The Simultaneous Coding of Heading and Path in Primate MSTd

Oliver W. Layton and N. Andrew Browning

Abstract—The spatio-temporal displacement of luminance patterns in a 2D image is called optic flow. Present biologically-inspired approaches to navigation that use optic flow largely focus on the problem of extracting the instantaneous direction of travel (heading) of a mobile agent. Computational models have demonstrated success in estimating heading in highly constrained environments whereby the agent is largely assumed to travel along straight paths. However, drivers competitively steer around curved road bends and humans have been shown capable of judging their future, possibly curved, path of travel in addition to instantaneous heading. The computation of the general future path of travel, which need not be straight, is of interest to mobile robotic, autonomous vehicle driving, and path planning applications, yet no biologically-inspired neural network model exists that provides mechanisms through which the future path may be estimated. We present a biologically inspired recurrent neural network, based on brain area MSTd, that can dynamically code both instantaneous heading and path simultaneously. We show that the model performs similarly to humans in judging heading and the curvature of the future path.

I. INTRODUCTION

OPTIC flow is defined as the spatial displacement of luminance over time on a 2D image. Present neural networks for visually-guided navigation and optimization methods can competently estimate the parameters in first-order optic flow [1] that are necessary for computing the instantaneous direction of travel (heading) from visual motion fields [2]. However, only estimating instantaneous heading when processing optic flow for the guidance of mobile robotic systems has its limitations. First, when a mobile agent travels along a circular path, the representation of path remains constant, whereas the instantaneous heading needs to be constantly recomputed. Thus, incorporating path estimation in a neural network for autonomous navigation has a computational advantage over heading. In the present paper, we use heading to refer to the instantaneous, tangential, straight course of travel, and path refers to the general future trajectory, which may be straight or curved. Second, path information is useful for future route planning and detecting potential interceptions with stationary and moving targets. Human drivers have been shown to safely negotiate steering around curvilinear road bends, even at high speeds [3], [4]. Therefore, anticipating and/or determining the future path of travel can prevent an autonomous agent from veering off course.

The ability of humans to navigate on varied terrain with ease has ignited much research interest. Research demonstrates that humans can judge their heading in rigid environments with little error (< 1°) [5], and when moving their eyes with error that depends on the eye rotation rate [6]. In the presence of independently moving objects, human heading bias depends on whether the object approaches or maintains a fixed distance with respect to the observer [7], [8]. Because humans estimate their heading with reasonable accuracy in a short amount of time, the computational mechanisms in the brain that afford this performance are of interest for neural network models that process visual motion for autonomous navigation. When an observer travels along a straight path in a rigid environment toward a frontoparallel plane without eye movements, the observer experiences a radially expanding optic flow field that contains a singularity known as the focus of expansion (FoE) that specifies the heading direction [9]. Neurophysiological experiments have shown that the dorsal medial superior temporal area (MSTd) in primates contains neurons that selectively respond to radially expanding optic flow fields similar to those viewed when traveling along a straight path [10], [11]. Neural network models have demonstrated that the activity of model MSTd neurons tuned to radial motion patterns yields the same patterns of heading errors as humans in rigid environments and in the presence of independently moving objects [7], [12], [13], indicating that MSTd may represent a crucial brain area for heading perception.

In the natural case of travel along a curvilinear path, instantaneous heading may be less important for guiding navigation than the perception of the future, potentially curved, path. Unlike the straight path case wherein the instantaneous heading specifies the future path, curvilinear navigation produces optic flow with translation, and rotation that is intrinsic to the pattern of motion on the human retina (i.e. not due to eye, heading, or body movements). In the present paper, we assume navigation is along a circular path and all rotation in the optic flow is intrinsic. Therefore, the optic flow contains information about the path curvature. When traveling along a circular path and the observer gaze is in the same direction as heading (i.e. looking where one is going), humans can accurately judge the path curvature [14], even in the presence of independently moving objects [15]. Li & Cheng showed that human judgments underestimate the curvature of the circular path when gaze is outside the circle and overestimate the curvature when gaze is on or inside the path [16].

No existing theory of path perception explains the underlying neural mechanisms or clarifies whether or not heading is involved. We present a neural network model of path perception motivated by primate brain area MSTd that encodes path through the maximally active subpopulation of units.
selective along a continuum of radially expansive, spiral, and center global motion patterns, and the two dimensional (2D) visuotopic location of the maximally active unit corresponds to the heading direction. Our neural network predicts that the representation of heading and path are intimately related and dynamically interact over time.

The rest of this paper is organized as follows: Section II presents a mechanism inspired by primate brain area MSTd that may simultaneously code heading and path. Section III introduces a neural network inspired by MSTd and shows derivations of the path conditions simulated in the network. Simulations results of conditions that resemble those in human psychophysical experiments of heading and path perception are shown in Section IV. Finally, Section V concludes the paper.

II. HEADING AND PATH IN MSTd

Due to the tendency of neurons in brain area MSTd to respond to large radial expansion or contraction, researchers have suspected that the primary functional role of MSTd in navigation is heading perception [10], [17]. Although most neurophysiological investigations of MSTd have focused on radial optic flow sensitivity, some studies demonstrate that many MSTd neurons exhibit highly selective tuning to spiral [18] and center [19], [20] optic flow patterns.

No functional role has been ascribed to spiral optic flow selectivity.

We propose that spiral optic flow selectivity affords MSTd with sensitivity to the future path of travel. When humans travel along a curvilinear path in a flat environment with a fixed direction of gaze, spiral optic flow is experienced that contains information about the path curvature. Neurons that are selective to different spirals that also vary in the curvature of their preferred spiral pattern would therefore be capable of extracting information about the future path. Figure 1 depicts our proposed spiral space functional organization of MSTd. Each rectangular cross-section in the volume spans the selectivity across 2D visuotopic space of neurons with same spiral tuning. For a fixed 2D location, the depth of the 3D volume corresponds to selectivity to different patterns with counter-clockwise (CCW) orientations in spiral space spanning radial fields (top), spirals (middle), and centers (bottom). There are also complementary selectivity volumes for clockwise (CW) patterns and contraction (not shown). The collection of blue arrows correspond to a vector field representation of the optic flow experienced by an observer moving along a circular path. The ‘heat map’ superimposed on each cross-section shows the spatial activity pattern of hypothetical neurons. We hypothesize that the spiral selectivity of the most active MSTd subpopulation across the volume depth codes path and its cross-sectional spatial location indicates the heading.

In the simple case of traveling along a straight path, we expect neurons on the radial side of the spiral selectivity continuum to be most active, indicating no path curvature.
and we anticipate the peak to be located in the 2D cross-section at the FoE, indicating the heading. Therefore, the population MSTd response in this example is the same as if there were only neurons selective to radial patterns. In the case of a circular path, we expect the spiral-selective neurons with spiral 'arms' that best match the path curvature to be most active.

The following neural network implements the mechanisms by which MSTd may exploit tuning in spiral space to simultaneously code path curvature and heading.

III. Model Description

In this section, we present a recurrent neural network model that simultaneously codes heading and path. The model consists of stages that correspond to primate primary visual cortex (V1) for local motion detection, medial temporal area (MT) for short range motion grouping, and MSTd for the sensitivity to large patterns of motion (Figure 2). In our neural network, units process optic flow and compete across spiral space, spiral orientation, and visuotopic location. Before we introduce the model stages, we summarize the psychophysical experiments of Li & Cheng (2011) that we simulate to compare our neural network performance with human judgments of path and heading.

A. Experimental Conditions

The displays were composed of 1000 dots distributed along a ground plane 1.41–50 m away in depth from the observer’s initial position. Subjects had a 110° horizontal field of view and fixated the center of the screen, slightly above the ground plane horizon, throughout the 1 sec simulated movement along the circular path. There were five experimental conditions that varied according to the simulated gaze direction of the observer: gaze along Z-axis (Figure 3a), gaze outside path (Figure 3b), gaze on target on path (Figure 3c), gaze along heading (Figure 3d), and gaze inside path (Figure 3e). In the Z-axis condition, no observer rotation occurred throughout the trial and the observer maintained gaze in the direction of the initial heading, which was defined as the z-axis of the world coordinate system. The subject observed radially expansive optic flow that contained a FoE that horizontally drifted over time. In the outside path condition, the subject was shown optic flow with simulated gaze on a target 15° outside the path with respect to the initial heading and \( r m \) away, where \( r \) represents the circular path radius. In the target on path condition, the simulated observer gaze was on a target located 30° with respect to the initial heading on the future path. The gaze along heading condition is the natural case wherein the observer always looks in the instantaneous direction in which he is going. Finally, the observer gaze in the inside path condition was on a target located 15° interior to the path from the initial heading and at a distance of \( r m \) away. Three different circle radii were tested: 28.5 m, 38.5 m, and 59 m. Subjects were asked to manipulate the horizontal position of a marker at the end of the trial such that it intersected the future path of travel. In separate experiments, Li & Cheng asked subjects to respond with their perceived heading at the end of the trial.

B. Model V1 & Optic Flow in Each Condition

Our neural network takes analytical optic flow as input corresponding to that viewed by a moving observer in each
Li & Cheng’s subjects of 1.51 m and ω = 3 r such that the observer will traverse 3 m around the circle when the trial concludes at t = 1, which was the case in the experiments of Li & Cheng. Hence, the initial position of the observer’s eye is given by

\[ p_h(0) = (r, \bar{y}, 0) \]  

and at time t the observer will have traversed an arc length of rωt about the circle (Figure 3a). In the outside path and inside path conditions, when gaze is on a simulated target, the target object has a position

\[ p_o = (x_o, \bar{y}, z_o). \]  

In the target on path condition, the target position can be parameterized with respect to the initial angle between the observer heading and the target θ(0) and the circle radius r (Figure 4b):

\[ p_o = (x_o, \bar{y}, z_o) = (r \cos(2\theta(0)), \bar{y}, r \sin(2\theta(0))) \]  

Let \( \bar{T} = (T_x, T_y, T_z) \) and \( \bar{R} = (R_z, R_y, R_z) \) represent the translational and rotational velocities of the observer, respectively. Note that \( \bar{R} \) incorporates all sources of rotation from the eye, head, and body. Consider a point P in the environment that has world coordinates \((X, Y, Z)\). We use a planar camera model with unit focal length to project P onto the retinal surface. The projection of P has a position \( p \) in retinal coordinates \((x, y)\) given by

\[ p = \left( \frac{x}{Z}, \frac{y}{Z} \right) = \frac{1}{Z} \begin{pmatrix} X \\ Y \end{pmatrix} \]  

Only points in the world within the simulated field of view are projected onto the retina. For a field of view angle of \( \phi = 110^\circ \), we require that the following two conditions hold:

\[ Z > 0 \]  

\[ Z | \tan(\Phi_0) | > X, \]  

for \( \Phi_0 = \pm \frac{\phi}{2} \). The instantaneous velocity of the projected point \((x, y)\) (i.e. first-order optic flow) that is the V1 representation in the model is given by [1]:

\[
\begin{pmatrix}
\frac{dx}{dt} \\
\frac{dy}{dt}
\end{pmatrix} = \frac{1}{Z} \begin{pmatrix}
-1 & 0 & x \\
0 & -1 & y
\end{pmatrix} \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix} + \begin{pmatrix}
x y \\
-(1 + x^2) & y \\
-xy & -x
\end{pmatrix} \begin{pmatrix} R_x \\ R_y \\ R_z \end{pmatrix}
\]  

Because the observer height does not change over time, \( T_y = 0 \) in Eq. 8. Across all simulated conditions, gaze is directed at the same eye height as the observer. Therefore, there are no source of rotation in the x or z directions, so \( R_x = R_z = 0 \) in all conditions.

The observer translation vector \( \bar{T} \) is given by:

\[ \bar{T} = \left( T_x, T_y, T_z \right) = \frac{\dot{p}_h(t)}{r} \]  

\[ = \left( -r \omega \sin(\omega t), 0, r \omega \cos(\omega t) \right) \]  

The rotational velocity about the y axis \( R_y \) differs across conditions due to variations in observer gaze. In the Z-axis
condition, gaze remained parallel to the Z axis and therefore \( R_y = 0 \). In the outside path condition, the observer gaze angle \( \theta(t) \) (Figure 4a) at time \( t \) with respect to the initial heading is

\[
\theta(t) = \arctan\left( \frac{\Delta z}{\Delta x} \right) = \arctan\left( \frac{x_o - x_h}{z_o - z_h} \right)
\]

(10)

\( R_y \) in the outside path condition is then defined as the rate of change of the rotation due to gaze:

\[
R_y = \frac{d}{dt} \theta(t) = \frac{\omega}{2}
\]

(11)

In the target on path condition, the parameterization of the target position with respect to the initial gaze angle (Eq. 4) affords the following definition for \( \theta(t) \):

\[
\theta(t) = \arctan\left( \frac{\Delta x}{\Delta z} \right) = \arctan\left( \frac{x_h - x_o}{z_h - z_o} \right)
\]

(12)

\( R_y \) in the target on path condition, is then given by:

\[
R_y = \frac{d}{dt} \theta(t) = \frac{\omega}{2}
\]

(13)

Because the observer rotates in the gaze along heading condition such that his gaze always remains tangent to the path, \( \theta(t) = \omega t \) and \( R_y = \omega \). The parameter values for \( \theta(t) \) and \( R_y \) in the inside path condition are found by Eqs. 10-11, and if the gaze angle differs in sign from that in the outside path condition, replace \( \Delta x \) with \(-\Delta x \).

Model MT smooths the optic flow vector field \((u_x, u_y)\) with a Gaussian filter, resulting in the vectors \((v_x, v_y)\) [13], [22].

C. Model MSTd

1) Template Match: Model units perform a template match on the pooled optic flow representation and dynamically interact in a competitive recurrent neural network. Templates vary across four dimensions: 1) spirality, 2) orientation, 3) horizontal offset, and 4) vertical offset. We generate spiral templates through a linear interpolation between radial and center templates [23]. The degree of interpolation is determined by the proportion \( \psi \), where \( 0 \leq \psi \leq 1 \). \( \psi = 0 \) corresponds to a center template, and \( \psi = 1 \) corresponds to a radial template. The spiral template space is a continuum that ranges from radial and center templates (Figure 1).

We created a neural network with 11500 MSTd neurons with motion pattern selectivities determined by the spiral templates. Each network unit compares its input to its preferred template, yielding a template match score that assesses the degree of similarity. The match score at time \( t \), \( M(s, a, x, y, t) \), for the network unit at location \((x, y)\) with preferred spirality \( a \) and orientation \( s \) is computed according to the following inner product:

\[
M(s, a, x, y, t) = \lambda \sum_{j,k} (x_j) e^{-}\frac{(x^2 + y^2)}{2} \times \left( \sum_{i} (v_{x,i}v_{y,i}) \frac{C(s,a,x,y) \circ (v_{x,i},v_{y,i})}{\sqrt{v_{x,i}^2 + v_{y,i}^2}} \right)
\]

(14)

Eq. 14 computes an inner product (i.e. cosine similarity) by performing component-wise multiplication, indicated by \( \circ \), between the input optic flow \((v_x,v_y)\) and the spiral template \( C \). The result is normalized by the \( L^2 \) norm of the optic flow vector and the vector components are summed. An exponential distance-dependent weighting is applied to give matches near the center of motion greater weight, following by the summing over all spatial locations to obtain a scalar match score. The parameter \( \lambda \) is set to \( \frac{1}{N} \), where \( N \) is the number of dots in the scene.

2) Network Dynamics: The ordinary differential equations that describe the temporal dynamics of neural network of MSTd densely sample the visual field and feature shunting competitive dynamics [24]. These equations perform a leaky integration of their inputs and simulate many known properties of neurons, such as divisive normalization [25], [26] and automatic gain control. Eq. 15 defines a dynamical competitive network that describes the activation of network unit \( S \) at spatial location \((x, y)\) that is selective to a spiral pattern with spirality \( a \) and orientation \( s \).

\[
\frac{dS_{s,a,x,y}}{dt} = \epsilon(-\alpha S_{s,a,x,y} + (\beta - S_{s,a,x,y})(S_{s,a,x,y}^2 + f(M(s, a, x, y, t))))
\]

(15)

Eq. 15 is a recurrent competitive field and is configured as a contrast-enhancing or winner-take-all network [24]. Competition between neurons in the network occurs across location and spiral template space. The constant \( \epsilon \) is defined as the inverse of the membrane time constant of the model neuron and scales how fast the neuron responds, \( \alpha \) signifies the passive decay rate, and \( \beta \) is the saturation upper bound of the model neuron. In Eq. 15, the inhibition model neurons receive from others in the network that have a different spiral pattern and orientation sensitivities is set to unity weight, and \( \mu_{s,y} \) differentially weights the spatial competition. The function \( f(w) \) in Eq. 15 is a sigmoidal transfer function defined as

\[
f(w) = \frac{(\left[w - \Gamma^{+}\right])^2}{\zeta + (\left[w - \Gamma^{+}\right])^2},
\]

(16)

where \([\cdot]^+\) indicates the half-wave rectification \( \max(\cdot, 0) \), \( \Gamma \) is a threshold on the input from model MT, and \( \zeta \) is a sigmoid shape parameter.
All simulations were run on a 8-core 2.66Ghz Mac Pro with 64 GB of memory using Mathematica 8. Parameter values listed in the text specify those that remained constant throughout all simulations. Table 1 summarizes parameters values that were used in configuring the MSTd dynamics.

Figure 6 summarizes model path estimation performance compared to humans in the experimental conditions of Li & Cheng. Figure 6a plots the maximum model activation (y-axis) yielded for each the subpopulations of network neurons tuned to different spiral templates (x-axis). Figure 6a shows the network activity at the end of the 1 sec trial, after which subjects would respond. The left (right) hand side of the x-axis corresponds to sensitivity to radial (center) patterns, and in between signifies spiral pattern selectivity. In the Z-axis condition and outside path condition, the MSTd neurons with radial pattern sensitivities ($\psi = 0$) elicited the highest activity. In the Z-axis condition, this result is expected since the optic flow at every frame of the input video is radially expansive without a rotational component. The similarity in the outside path condition and the Z-axis condition model activation explains why Li & Cheng report that humans responded as if they were moving along a straight path in the outside path condition.

Spiral subpopulations in the model yielded the maximal activation in the on path condition and along heading condition. The inside path condition produced a peak in the spiral continuum closer to model neurons tuned to center templates. The ordinal distribution of the peaks in Figure 6a are consistent with a response to the increasing curvature in the retinal optic flow patterns throughout the respective experimental conditions. Neurons sensitive to spiral patterns are capable of extracting this information in the optic flow.

Figure 6b compares the position of the peak subpopulation response in the network along the spiral continuum to human path error judgments. We calibrated the model path error readout such that the network response distribution in the along heading condition constituted zero path error. We make this assumption because human subjects in the experiments of Li & Cheng only on average yielded zero path error irrespective of the path curvature or structure of the environment in the along heading condition, the condition is the most ecologically-relevant case tested wherein gaze changes in tandem with heading direction, the network activation due to the optic flow in this condition (Figure 6a; blue) is therefore more likely to approximate that during actual curvilinear navigation. To produce path error readouts in the model, we subtracted the position of the peak in each experimental condition from the location in the along heading condition in the spiral space continuum. Negative and positive path errors are defined as an underestimation and overestimation of path curvature, respectively. The ordinal distribution of path errors yielded by the model (green) exactly match that of human subjects (blue). The model also produced comparable path error signs and magnitudes to human judgments: the model underestimated path curvature in all condition except along heading and inside path, exhibited no bias in the along heading condition, and overestimated the path curvature in the inside path condition. Our neural network predicts
that humans responded with large underestimations of path curvature in the Z-axis condition and outside path condition because the distribution of activity in the network is similar and yields the same peak in the subpopulation of radial pattern neurons. Curvature is overestimated in the model because the location of the network activity peak in the inside path condition falls closer to the center side of the spiral continuum than that in the along heading condition.

Vestibular input about velocity, acceleration, and jerk may play an important role to reduce path estimation error. Li & Cheng (2011) measured human path errors when subjects passively viewed optic flow. Were subjects to walk or drive, perhaps in a virtual reality environment, along a circular path and judge the path curvature, we anticipate that path error would be greatly reduced. Heading sensitive neurons in primate area MSTd are not only tuned to visual optic flow, but also vestibular inputs [11]. It may be the case that vestibular signals are combined with those derived from visual optic flow to mitigate heading and path errors.

Human subjects in the experiments of Li & Cheng also demonstrated small heading bias in the outside path condition, on path condition, and inside path condition. The human pattern of bias follows that of the path errors: the on path and on path conditions yielded slight bias away from the direction of path curvature and the inside path condition resulted in larger bias in the direction of path curvature. Figure 7a compares model heading errors with those produced by humans. Positive and negative heading errors correspond to heading judgments biased in the direct of and the direction opposite of the path curvature, respectively. Model readouts of heading are taken as the spatial displacement of the peak network subpopulation activity compared to the peak location in the along heading condition. Because the experimental conditions varied due to gaze about the y-axis, we considered peak displacements along the x-axis (parallel to the horizon). The magnitude and sign of model heading
errors agreed well with those of human subjects.

Our simulation results indicate that heading and path curvature can be encoded in a single neural network that performs a template match on optic flow input. This confers a number of advantages for autonomous navigation applications. First, the simultaneous coding of path and heading can reduce the computation time needed for guiding mobile robotic agents. Second, our approach estimates path and heading in parallel based on the input optic flow. Path planning need not rely on heading estimations, which may contain bias. Third, the simultaneous coding of path and heading results in a smaller network size and less memory overhead compared to computing heading and path in separate network layers.

Finally, we study the stability of the path and heading representation over time. Figure 7b plots the spatial position (x-axis) of the network activity peak (white; indicated by orange line) over time (y-axis) in the along heading condition. Over the course of the 1 sec trial, the spatial position of the maximally active network subpopulation does not change despite large local fluctuations in the optic flow. This invariance is consistent with the fact that the curvature of the circular path did not change during the trial. Thus, our network can produce a stable response that parsimoniously anticipates the future path of travel. Figure 7c shows the peak position of the maximally active subpopulation in the network over time in the Z-axis condition. In this case, the radial subpopulation is the most active, and the peak drifts from right to left over time, which is consistent with the change in heading over time. The radial subpopulation is the most active throughout the entire trial, indicating that the path is straight. Unlike in the along heading condition, the spatial position of the peak displaces over time.

V. CONCLUSIONS

We introduced a biologically-inspired neural network that codes path curvature by the maximally active neuron subpopulation along the spiral tuning continuum, and the visuotopic position of this model neuron indicates the heading direction. The pattern of activity in Figure 6a supports our hypothesis that a single network can extract path curvature and heading information in the retinal optic flow. The network performs similarly to human subjects when judging the curvature of the future path of travel (Figure 6b and Figure 7a). We showed that the heading estimates dynamically update over time. Finally, we showed that responses to path curvature along constant-curvature curvilinear paths are stable and peak activation in the network spatially displaces as instantaneous heading is expected to change.

REFERENCES