magenta) or maximal activity (green; as one of these locations typically represented the border between each gradient direction) we split the velocity field from which we sampled the characteristic gradient direction into two fields (black outline). Finally, unit 38 in panel B shows inseparable coding both in the compensatory (T-E; -45° gradient) and anti-compensatory ways (T+E; -45° gradient), depending on the direction of the target velocity.

In Figure 6C and D, we show the distributions of the separability indices for each layer (Buneo et al. 2002; Pesaran et al. 2006; Blohm et al. 2009; Blohm 2012; Bremner and Andersen 2012).
Each polar histogram is labeled at 0º, 90º, 180º and 270º as coding for retinal target motion (T), coding in an anti-compensatory way (T+E), coding for eye motion (E) and coding in a compensatory way (T-E), respectively. Figure 6C shows that first hidden layer units showed a mostly separable (target) coding, whereas many second hidden layer units (Figure 6D) exhibited some target coding but with a large proportion of units displaying inseparable coding, intermediate between purely retinal (T) and spatially accurate (T-E). Finally, this gradient analysis revealed very few units in the first hidden layer and some units in the second hidden layer coding motion inputs in a spatially correct fashion (T-E), which is consistent with other visual input simulations (see Figures 6B, 7 and 8). Group level t-tests on the average gradient directions across network sizes suggested a retinal coding in the first hidden layer (t(7) = 1.00, p = 0.35, 95% CI [-7.80, 19.1]), but not as strongly in the second hidden layer, as given by its wide 95% CI (t(7) = 1.29, p = 0.24, 95% CI [-37.0, 125]), with average gradients for the first hidden layer from -9.8° (16 HLU network) to 38.3° (25 HLU network), though this mean gradient was not representative of the unit gradients (dotted radius in Fig. 9: t(99) = 2.09, p < 0.05), and average gradients for the second hidden layer from -133.4° (49 HLU network) to 144.5° (25 HLU network). In the first hidden layer, SDs ranged from 59.5° for the 100 HLU network to 98.1° for the 25 HLU network while, in the second hidden layer, SDs ranged from 88.6° for the 100 HLU network to 111.1° for the 9 HLU network. These results suggest that, although the first hidden layer units typically coded in a retinal, target-dependent fashion, the second hidden layer units also exhibited a dependence on eye speed, indicating a non-retinal, intermediate code. We will discuss these results in greater detail and their implications for pursuit-related areas of the brain in the DISCUSSION.
To this point, our simulated electrophysiological experiments have revealed converging findings concerning the input reference frame of our network HLUs: (1) first hidden layer units code inputs almost exclusively according to a retinal reference frame with mainly separable retinal and eye motion signals, and (2) second hidden layer units code inputs according to both a retinal reference frame and an intermediate reference frame with inseparable motion coding.

Output properties: Motor fields

We next sought to determine the reference frames in which HLUs code motor output. One might assume that units’ visual and motor tuning should be aligned in a network performing visuomotor transformations (such that the input and output reference frames would be the same); however, there is evidence supporting the idea that network units involved in visuomotor transformations code visual input and motor output tunings according to different reference frames (Salinas and Abbott 1995; Pouget and Sejnowski 1997; Blohm et al. 2012), i.e. if a unit is involved in transforming information, then the input and output codes should differ.

For example, one can consider head roll-induced OCR. As we investigated earlier in the context of visual input tuning properties, units must be modulated by head roll-induced OCR in order to compensate for the spatial misalignments between retinal input and the required pursuit output. We can take an alternative approach as well, noting that the network’s motor output vector (its motor contribution) must also compensate for spatial misalignments between retinal velocity input and the required pursuit output. We computed how each unit’s activity was correlated with the 3D motor output, known as the motor field of each unit (Smith and Crawford 2005; Keith et al. 2007; Blohm et al. 2009). To discern motor field reference frames, we mapped
units’ motor fields and determined how they were modulated specifically by head roll-induced
OCR, oblique gaze and the half-angle rule.

In contrast to VTCs, which specified the activation of units to every possible 2D retinal
velocity, the motor field is each unit’s activity for every possible motor output, spanning all three
spatial dimensions. Essentially, motor fields describe the motor output tuning of each neuron. As
such, we can examine how the motor field of each unit changes with eye and/or head inputs in
order to see the unit’s motor field reference frame. Figure 7 presents the motor field for an
example unit from the first hidden layer (#16) of the 100 HLU network. Unit activations are
shown in 3D, head-fixed space as viewed from behind (upper left, frontoparallel horizontal-
vertical plane), from below (lower left, transverse horizontal-torsional plane) and from the side
(upper right, sagittal vertical-torsional plane). Because the same motor outputs can be produced
from an infinite combination of inputs, activations were binned based solely on motor output —
collapsing across all retinal, eye and head input signals giving rise to that output. This resulted in
a distribution of activities for each bin, and in Figure 7A we present the color-coded average
activities within each bin. In this way, each viewpoint of the motor field resembled the unit’s true
output-aligned activation. We also summarize the activation profile along each axis in the plots
adjacent to each pseudo-color plot axis (error bars represent SE across each dimension). To
capture the three-dimensional preferred motor tuning of each unit, we computed the activity-
weighted 3D motor field center-of-mass (COM\textsubscript{MF}), which is represented in Figure 7A as a white
disk. We also computed the 2D frontoparallel motor field (by averaging across torsional output
velocities), which we show in Figure 7B plotted on polar axes for the same HLU (#16). Note that
the tuning shown in the 2D frontoparallel motor field is representative of the tuning for the
frontoparallel plane (Figure 7A, upper left). The red bar represents the preferred motor tuning
direction of this unit ($PD_{MF}$) and the black bars represent the SD of activity for each directional
output bin.

When determining the motor field reference frames of units, we examined how shifts
either of the COM$_{MF}$ or of the $PD_{MF}$ accounted for the spatial requirements of the visuomotor
transformation, analogous to how we used shifts of the preferred visual tuning to delineate the
input reference frames of units. In this section we will review the findings of three experimental
simulations that we used to find the motor field reference frame of each HLU, each representing
the corresponding simulation performed in the Input Properties section.

Motor field reference frame analyses for experiments 1, 2 and 3

We have thus far determined that, in general, units in the first hidden layer of our network code
the visual input independently of OCR while the units in the second hidden layer code visual
input while partially compensating for the 2D rotational effects of OCR (by accounting for head
roll). However, because our network compensates for OCR to produce spatially correct smooth
pursuit (see performance in Figure 3), we must instead examine how the outputs of each unit
accounted for OCR in order to fully characterize the mechanisms underlying the transformation.

To see how the network accounted for 2D retinal rotations due to OCR, we examined how each
unit’s $PD_{MF}$ was modulated by head roll-induced OCR.

For a unit’s output to compensate for head roll-induced OCR its motor contribution
should remain consistent regardless of the OCR angle. Conversely, a unit that coded its output
according to retinal coordinates would exhibit shifts in an equal and opposite direction relative to
the signal for which it must compensate. We present the effects of OCR on the frontoparallel
motor field for two example units (63 and 88) from the first and second hidden layers of our 100
HLU network in Figure 8A (respectively). Here, the pseudo-color plots represent the change in $P_{D_{MF}}$ across changes in head roll (abscissa) and OCR (ordinate), using the identical conventions as the pseudo-color plots in Figure 5A. Across changes in head roll and OCR, for unit 63 in the first hidden layer there was a shift of the $P_{D_{MF}}$ by approximately an equal and opposite angle of head roll and OCR, indicating that this unit coded motor outputs according to a retinal frame, as confirmed by multiple regression analysis (see red point in Figure 8B). Alternatively, for unit 88 in the second hidden layer there was no shift of the $P_{D_{MF}}$ across head roll and OCR, indicating that this particular unit coded according to an approximately spatial frame (see blue point in Figure 8B). Repeating this multiple regression analysis for all network units yielded similar results between layers. Figure 8B shows the distributions of these regression gains, along with their comparisons with the spatially correct prediction (intersection of solid lines) and retinal prediction (intersection of dashed lines).

This analysis revealed that units in each layer coded motor output according to both the spatial and retinal predictions, but also according to frames intermediate to both retinal and spatial. The various reference frames used by units can be seen in Figure 8B, which shows units in both hidden layers clustering around either the retinal or spatial hypotheses (see intersections of dashed lines and of solid lines, respectively), and several units coding for intermediate reference frames. Because of this clear bimodality, using statistical analyses based on median gains was not as revealing as in previous analyses, though for both layers these statistical analyses agree with the qualitative observation that units in both layers coded neither according to retinal nor spatial frames during both head roll and ocular torsion. However, these analyses could not distinguish between the reference frame distributions of each layer (for both layers and both head roll and ocular torsion, all group-level t-tests $p < 0.01$ and there was no consistent
change in variability between layers of any network size). Like in the visual tuning analysis, the concomitant head roll and OCR dependence suggests that units in both layers showed evidence of a learned internal model of the head roll-OCR interaction.

In each of the other two experiments (oblique eye orientations and half-angle rule), HLUs exhibited motor field coding frames consistent with those found in the head roll-OCR simulation, as shown in the histograms in Figures 8C and 8D (following the same conventions as Figure 8B). Thus, HLUs displayed a variety of reference frames that are – at the group level – sufficient to recover spatially accurate motor plans (since our networks produce accurate movements).

**Output properties: Simulated microstimulation**

In contrast with the motor field analysis, simulated microstimulation allowed us to investigate each unit’s specific contribution to motor output while also activating a unit’s downstream connectivity, instead of only correlating its activity with the output. To do this, unit-by-unit we examined how increases in each unit’s activity affected network output by first setting the required network output to zero (by foveating the retinal target in the network inputs), then “evoking” network outputs by artificially setting each unit’s activity to 5. Thus the output vectors corresponded to an amplification of each unit’s normal contribution to the network output. We then repeated this process under three transformation contexts (i.e. head roll-induced OCR, retinal rotations from oblique gazes and the half-angle rule) and compared changes in the network output vectors to those required for each transformation to reveal the reference frame of each unit’s motor output vector.

When determining the output reference frames of units, we examined how shifts of the evoked network output accounted for the spatial requirements of the visuomotor transformation,
analogous to how we used shifts of units’ preferred visual and motor field tunings to delineate the unit-by-unit input and motor field reference frames. In this section we will review the findings of three experimental simulations that we used to find the output reference frame of each HLU, each representing the corresponding simulations in the input and motor field reference frame analyses.

Simulated microstimulation reference frame analyses for experiments 1, 2 and 3

In addition to our motor field analysis, which found that units compensated for head roll-induced OCR primarily according to mixed, intermediate output reference frames in each hidden layer, we also wanted to see how changing the eye-head geometry influenced units’ contributions to motor output and how these effects compensated for OCR while accounting for downstream network connectivity. To do so, we used simulated microstimulation and compared changes to the “evoked” network output to those required for a spatial or retinal coding.

Typical evoked gaze velocities across head roll and OCR for units in the first (15, blue) and second (93, green) hidden layers can be seen in Figure 9A. Each point represents the tip of the on-screen gaze velocity vector evoked by microstimulation at a specified head roll-OCR combination. As such, the grid-like patterns seen in these plots represent the entire head roll-OCR space across which we performed multiple regression analysis, as described in the following paragraph.

The network had to compensate for OCR under conditions of head roll in order to produce spatially correct pursuit – therefore, because our network generated spatially accurate pursuit, a complete compensation would reflect regression gains between OCR (or head roll) and microstimulation output equal to 0 while no compensation (implying a retinal coding of outputs)
would correspond to regression gains equal to -1. However, as we have discussed in earlier sections, because head roll and OCR were correlated in our training set, either signal could theoretically be used by the network to carry out the transformation. To account for this potential effect we performed multiple regression analysis between the required rotation angles (head roll and OCR) and the evoked eye movement directions. For the typical units presented in Figure 9A, one can easily see that the unit from the first hidden layer (15) exhibited a wider spatial distribution of evoked movements across head roll and OCR compared to those evoked when the unit from the second hidden layer (93) was stimulated. This result is indicative of an intermediate coding by unit 15, as its evoked movements are modulated by head roll and OCR but not in a directly compensatory way. On the other hand, the narrow field of evoked movements for unit 93 is indicative of a spatial coding. These units were representative of the regression findings for first and second hidden layer units.

Figure 9B shows the resulting multiple regression gains of evoked movement shifts relative to head roll and OCR. We performed these regression analyses on a unit-by-unit basis for each network size and show the resulting histograms, along with the spatial (intersection of solid lines) predictions, in Figure 9B. This analysis revealed that first hidden layer units coded motor outputs according to either spatial or intermediate reference frames while second hidden layer units coded outputs mostly according to a spatial reference frame, though with some intermediate coding also. Across head roll, group level t-tests on median gains were not significantly different from the spatial hypothesis (hidden layer 1: \( t(7) = -0.48, p = 0.65, 95\% \text{ CI } [-0.01, 0.01] \); hidden layer 2: \( t(7) = -1.18, p = 0.28, 95\% \text{ CI } [-0.01, 0.004] \)). Similarly, across OCR, group level t-tests on median gains were not significantly different from the spatial hypothesis (hidden layer 1: \( t(7) = -0.29, p = 0.78, 95\% \text{ CI } [-0.03, 0.02] \); hidden layer 2: \( t(7) =
0.63, p = 0.55, 95% CI [-0.01, 0.01]). However, for both head roll and ocular torsion the interquartile gain distances from the first hidden layer were significantly larger (compared to those of the second hidden layer) across all network sizes (head roll: t(7) = 3.92, p < 0.01, 95% CI [0.03, 0.12]; ocular torsion: t(7) = 5.50, p < 0.01, 95% CI [0.10, 0.24]), suggesting that the first hidden layer was coding according to a more intermediate frame than the second hidden layer. Therefore, when retinal signals were rotated relative to the head due to OCR (Blohm and Lefèvre 2010), this analysis revealed that, when accounting for network connectivity, first hidden layer units coded motor outputs according to both spatial and intermediate frames while second hidden layer units coded motor outputs according to a primarily spatial reference frame. Additionally, like in the other head roll-induced OCR simulations, the finding of an effect for both head roll and OCR suggests at least a partial learning of the head roll-OCR interaction.

In each of the other two experiments (oblique eye orientations and half-angle rule), HLUs exhibited microstimulation coding frames consistent with those found in the head roll-OCR simulation, as shown in the histograms in Figures 9C and 9D (following the same conventions as Figure 9B).

**Summary of gradual visuomotor transformation**

Using each of these analyses, we determined the input and output coding schemes of each unit within our network, enabling us to trace the visuomotor transformation from retinal input to spatially correct motor output. Figure 10 represents a summary of all analyses and hidden layer unit gains, plotted as quartiles relative to the retinal and spatial predictions for visual tuning (orange), motor field (purple) and microstimulation analyses (blue). We summarize the HLU reference frames for each network size (see legend, top right) and collapsed across each
experimental simulation (head roll-induced OCR, oblique gaze-induced retinal rotations or the
half-angle rule). We found that units in the first hidden layer of our network coded visual
velocity inputs according to a retinal reference frame while units in the second hidden layer
coded visual velocity inputs according to both a retinal frame and a mixed, intermediate frame.
Additionally, motor field analyses revealed that the motor tunings of units in the first and second
hidden layers were coded according to retinal, spatial and several mixed, intermediate frames.
When accounting for downstream connectivity, simulated microstimulation analyses revealed
that units in the first hidden layer coded motor outputs according to both spatial and intermediate
reference frames, while units in the second hidden layer coded motor outputs primarily according
to a spatial reference frame. Therefore, one of the key characteristics that we observed in both
our input and output analyses is that, in either case, units in at least one layer coded according to
various mixed reference frames, neither in retinal nor spatial coordinates.

To see if these intermediate coding schemes were consistent on a per-unit basis, we
performed a meta-analysis of unit gains in the layers showing intermediate properties (in the 100
HLU network). In the first hidden layer, we found a significant correlation between units’ motor
field reference frames during the oblique gaze simulation and the half-angle rule simulation
(slope = -0.19, r² = 0.24, p < 0.01). For the second hidden layer, we found significant
correlations between units’ visual tuning reference frames during the oblique gaze simulation
and the half-angle rule simulation (slope = 0.26, r² = 0.18, p < 0.01), and between units’ motor
field reference frames during the oblique gaze simulation and the half-angle rule simulation
(slope = 0.14, r² = 0.05, p < 0.05). These significant correlations between units suggest that, in
some special cases, a given unit utilized a common intermediate reference frame which
generalized across visuomotor contexts. However, these three significant cases represent a
minority of all 12 possible comparisons between intermediately coding network layers, suggesting that units code according to a reference frame which depends on the exact context of the visuomotor transformation. It would be interesting to see how the input and output properties outlined in this report correlate with the neurophysiological properties of neurons in areas of the brain thought to be involved in the visuomotor transformation for smooth pursuit (Blohm and Lefèvre 2010), such as MT and MST. The simulation results presented here provide several testable predictions for the neurophysiological properties electrophysiologists might expect to find when investigating the responses of neurons involved in smooth pursuit under different 3D transformational contexts.

Discussion

We designed and trained a physiologically-inspired four-layer, feed-forward network model to simulate the transformation of retinal motion signals into spatially correct smooth pursuit commands. After training had completed, we found that the feed-forward model could perform the spatially correct 3D visuomotor velocity transformation for smooth pursuit while obeying Listing’s law (Blohm and Lefèvre 2010). To carry out this transformation units exhibited both gain modulation and tuning shifts – two properties observed in electrophysiological studies of pursuit-associated neurons (Inaba et al. 2007; Chukoskie and Movshon 2009; Fujiwara et al. 2011; Inaba et al. 2011). Retinal velocity tuning, motor tuning, and simulated microstimulation analyses revealed that hidden layer units carried out the transformation in a gradual fashion from retinal (first hidden layer inputs) to spatial (second hidden layer microstimulation outputs). Thus, we suggest that our network units used gain modulation to differentially weight each unit’s input-output contribution to the transformation, allowing the network to generate a spatially
accurate pursuit command. To our knowledge, these findings are consistent with all known
electrophysiological properties of pursuit-related areas of the brain, and could thus provide a
mechanistic explanation for the presence of eye and head orientation signals in these areas. We
will discuss these points in greater detail below.

General discussion

We trained our network model to perform the general transformation of 2D retinal signals for
pursuit with any combination of 3D eye and head geometries and found that, to do this, the
network utilized several emergent computational strategies. First, the HLUs carried out the
complex, nonlinear pursuit transformation in a distributed fashion. As our tuning shift analyses
revealed, units coded input motion according to various reference frames that were neither retinal
nor spatial (mixed intermediate frames) while they also coded motor outputs according to a
similar though more consistently spatial frame. Therefore, one single unit alone could not
account for the full, spatially correct 3D transformation of retinal motion into a head-centered
pursuit command.

Both in our network and in the real brain, intermediate reference frames are often
reported (Chang and Snyder 2009; McGuire and Sabes 2011). Those can even lie beyond the
limits of the expected range (e.g. over-compensating or anti-compensating) and often the
question about their meaning arises. When performing such reference frame analyses on (real or
simulated) neural data, one should always keep in mind that neurons do not care about reference
frames; they participate in a certain computation and all that counts is the end result, regardless
of how it is achieved. Thus, artificially assigning a reference frame to a unit is questionable but
unfortunately correlating neural activity with measurable quantities from the physical world is
the only way we have to probe brain function. Therefore, reference frame analyses can be used to
probe network mechanisms that cannot otherwise be quantified, in particular when investigating
sensory-to-motor transformations.

We also found that HLUs’ tuning curves were modulated by eye and head orientation and
velocity signals in a way that is consistent with gain field theory for reference frame
transformations (Blohm and Crawford 2009). Thought to be the optimal way for feed-forward
network models to compute nonlinear sensorimotor transformations (Zipser and Andersen 1988;
Salinas and Abbot 1995; Salinas and Abbott 1996; Pouget and Sejnowski 1997; Xing and
Andersen 2000; Smith and Crawford 2005; Blohm et al. 2009; Chang et al. 2009; Blohm 2012),
the finding that gain fields were also used to account for the requirements of Listing’s law
suggests that the brain could theoretically utilize gain fields when adhering to Listing’s law
during the transformation for smooth pursuit (Blohm and Lefèvre 2010).

Network units also shared many activation properties with neurons recorded in areas MT
and MST. First, the visual tuning of HLUs were complex, often with more than one peak of
activity representing more than one “preferred” direction and areas of lower activity between
those peaks (see Figure 3). Neurophysiological recordings have found very similar velocity
tuning properties in area MT neurons (Richert et al. 2013), suggesting that the network units
might encode visual inputs in a similar fashion as MT neurons. In our network we also observed
the use of gain modulation to generate smooth pursuit, similar to the gain fields that have been
found in electrophysiological recordings from areas MT and MST during pursuit tasks. For
example, visual and pursuit-related activity of MT and MST neurons are gain modulated by eye
orientation (Bremmer et al. 1997; Lee et al. 2011) and by pursuit velocity (Chukoskie and
Movshon 2009; Inaba et al. 2007; Inaba et al. 2011). Additionally, gain fields have been
implicated in the reference frame transformations for reaching (Galletti et al. 1995; Batista et al. 1999; Blohm et al. 2009; Chang et al. 2009) and for depth (Bhattacharyya et al. 1999; Ferraina et al. 2009; Blohm 2012), as parietal neurons are gain-modulated in response to eye orientation changes and changes in hand position (Galletti et al. 1995; Batista et al. 1999; Chang et al. 2009; Ferraina et al. 2009; Bhattacharyya et al. 2009). If the brain indeed carries out the visuomotor velocity transformation for pursuit in the same way as our network model, the agreement between emergent properties of our network model and electrophysiological findings suggests that gain modulation in areas MT and MST may play a key role in integrating extraretinal signals into the motor plan.

Analogous to the behavior of units within the second hidden layer, shifts of neuronal visual tuning functions have been observed in area MST. Specifically, shifts have been observed during pursuit under conditions of self-motion (Bradley et al. 1996; Page and Duffy 1999; Shenoy et al. 1999; Shenoy et al. 2002) and during pursuit of an on-screen stimulus at a fixed depth (Inaba et al. 2007; Chukoskie and Movshon 2009; Inaba et al. 2011). Tuning shifts have also been reported after changes in eye, head and body orientation in area MST (Fujiwara et al. 2011) and in the frontal eye fields (FEF; Kurkin et al. 2007). Furthermore, these shifts corresponded to coding schemes in not only retinal and spatial frames, but also in mixed, intermediate reference frames, which is not unlike the mixed coding schemes observed across eye orientation shifts in area MST (Bremmer et al. 1997), in the lateral intraparietal area (LIP) during saccades to remembered auditory locations (Stricanne et al. 1996), in the ventral intraparietal area (VIP) across eye orientation shifts (Duhamel et al. 1997), in the dorsal premotor cortex during reach (Batista et al. 2007) and in previous computational work (Xing and Andersen 2000; De Meyer and Spratling 2013). Taken together, these findings suggest that shifts
of neuronal tunings in areas MST and FEF might represent a distributed mechanism for the compensation for distortions and rotations to retinal information, using properties that the second layer of our network model reproduced spontaneously. Interestingly, previous feed-forward network models (Salinas and Abbott 1995; Salinas and Abbott 1996) have also found that the presence of these tuning shifts are indicative of gain modulation at earlier processing stages of coordinate transformations.

While, to our knowledge, detailed mapping of neuronal motor fields has not been performed for MT or MST neurons, there have been several studies (Komatsu and Wurtz 1989; Groh et al. 1997; Britten and van Wezel 1998; Born et al. 2000; Ilg and Schumann 2007) which investigated microstimulation of neurons in these areas and the effects on smooth pursuit. Another set of studies (Salzman et al. 1990; Salzman et al. 1992; Celebrini and Newsome 1995; Britten and van Wezel 1998) has investigated the effects of microstimulation on perceived heading or motion direction. Microstimulation of area MT neurons has been shown to bias both perception of motion towards the preferred retinal tuning of those neurons (Salzman et al. 1990; Salzman et al. 1992; Born et al. 2000) and pursuit velocity in the anti-preferred retinal velocity direction, or the pursuit direction typically required to minimize retinal slip (Komatsu and Wurtz 1989; Groh et al. 1997; Born et al. 2000). Similarly, simulated microstimulation of our first hidden layer units elicited network output velocities that were biased towards the units’ preferred retinal velocity tunings. On the other hand, microstimulation of area MST neurons has been shown to bias heading perception (Britten and van Wezel 1998) and motion direction perception (Celebrini and Newsome 1995) as well as pursuit direction toward the preferred spatial tuning (Ilg and Schumann 2007) or toward the recording site of those neurons (Komatsu and Wurtz 1989), which aligns well with the spatial constancy of the network outputs when we stimulated
units in the second hidden layer. Additionally, microstimulation studies of areas FEF (Tanaka and Lisberger 2002; Gottlieb et al. 1993), SEF (Missal and Heinen 2004) and the cerebellar vermis (Krauzlis and Miles 1998) have each found that pursuit outputs correspond to the stimulated neurons’ spatial preferences, suggesting that these areas could alternatively be considered functional analogues of the second hidden layer of the network model.

Another emergent property of the network is that compensation for head roll-induced OCR depended on both head roll and OCR signals during the visual tuning, motor field and microstimulation reference frame multiple regression analyses. This suggests that the network was able to learn an internal model of the head roll-OCR interaction, even though eye orientation was only partially dependent on head orientation in the training set. If the brain also uses an internal model of head roll-induced OCR to compensate for ocular torsion, this might explain why there is little direct electrophysiological evidence of compensation for OCR in pursuit-related neurons (Kurkin et al. 2007; Fujiwara et al. 2011).

The presence (or absence) of ocular torsional signals in cortex raises some questions about how the brain might plan geometrically correct pursuit movements (Blohm and Lefèvre 2010). Theoretically, in order to correctly interpret retinal input an estimate of ocular torsion must be, and is, incorporated (Blohm and Lefèvre 2010; Murdison et al. 2013; Klier and Crawford 1998; Leclercq et al. 2013). Traditional thinking is that a 2D motor command is transformed into a 3D movement at the level of the brainstem and/or extraocular muscles, and while this might be true, this does not preclude the visual system from requiring knowledge of torsion in order to correctly interpret primary visual information (for review see Klier et al. 2013). But what is the source of these signals? A simple explanation could be that cortical torsional signals might simply have gone undetected in electrophysiological studies; an effect
potentially due to the fact that ocular torsional signals (orientation and velocity) are typically small in magnitude, making them more difficult to detect. In this case, these explicit signals would presumably be included in any 3D motor plan. Another potential explanation comes from the idea that ocular torsion does not have to be explicitly coded, but could instead be the result of an internal model of head and eye orientations and velocities (i.e. based on the other signals present). As long as cortex has a model of the interactions between eye and head signals, it could form an implicit estimate of ocular torsion. We accounted for the correlated nature of the head and eyes when we analyzed both head roll and ocular torsion as free variables in our regression models, and found that our networks spontaneously used eye-head interactions to their advantage when generating pursuit movements. Therefore, our model predicts that the brain would utilize all available signals about 3D eye and head geometry, regardless of their source, when generating pursuit movements.

Comparison to previous models

Gain modulation is an efficient way for feed-forward networks to compute complex coordinate transformations (e.g. Blohm and Crawford 2009). The gain field mechanism used by our network echoes previous feed-forward network models that also employ gain fields to perform coordinate transformations (Zipser and Andersen 1988; Salinas and Abbott 1995; Salinas and Abbott 1996; Pouget and Sejnowski 1997; Xing and Andersen 2000; Deneve et al. 2001; Smith and Crawford 2005; Keith et al. 2007; Blohm et al. 2009; Chang et al. 2009; Blohm 2012). Here, we show that the transformation of 2D retinal velocity (and position) signals into 3D smooth pursuit movement plans can be carried out by a physiologically-inspired feed-forward network model using gain modulation. This finding further supports the notion of gain fields being the primary

Though there have also been several models of the roles of areas MT and MST during smooth pursuit (Pack et al. 2001; Furman and Gur 2003; Furman and Gur 2005; Shibata et al. 2005), during coordinated saccades and pursuit (Grossberg et al. 2012), during motion perception (Cameron et al. 1998; Furman and Gur 2005) and in reference frame transformations (Dicke and Thier 1999), our model represents the first time that the pursuit transformation has been performed in three dimensions, as previous models including a pursuit component have only carried out transformations in one (Pack et al. 2001; Shibata et al. 2005; Grossberg et al. 2012) or two dimensions (Furman and Gur 2003; Furman and Gur 2005). These models could not perform the general, 3D transformation that our model performs here which accounts for 2D retinal signals, 3D eye orientation and velocity and 3D head orientation and velocity, and instead they only accounted for the vector summation of retinal and extraretinal signals.

Additionally, some of these network studies relied on pre-determined connectivity between model neurons in MT, MST, visual and/or motor cortical areas (Pack et al. 2001; Furman and Gur 2003; Furman and Gur 2005; Grossberg et al. 2012), whereas our network only relied on a pre-determined processing architecture of three feed-forward connection matrices, which were self-organized during training, between four processing layers. This architecture allows for a more general description of the mechanisms underlying transformations in the brain, and does not limit it to neurons in pre-specified areas such as MT and MST, though here we argue that our model accounts for the activation properties of neurons in these areas solely as a
result of learning the 3D transformation for pursuit. Finally, to our knowledge, our network model is the first to account for the combined coding of retinal position and velocity, as found in MT (Gattass and Gross 1981; Richert et al. 2013), though only in the context of smooth pursuit.

To our knowledge, only one other model created by Smith and Crawford (2001) performed the 3D transformation for saccade generation using a neural network framework similar to ours, though there were several differences between their study and ours. These differences consisted of distinct saccade and pursuit pathways, model design and major findings. First, they investigated the position transformation underlying saccade generation with only retinal error and gaze orientation inputs (generating a motor error output) whereas we investigated the velocity transformation for smooth pursuit with retinal, eye and head orientation and motion inputs (generating an eye velocity output). While these pathways share some neural circuitry, they are ultimately distinct (for review, see Krauzlis 2004), especially when also considering head orientation and motion signals. Aside from the design choices coming from the differences between saccade and pursuit generation (e.g. their use of retinal position compared to our retinal position and velocity), they utilized different eye orientation codes. We used push-pull coordinates (similar to motor neurons guiding the head-centered pairs of extraocular muscles) while they utilized 3D angular vector coordinates (representing the spatial vectors required for movement). Finally, the analyses they used to investigate how units in their model’s hidden layer carried out the reference frame transformation consisted of simulated microstimulation, simulated lesions and computing the sensitivity vectors for each unit, with no investigation of gain fields as we did here. Their main findings were also different from ours; instead of units carrying out the transformation in a fully distributed fashion as ours did, their units spontaneously organized into three major classes, each being responsible for a different
aspect of the transformation. As such, our model provides novel predictions for circuitry underlying the generation of smooth pursuit movements while accounting for eye and head geometries, and these differences suggest that these mechanistic properties are distinct from those underlying the saccadic transformation.

Predictions and limitations

Because our network model performs the general 3D transformation of 2D retinal signals for smooth pursuit, it can be used to make many testable predictions about the neurophysiological properties of areas involved in the transformation for smooth pursuit. As we observed, the network model uses gain modulation to accomplish the transformation – a property which has been observed in numerous visuomotor areas in the brain (Bremmer et al. 1997; Lee et al. 2011; Chukoskie and Movshon 2009; Inaba et al. 2007; Inaba et al. 2011; Galletti et al. 1995; Batista et al. 1999; Chang et al. 2009; Ferraina et al. 2009; Bhattacharyya et al. 1999), some of which are involved in pursuit (Bradley et al. 1996; Bremmer et al. 1997; Page and Duffy 1999; Shenoy et al. 1999; Shenoy et al. 2002; Inaba et al. 2007; Chukoskie and Movshon 2009; Inaba et al. 2011; Fujiwara et al. 2011; Lee et al. 2011) – suggesting that the brain may use a similar mechanism for carrying out the transformation. Under this assumption, we hypothesized that areas MT and MST are represented by our model’s first and second hidden layers, respectively, though in the real brain the transformation may be carried out by numerous other areas. If this is in fact the case, the model predictions would still hold but we would expect a more gradual transformation throughout these areas using gain modulation.

One of our model’s main predictions regards the task-dependent, mixed coding of visual inputs and motor outputs in areas involved in the transformation. Because we found that units in
the first hidden layer code retinal signals independently of extraretinal signals, our model predicts that neurons early in the processing of the transformation should exhibit retinal velocity tuning which is purely gain modulated by eye and head orientation. On the other hand, because we found that the visual tunings of units in the second hidden layer shift with extraretinal changes, our model predicts that neurons in subsequent processing stages should be similarly dependent on eye and head orientation. This prediction, though present in each of our analyses, is clearly exemplified in our separability analysis (Buneo et al. 2002; Pesaran et al. 2006; Bremner and Andersen 2012) during OKN (Fig. 9), revealing that units in the first hidden layer typically code according to retinal target signals, while second hidden layer units typically code according to a mixture of retinal and eye-in-head motion signals. This prediction for our first hidden layer fits well with the findings in areas MT (Inaba et al. 2007; Chukoskie and Movshon 2009; Inaba et al. 2011); however, the separability of retinal and extraretinal signals has never been explicitly tested in area MST. Despite this, Lee and colleagues (2011) found that MST neurons account for eye motion when coding heading direction during pursuit, a finding which is compatible with the inseparability of retinal and eye-in-head motion signals in our model. Taken together, this electrophysiological evidence (Bradley et al. 1996; Bremmer et al. 1997; Page and Duffy 1999; Shenoy et al. 1999; Shenoy et al. 2002; Inaba et al. 2007; Chukoskie and Movshon 2009; Inaba et al. 2011; Fujiwara et al. 2011; Lee et al. 2011) and the findings of our model suggest that the 3D transformation for pursuit could be fully accounted for by areas MT and MST.

Moreover, we found that the intermediate coding schemes of units in each simulation were usually uncorrelated with one another (in 9 of 12 possible comparisons between simulation gains), implying that the exact contribution of each unit to the transformation depended on the current task. Task-dependent coding has been theorized to be used by neurons involved in
several different transformations (Pouget and Sejnowski 1997) and in fact may be an efficient way for the brain to carry out complex transformations across multiple areas (e.g. Boussaoud and Bremmer 1999). Thus, our network model predicts that networks of real neurons could use similar intermediate coding schemes when performing transformations in a distributed fashion. If MT and MST are functionally equivalent to the layers of our network, our simulations predict that the visual tunings and motor tunings should shift with eye and head orientation, though not necessarily according to any specific reference frame nor consistently across tasks.

When considering these predictions, it is important to consider the limitations of our model – some of which are similar to those of previous feed-forward network models (Blohm et al. 2009; Blohm 2012). First, the network only performs the transformation for the initiation, or “open-loop”, portion of smooth pursuit (Blohm and Lefèvre 2010; e.g. Ilg 2008 or Lisberger 2010). Therefore, the transformation during the minimization of retinal slip during ongoing smooth pursuit (i.e. once pursuit is driven primarily by extraretinal signals) is beyond the scope of this model. Next, in the brain sensory information about the eyes and the head either arrive to visuomotor areas via proprioception, vestibular inputs or efference copies, and we did not distinguish between these possibilities here. As a result, if, for example, head velocity is coded both via proprioception and efference copies the CNS would have to solve the multisensory integration problem (while also accounting for relative delays), which we did not consider for purposes of this model (Sober and Sabes 2003; Sober and Sabes 2005; McGuire and Sabes 2009; Burns and Blohm 2010). Additionally, we constructed our network model to be fully feed-forward, rate-based and static – three assumptions which are almost never valid in the brain. In the brain there are recurrent connections between neurons, neurons follow spike codes and perform computations in a time-dependent, dynamic way. Also, the network architecture does
not follow true cortical structure. We chose to implement two hidden layers, but in the brain the transformation might be carried out across several neural areas, potentially influencing the performance of the network (Hermundstad et al. 2011).

Simulated microstimulation also presented two main limitations. Because we did not model downstream pursuit circuitry such as the omnipause neurons (OPNs), our model does not employ any gating mechanisms (e.g. the OPNs, whose activities are lowered during pursuit) that might potentially account for the fact that microstimulation effects are seen only during ongoing pursuit (Komatsu and Wurtz 1989; Groh et al. 1997; Born et al. 2000; Ilg and Schumann 2007). However, microstimulation-evoked smooth pursuit under a constant velocity (in the visual preferred direction) produced qualitatively identical network output vectors to those evoked from fixation, though offset by the initial eye velocity. Additionally, MT has been found to contain both local motion and wide field motion detectors (Born and Tootell 1992; Born 2000), each of which evoke distinct eye movements when stimulated, potentially due to the respective preferences for target and background motion (Born et al. 2000) and their respective projections to the either the dorsal or lateral subregion of MST (Berezovskii and Born 2000; Komatsu and Wurtz 1988). Our self-organizing network model could not reproduce these neuron types because the simulated pursuit task used to generate the training set contained only a point motion target which was pursued in complete darkness (i.e. there was no background motion in the opposite direction projected onto the retina), thus presenting one potential area for expansion.

Finally, the training method used might have slightly influenced the detailed emergent properties of the network, though we expect qualitatively similar results using other training algorithms (Blohm et al. 2009). Thus, these limitations provide several potential extensions of our model for future work. Astonishingly, despite these limitations and abstractions from the real
cortical network, there was a striking similarity between our model units and known pursuit 
neuron properties in the brain. This not only validates our approach but also indicates that such 
simple feed-forward models might be good tools to understand the principal mechanisms 
underlying sensory-to-motor transformations for pursuit.

In summary, we have shown that a simple feed-forward network model can carry out the 
3D, spatially correct transformation underlying smooth pursuit while following Listing’s law 
(Blohm and Lefèvre 2010). The network model does so using eye- and head-dependent gain 
modulation to weight visual tuning, resulting in shifts of downstream visuomotor tuning, and 
generating a motor plan which fits the spatial task requirements. Thus, we provide a mechanistic 
explanation for how this transformation could be performed by the brain, and suggest that areas 
MT and MST utilize the powerful computational means of gain modulation and tuning shifts to 
do so (Bremmer et al. 1997; Inaba et al. 2007; Chukoskie and Movshon 2009; Fujiwara et al. 
2011; Inaba et al. 2011). Importantly, the model simulations we present here provide several 
testable predictions for the neurophysiological properties which might be present in any area 
involved in the 3D transformation of retinal signals for smooth pursuit.
Acknowledgments

The authors would like to thank the reviewers for their insightful and helpful comments on this manuscript. This work was supported by NSERC (Canada), CFI (Canada), the Botterell Fund (Queen’s University, Kingston, ON, Canada) and ORF (Canada).

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
References


Berezovskii VK and Born RT. Specificity of Projections from Wide-Field and Local Motion-Processing Regions within the Middle Temporal Visual Area of the Owl Monkey. *J Neurosci* 20: 3: 1157-1169, 2000.


Buneo Ca, Jarvis MR, Batista AP and Andersen RA. Direct visuomotor transformations for


Klier EM, Blohm G and Crawford JD. Neural Mechanisms of Eye Movements: Three-


Lee B, Pesaran B and Andersen RA. Area MSTd neurons encode visual stimuli in eye


McGuire LMM and Sabes PN. Heterogeneous representations in the superior parietal lobule are common across reaches to visual and proprioceptive targets. *J Neurosci* 31: 18: 6661-73, 2011.


Murdison TS, Paré-Bingley CP and Blohm G. Evidence for a retinal velocity memory


Naka KI and Rushton WAH. S-Potentials from colour units in the retina of fish (cyprinidae). *J Physiol (Lond)* 185: 536-555, 1966b.

Naka KI and Rushton WAH. S-Potentials from luminosity units in the retina of fish (cyprinidae). *J Physiol (Lond)* 185: 587-599, 1966c.


Ono H and Barbeito R. The cyclopean eye vs. the sighting-dominant eye as the center of visual direction. *Percept Psychophys* 32: 3: 201-10, 1982.


Smith MA and Crawford JD. Self-organizing task modules and explicit coordinate systems in


**Stricanne B, Andersen RA and Mazzoni P.** Eye-Centered, Head-Centered, and Intermediate Coding of Remembered Sound Locations in Area LIP. *J Neurophysiol* 76: 2071-2076, 1996.


**Tweed D and Vilis T.** Implications of rotational kinematics for the oculomotor system in three


Figure captions

Figure 1: Overview of predicted effects. A Head roll-induced ocular counter-roll (OCR). During head roll, the eyes counter-rotate by a small gain (10% here for illustration purposes), causing misalignments between the spatial target direction (black) when projected onto the retina (red) and head-fixed axes (blue). These rotations must be accounted for by the brain, otherwise they lead to smooth eye movements trajectories rotated by OCR ($\theta_{OCR}$, red arrow). B Oblique gaze-induced retinal rotations. When the eyes pursue moving targets starting from orientations along an oblique gaze vector (along the 45° direction here), the retinal movement vector becomes increasingly rotated (grayscale arrows) as a result of the spherical projection geometry of images onto the back of the eyes. These retinal rotations predict pursuit vectors (red) that are rotated by this same angle ($\theta$) from each oblique orientation. C Half-angle rule. When the eyes pursue targets moving along the horizontal direction at increasingly eccentric vertical gaze orientations (black $\theta$), Listing’s law requires that the axis of rotation has a torsional component equal to one half of the eccentricity ($\theta/2$, grayscales axes). If the eyes did not obey the half-angle rule, one might expect the axis of rotation to contain torsional components orthogonal to the eccentricity of eye orientation (red $\theta$, light to dark red axes), as the retinal input remains constant regardless of vertical orientation (single black retinal vector). D Optokinetic nystagmus (OKN). In order to pursue a target following OKN, the brain must perform a vector addition of the current gaze speed ($s_{g,0}$, green) and the retinal target speed ($s_{t,r}$) in order to reconstruct the on-screen target speed ($\hat{s}_{t,s}$).

Figure 2: Geometrical model and neural network model. A Single trial of the simulated pursuit task comprising our geometrical model used to generate retinal and extraretinal signals for the training set. Using the depicted set-up, the model computes the smooth pursuit command required to minimize the velocity of a frontoparallel on-screen target, projected onto the retina, given various gaze positions, eye and head orientations and velocities. In this illustration (from the subject’s perspective), the x-axis (horizontal) points to the right of the subject, the y-axis (depth) points into screen and the z-axis (vertical) points towards the ceiling. B Network model architecture. We trained a rate-based four-layer (two hidden), fully feed-forward network model to produce 3D, spatially correct smooth pursuit velocity commands (output layer) from five input populations: (1) 2D retinal position and velocity signals, (2) 3D eye-in-head orientation and (3) velocity signals and (4) 3D head-on-shoulders orientation and (5) velocity signals. All weight matrices ($w_{in}, w_{HL}, w_{out}$) were adjusted during training and we trained eight network sizes (9-100 units in each hidden layer).

Figure 3: Network performance. A All network sizes adequately performed the 3D transformation, according to a regression analysis between the observed 3D compensation index and the predicted 3D compensation index (Blohm and Crawford 2007), based on 10,000 different test simulation points (previously unseen by the network), shown here for only the 100 HLU network for clarity. Also shown are the regression fits (dashed color-matched lines) for each network size. B 3D compensation error, represented by the component of the observed 3D compensation orthogonal to the predicted compensation, revealed that all network sizes (color-matched stair histograms) adequately minimized 3D error. C Distributions of half-angle rule velocity axis “torsional tilt” error (°) for each network size (color-matched stair histograms).
revealed that all networks adequately adhered to Listing’s law.

**Figure 4:** Gain modulation and visual tuning shifts. **A** (left) Gain-modulated unit from the first hidden layer of the 100 HLU network (#85) with eye velocity gain fields. Note that for all eye velocities the PD (white bar), COM (white circle), minimum (downward magenta triangle) and maximum (upward green triangle) remain in constant locations. (right) Unit from the second hidden layer of the 100 HLU network (#72) exhibiting eye velocity gain fields. **B** (left) Summary of gain modulation in the first hidden layers of each network size. In general, the first hidden layer units exhibited larger gain modulations than units in the second hidden layer (see right panel). Each box represents the minimum and maximum of gain modulation in the horizontal and vertical eye velocity directions, and the color-matched histograms along the borders reveal the distributions of gain modulations for each network size. Note that this analysis only accounted for fluctuations in overall activity (not tuning directions). (right) Summary of gain modulation in the second hidden layers of each network size. In general, the second hidden layer units exhibited smaller gain modulations than units in the second hidden layer (see left panel). Conventions are the same as in the left-hand panel.

**Figure 5:** Visual tuning properties. **A** Visual tuning modulations of typical units in the first (#1) and second hidden layers (#7) of the 100 HLU network during head roll-induced OCR. Across head roll angles of -40° to 40° (pseudo-color plot, abscissa), and across ocular torsion angles of -28° to 28° (pseudo-color plot, ordinate). The preferred tuning of the units (shown in polar coordinates projected along each axis according to red and blue color scales) shift by small amounts for the first hidden layer unit (0.5° with head roll and 0.02° with OCR) and larger amounts for the second hidden layer (8° with head roll and -8° with OCR). **B** Comparing these shifts to those required to compensate for head roll and OCR (in red and blue for first and second hidden layer units, respectively), we see that the shifts of the first hidden layer unit indicate a nearly retinal coding (regression gains = 0, dashed lines) and the shifts of the second hidden layer unit indicate an intermediate coding frame. This same analysis was performed to observe unit by unit compensation for head roll. Also shown are the distributions of compensation gains across head roll and ocular torsion for the 100 HLU network (grey histogram for first hidden layer units; black stair histogram for second hidden layer units), with quartiles (tick marks representing 25%, 50% and 75% gain percentiles) for each layer of each network size above histogram (color-matched, with lighter shades corresponding to first hidden layer quartiles and darker shades corresponding to second hidden layer quartiles). This histogram and these quartiles show a consistent narrow distribution of first hidden layer units around the retinal gain (dashed vertical), while second hidden layer units show a consistently wider distribution, indicating more units coding to a frame intermediate to retinal and spatial (solid vertical). **C** and **D** Summaries of compensatory gain distributions for hidden layer units in first and second hidden layers during oblique gaze (C) and half-angle rule (D) simulations. These distributions revealed consistent, retinal coding by first hidden layer units and intermediate coding by second hidden layer units. Color scheme and plotting conventions are identical to those in panel B.

**Figure 6:** Visual tuning properties: OKN separability analysis. **A** Retinal target speed versus
eye speed normalized activity gradients for two typical target-dependent (T) units in the first
hidden layer of the 100 HLU network (#s 74 and 23) and B two units in the second hidden layer
(#s 38 and 72) with both separable (T and E) and inseparable (T-E and T+E) characteristics. Because zero retinal velocity induced zero eye movement vectors, we started the gradient
analysis at speeds greater than 10°/s, and analyzed positive and negative target speeds separately.

C Direction-binned polar histogram showing the distribution of gradients in first and D second
hidden layer units. The colored radii indicate the mean gradient directions for each network, and
the dashed radius (for the first hidden layer of the 25 HLU network) represents a mean gradient
from which unit gradients were significantly different (t(99) = 2.09, p < 0.05).

Figure 7: Typical HLU 3D and 2D motor field. A Depiction of the 3D motor field for a typical
unit from the first hidden layer of the 100 HLU network (#16). In the pseudo-color plots, we
show three orthogonal views: (top left) from the back, revealing activities correlated to outputs in
the frontoparallel (x-z) plane; (top right) from the side, revealing activities correlated to outputs
in the sagittal (y-z) plane; (bottom left) from the bottom, revealing activities correlated to outputs
in the transverse (x-y) plane. Also shown in each plot is the locus of the activity-weighted
COM_{MF} (white circle). B 2D motor tuning curve for the same unit from the 100 HLU network
(#16). The radial distance represents the activation of the unit associated with each directional
bin and the error bars represent the SE within each bin. The red bar represents the preferred
motor tuning direction (PD_{MF}). Note the similarity between the 2D tuning curve and the area of
high activation in the 3D motor field representation in panel A.

Figure 8: Motor field properties. A Shifts of a representative first hidden layer unit’s motor
tuning (top left, #63) and second hidden layer unit’s motor tuning (top right, #88) from the 100
HLU network under conditions of head roll-induced OCR. The pseudo-color plots here represent
the shift of motor field tuning similar to those in Figure 5A, but the polar tuning plots were
omitted here for clarity. B Summary of compensatory gain distributions for hidden layer units in
first and second hidden layers across head roll and OCR. These distributions revealed retinal
(unit #63), spatial (unit #88), and mixed, intermediate coding by both first and second hidden
layer units. C and D Summary of compensatory gain distributions for hidden layer units in first
and second hidden layers during oblique gaze (C) and half-angle rule (D) simulations. These
distributions revealed mixed, intermediate coding by both first and second hidden layer units.
Color scheme and plotting conventions are identical to those in previous figures.

Figure 9: Microstimulation properties. A Example evoked gaze velocity vector endpoints,
projected onto the screen for units #15 and #93 from the first and second hidden layers of the 100
HLU network, respectively, under conditions of head roll-induced OCR. Also shown are the axes
representing head roll (blue) and OCR (red), which run along the outer edges of each “field” of
evoked pursuit movements. B, C and D Summaries of microstimulation compensatory gain
distributions for hidden layer units in first and second hidden layers during head roll-induced
OCR (B) oblique gaze (C) and half-angle rule (D) simulations. Each of these distributions
revealed mainly spatial gaze (C) and half-angle rule (D) simulations. These of evoked pursuit movements. B, C and D Summaries of microstimulation compensatory gain
distributions for hidden layer units in first and second hidden layers during head roll-induced
OCR (B) oblique gaze (C) and half-angle rule (D) simulations. Each of these distributions
revealed mainly spatial coding by the second hidden layer units, with some first and second
hidden layer units coding according to a mixed, intermediate frame as well. Color scheme and
plotting conventions are identical to those in previous figures.

**Figure 10: Summary of results across simulations and network sizes.** We present the quartile locations for the units of each hidden layer (first along horizontal axis, second along vertical axis) for each network size (see legend), relative to the retinal (dashed lines) and spatial (solid lines) predictions, for visual tuning shifts (orange quartiles), motor field properties (purple quartiles) and microstimulation properties (blue quartiles). Also shown along each axis (representing each hidden layer) are the distributions of all unit gains in each analysis (color coded), collapsed across all network sizes.
A head roll-induced OCR

B oblique gaze

C half-angle rule

D optokinetic nystagmus

\[ \theta_{\text{OCR}} = \theta_{\text{head}} + \theta_{\text{retinal}} \]

predictions spatial retina

\[ s_{\theta} = s_{\theta_0} + s_{\theta_f} \]

\[ s_{\theta_f} = s_{\theta_0} - s_{\theta_0} \]
100 HLU network

A

first hidden layer

90°

180°

0°

270°

PD shift 0.5°

OCR (°)

head roll (°)

PD shift 0.02°

second hidden layer

90°

180°

0°

270°

PD shift 8°

OCR (°)

head roll (°)

PD shift -8°

B

gain RE: head roll

HL1: r = -0.33, p < 0.001

HL2: r = 0.70, p < 0.001

C

gains re: oblique gaze

D

gains re: half-angle rule
A 100 HLU net
HL1 #16

activation

BACK

SIDE

BOTTOM

vertical velocity (g)

torsional velocity (°/s)

horizontal velocity (°/s)

B

act. = 1

90°

270°

180°
100 HLU network

A

first hidden layer  second hidden layer

head roll (°)

OCR (°)

#15

#93

B

HL1: $r = 0.54$, $p < 0.001$

HL2: $r = 0.97$, $p < 0.001$

C

D

number of units

gains re: oblique gaze

gains re: half-angle rule