2015 Special Issue

A GPU-accelerated cortical neural network model for visually guided robot navigation

Michael Beyeler [a,*], Nicolas Oros [b], Nikil Dutt [a,b], Jeffrey L. Krichmar [a,b,1]

[a] Department of Computer Science, University of California Irvine, Irvine, CA, USA
[b] Department of Cognitive Sciences, University of California Irvine, Irvine, CA, USA

ARTICLE INFO

ABSTRACT

Humans and other terrestrial animals use vision to traverse novel cluttered environments with apparent ease. On one hand, although much is known about the behavioral dynamics of steering in humans, it remains unclear how relevant perceptual variables might be represented in the brain. On the other hand, although a wealth of data exists about the neural circuitry that is concerned with the perception of self-motion variables such as the current direction of travel, little research has been devoted to investigating how this neural circuitry may relate to active steering control. Here we present a cortical neural network model for visually guided navigation that has been embodied on a physical robot exploring a real-world environment. The model includes a rate based motion energy model for area V1, and a spiking neural network model for cortical area MT. The model generates a cortical representation of optic flow, determines the position of objects based on motion discontinuities, and combines these signals with the representation of a goal location to produce motor commands that successfully steer the robot around obstacles toward the goal. The model produces robot trajectories that closely match human behavioral data. This study demonstrates how neural signals in a model of cortical area MT might provide sufficient motion information to steer a physical robot on human-like paths around obstacles in a real-world environment, and exemplifies the importance of embodiment, as behavior is deeply coupled not only with the underlying model of brain function, but also with the anatomical constraints of the physical body it controls.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Animals are capable of traversing novel cluttered environments with apparent ease. In particular, terrestrial mammals use vision to routinely scan the environment, avoid obstacles, and approach goals. Such locomotor tasks require the simultaneous integration of perceptual variables within the visual scene in order to rapidly execute a series of finely tuned motor commands. Although there have been studies on the behavioral dynamics of steering in humans (Fajen & Warren, 2003; Wilkie & Wann, 2003), it remains unclear how the perceptual variables highlighted therein might be represented in the brain. Only recently did researchers begin to ask whether individual behavioral performance is reflected in specific cortical regions (Billington, Wilkie, & Wann, 2013; Field, Wilkie, & Wann, 2007). Despite a wealth of data regarding the neural circuitry concerned with the perception of self-motion variables such as the current direction of travel ("heading") (Britten & van Wezel, 1998; Duffy & Wurtz, 1997; Gu, Watkins, Angelaki, & DeAngelis, 2006) or the perceived position and speed of objects (Eifuku & Wurtz, 1998; Tanaka, Sugita, Moriya, & Saito, 1993), little research has been devoted to investigating how this neural circuitry may relate to active steering control.

Visually guided navigation has traditionally attracted much attention from the domains of both vision and control, producing countless computational models with excellent navigation and localization performance (for a recent survey see Bonin-Font, Ortiz, & Oliver, 2008). However, the majority of these models have taken a computer science or engineering approach, without regard to biological or psychophysical fidelity. To date, only a handful of neural network models have been able to explain the trajectories taken by humans to steer around stationary objects toward a goal (Browning, Grossberg, & Mingolla, 2009; Elder, Grossberg,

Fig. 1. "Le Carl" Android based robot, which was constructed from the chassis of an R/C car. The task of the robot was to navigate to a visually salient target (bright yellow foam ball) while avoiding an obstacle (blue recycle bin) along the way. The robot's position throughout the task was recorded by an overhead camera that tracked the position of the green marker. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

& Mingolla, 2009), and none of them have been tested in real-world environments. The real world is the ultimate test bench for a model that is trying to link perception to action, because even carefully devised simulated experiments typically fail to transfer to real-world settings. Real environments are rich, multimodal, and noisy; an artificial design of such an environment would be computationally intensive and difficult to simulate (Krichmar & Edelman, 2006). Yet real-world integration is often prohibited due to engineering requirements, programming intricacies, and the sheer computational cost that come with large-scale biological models. Instead, such models often find application only in constrained or virtual environments, which may severely limit their explanatory power when it comes to generalizing findings to real-world conditions.

In this paper, we present a cortical neural network model for visually guided navigation that has been embodied on a physical robot exploring a real-world environment (see Fig. 1). All experiments were performed by "Le Carl", an Android Based Robotics (ABR) platform (Oros & Krichmar, 2013b) constructed from the chassis of an R/C car. An Android phone, mounted on the robot, was used to stream image frames while the robot was traversing a hallway of an office building. The images were sent via WiFi to a remote machine hosting the cortical neural network model. The neural network model generated a cortical representation of dense optic flow and determined the position of objects based on motion discontinuities. These signals then interacted with the representation of a goal location to produce motor commands that successfully steered the robot around obstacles toward the goal. The neural network model produced behavioral trajectories that not only closely matched human behavioral data (Fajen & Warren, 2003), but were also surprisingly robust.

To the best of our knowledge, this is the first cortical neural network model demonstrating human-like smooth paths in real-time in a real-world environment. The present model is inspired by the more comprehensive STARS model (Elder et al., 2009) and its successor, the ViSTARS model (Browning et al., 2009). All three models share some similarities, such as the use of cortical motion processing mechanisms in area MT to induce repeller-like obstacle avoidance behavior. Notable differences are that the present model uses a widely accepted model of cortical motion perception (Simoncelli & Heeger, 1998) to generate neural motion signals, that it combines these signals with a much simpler steering

control scheme, and that it does not attempt to model self-motion processing in higher-order parietal areas.

It should also be noted that area MT of the present model, which was responsible for generating the motion output, was composed of thousands of simulated biologically realistic excitatory and inhibitory spiking neurons (Izhikevich, 2003) with synaptic conductances for AMPA, NMDA, GABAa, and GABAb. Building neurorobotics agents whose behavior depends on a spiking neuron model that respects the detailed temporal dynamics of neuronal and synaptic integration of biological neurons may facilitate the investigation of how biologically plausible computation can give rise to adaptive behavior in space and time. In addition, spiking neurons are compatible with recent neuromorphic architectures (Boahen, 2006; Cassidy et al., 2014; Khan et al., 2008; Schemmel et al., 2010; Srinivasa & Cruz-Albrecht, 2012) and neuromorphic sensors (Lichtsteiner, Posch, & Delbruck, 2008; Liu, van Schaik, Minch, & Delbruck, 2010; Wen & Boahen, 2009). Thus, developing complex spiking networks that display cognitive functions or learn behavioral abilities through autonomous interaction may also represent an important step toward realizing functional large-scale networks on neuromorphic hardware.

Overall, this study demonstrates how neural signals in a cortical model of visual motion perception can respond adaptively to natural scenes, and how the generated motion signals are sufficient to steer a physical robot on human-like paths around obstacles. In addition, this study exemplifies the importance of embodiment for the validation of brain-inspired models, as behavior is deeply coupled not only with the underlying model of brain function, but also with the anatomical constraints of the physical body it controls.

2. Methods

2.1. The robotic platform

In order to engineer a system capable of real-time execution and real-world integration of large-scale biological models, we needed to address several technical challenges as outlined below.

First, the Android Based Robotics (ABR) framework was used as a flexible and inexpensive open-source robotics platform. In specific, we made use of the "Le Carl" robot, an ABR platform (Oros & Krichmar, 2012, 2013a, 2013b) constructed from the chassis of an R/C car (see Fig. 2). The main controller of the platform was an Android cell phone (Samsung Galaxy S3), which mediated communication via Bluetooth to an IOIO electronic board (SparkFun Electronics, http://www.sparkfun.com), which in turn sent PWM commands to the actuators and speed controllers of the R/C car. Since today's smartphones are equipped with a number of sensors and processors, it is possible to execute computationally demanding software based on real-time sensory data directly on the phone. The first neurorobotics study to utilize the ABR platform implemented a neural network based on neuromodulated attentional pathways to perform a reversal learning task by fusing a number of sensory inputs such as GPS location, a compass reading, and the values of on-board IR sensors (Oros & Krichmar, 2012). However, the complexity and sheer size of the present model did not allow for on-board processing, and instead required hosting the computationally expensive components on a remote machine.

Second, an Android app (labeled "ABR client" in Fig. 2) was developed to allow the collection of sensory data and communication with a remote machine (labeled "ABR server" in Fig. 2) via WiFi and 3G. The only sensor used in the present study was the Android phone's camera, which collected 320×240 pixel images while the robot was behaving. These images were then sent via UDP to the ABR server, at a rate of roughly 20 frames per second. The neural network processed the stream of incoming images in real-time
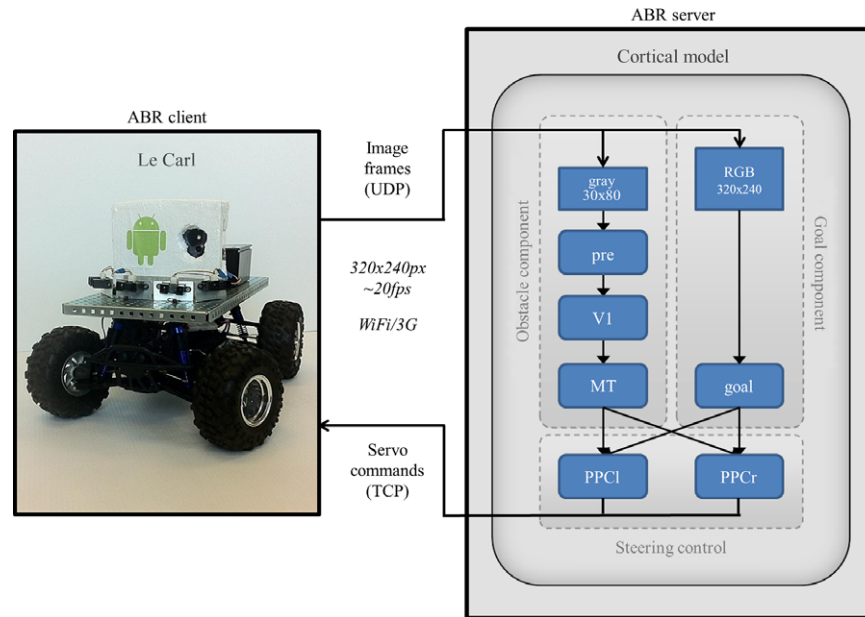
Fig. 2. Technical setup. An Android app (ABR client) was used to record 320 × 240 images at 20 fps and send them to a remote machine (ABR server) hosting a cortical model made of two processing streams: an obstacle component responsible for inferring the relative position and size of nearby obstacles by means of motion discontinuities, and a goal component responsible for inferring the relative position and size of a goal object by means of color blob detection. These streams were then fused in a model of the posterior parietal cortex to generate steering commands that were sent back to the ABR platform.

and generated motor commands, which were received by the ABR client app running on the phone via TCP. The ABR client communicated motor commands directly to the corresponding actuators and speed controllers of the R/C car via the IOIO board.

Third, we developed a Qt software interface that allowed integration of the ABR server with different C/C++ libraries such as CUDA, OpenCV, and the spiking neural network (SNN) simulator CARLsim (Beyeler, Carlson, Shuo-Chou, Dutt, & Krichmar, 2015; Nageswaran, Dutt, Krichmar, Nicolau, & Veidenbaum, 2009; Richert, Nageswaran, Dutt, & Krichmar, 2011). In order to adhere to the real-time constraints of the system, all computationally intensive parts of the model were accelerated on a GPU (i.e., a single NVIDIA GTX 780 with 3 GB of memory) using the CUDA programming framework.

2.2. The cortical neural network model

The high-level architecture of the cortical neural network model is shown in Fig. 2. The model was based on a cortical model of V1 and MT that was previously shown to produce direction and speed tuning curves comparable to electrophysiological results in response to synthetic visual stimuli such as sinusoidal gratings and plaids (Beyeler, Richert, Dutt, & Krichmar, 2014). The model used an efficient GPU implementation of the motion energy model (Simoncelli & Heeger, 1998) to generate cortical representations of motion in a model of the primary visual cortex (V1). Spiking neurons in a model of the middle temporal (MT) area then located nearby obstacles by means of motion discontinuities (labeled "Obstacle component" in Fig. 2). The MT motion signals projected to a simulated posterior parietal cortex (PPC), where they interacted with the representation of a goal location (labeled "Goal component" in Fig. 2) to produce motor commands to steer the robot around obstacles toward a goal. The following subsections will explain the model in detail.

2.2.1. Visual input

Inputs to the model were provided by the built-in camera of an Android phone (Samsung Galaxy S3) mounted on the

ABR platform. The phone took 320 × 240 pixel snapshots of the environment at a rate of roughly 20 frames per second. These frames entered the model in two ways: First, mimicking properties of visual processing in the magnocellular pathway, a grayscale, down-scaled version of the lower half of the frame (80 × 30 pixels) was sent to the network processing visual motion (obstacle component). This pathway consisted of a preprocessing stage as well as a model of V1, MT, and the PPC. Second, mimicking properties of visual processing in the parvocellular pathway, the originally collected RGB frame (320 × 240 pixels) was sent to a model component concerned with locating a visually salient target in the scene (goal component). Because a neural implementation of this pathway was considered out of scope for the present study, the goal object was located using OpenCV-based segmentation in HSV color space. The location of the goal in the visual field was then communicated to the PPC.

2.2.2. Preprocessing

The first stage of the obstacle component pathway (labeled "pre" in Fig. 2) enhanced contrast and normalized the input image using OpenCV's standard implementation of contrast-limited adaptive histogram equalization (CLAHE). The processed frame was then sent to the V1 stage.

2.2.3. Primary visual cortex (V1)

The second stage of the obstacle component pathway (labeled "V1" in Fig. 2) used the motion energy model to implement model neurons that responded to a preferred direction and speed of motion (Simoncelli & Heeger, 1998). In short, the motion energy model used a bank of linear space–time oriented filters to model the receptive field of directionally selective simple cells in V1. The filter responses were half-rectified, squared, and normalized within a large spatial Gaussian envelope. The output of this stage was equivalent to rate-based activity of V1 complex cells, whose responses were computed as local weighted averages of simple cell responses in a Gaussian neighborhood of $\sigma_{V1c} = 1.6$ pixels (see Fig. 3). The resulting size of the receptive fields was roughly one degree of the visual field (considering that the horizontal field of
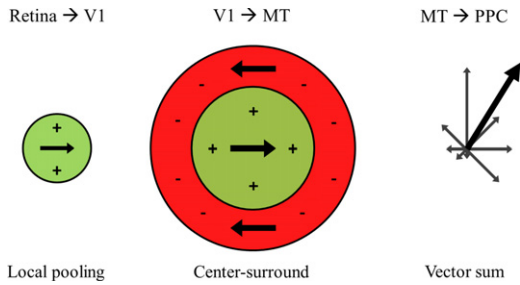
Retina → V1   V1 → MT   MT → PPC

Local pooling   Center-surround   Vector sum

Fig. 3. Schematic of the spatial receptive fields in the network. V1 neurons pooled retinal afferents and computed directional responses according to the motion energy model (Simoncelli & Heeger, 1998). The receptive fields of MT neurons had a circular center preferring motion in a particular direction, surrounded by a region preferring motion in the anti-preferred direction, implemented as a difference of Gaussians. PPC neurons computed a net vector response weighted by the firing rates of MT neurons.

view of a Samsung Galaxy S3 is roughly 60°), which is in agreement with electrophysiological evidence from recordings in macaque V1 (Freeman & Simoncelli, 2011). V1 complex cells responded to eight different directions of motion (in 45° increments) at a speed of 1.5 pixels per frame. We interpreted these activity values as neuronal mean firing rates, which were scaled to match the contrast sensitivity function of V1 complex cells (see Beyeler et al., 2014). Based on these mean firing rates we then generated Poisson spike trains (of 50 ms duration), which served as the spiking input to Izhikevich neurons representing cells in area MT. More information about the exact implementation can be found in Beyeler et al. (2014).

2.2.4. Middle temporal (MT) area

The simulated area MT (labeled "MT" in Fig. 2) consisted of 40,000 Izhikevich spiking neurons and roughly 1,700,000 conductance-based synapses, which aimed to extract the position and perceived size of any nearby obstacles by means of detecting motion discontinuities. Neurons in MT received optic flow-like input from V1 cells, thus inheriting their speed and direction preferences (Born & Bradley, 2005). Their spatial receptive fields had a circular center preferring motion in a particular direction, surrounded by a region preferring motion in the anti-preferred direction (Allman, Miezin, & McGuinness, 1985; Born, 2000). These receptive fields were implemented as a difference of Gaussians: Excitatory neurons in MT received input from Poisson spike generators in V1 in a narrow spatial neighborhood ($\sigma_{MTe} = 6.0$ pixels) and from inhibitory neurons in MT in a significantly larger spatial neighborhood ($\sigma_{MTi} = 12.0$ pixels; see Fig. 3), which are comparable in size to receptive fields of neurons in layers 4 and 6 of macaque MT (Raiguel, Van Hulle, Xiao, Marcar, & Orban, 1995). The weights and connection probabilities scaled with distance according to the Gaussian distribution. The maximum excitatory weight (at the center of a Gaussian kernel) was 0.01 and the maximum inhibitory weight was 0.0015 (see (4)). As a result of their intricate spatial receptive fields, excitatory neurons in MT were maximally activated by motion discontinuities in the optic flow field, which is thought to be of use for detecting object motion (Allman et al., 1985) and (Bradley & Andersen, 1998). Note that cortical cells with such a receptive field organization usually exhibit different binocular disparity preferences in their center and surround regions (Bradley & Andersen, 1998). However, because the model had access to only a single camera, we were unable to exploit disparity information (for more information please refer to Section "Model limitations").

All neurons in MT were modeled as Izhikevich spiking neurons (Izhikevich, 2003). The Izhikevich model aims to reduce

Hodgkin–Huxley-type neuronal models to a two-dimensional system of ordinary differential equations,

$$\frac{dv(t)}{dt} = 0.04v^2(t) + 5v(t) + 140 - u(t) + i_{\text{syn}}(t) \tag{1}$$

$$\frac{du(t)}{dt} = a(bv(t) - u(t)) \tag{2}$$

$$v(v > 30) = c \quad \text{and} \quad u(v > 30) = u - d. \tag{3}$$

Here, (1) described the membrane potential v for a given synaptic current i_{syn}, whereas (2) described a recovery variable u; the parameter a was the rate constant of the recovery variable, and b described the sensitivity of the recovery variable to the subthreshold fluctuations of the membrane potential. Whenever v reached peak ($v_{\text{cutoff}} = +30$), the membrane potential was reset to c and the value of the recovery variable was decreased by d (see (3)). The inclusion of u in the model allowed for the simulation of typical spike patterns observed in biological neurons. The four parameters a, b, c, and d can be set to simulate different types of neurons. All excitatory neurons were modeled as regular spiking (RS) neurons (class 1 excitable, $a = 0.02$, $b = 0.2$, $c = -65$, $d = 8$), and all inhibitory neurons were modeled as fast spiking (FS) neurons (class 2 excitable, $a = 0.1$, $b = 0.2$, $c = -65$, $d = 2$) (Izhikevich, 2003).

Ionic currents were modeled as dynamic synaptic channels with zero rise time and exponential decay:

$$\frac{dg_r(t)}{dt} = -\frac{1}{\tau_r}g_r(t) + \eta_r w \sum_i \delta(t - t_i), \tag{4}$$

where δ was the Dirac delta, the sum was over all presynaptic spikes arriving at time t_i, w was the weight of the synapse (hand-selected to be 0.01 at the center of an excitatory Gaussian kernel, and 0.0015 at the center of an inhibitory Gaussian kernel), τ_r was its decay time constant, η_r was a receptor-specific efficacy (or synaptic gain), and the subscript r denoted the receptor type; that is, AMPA (fast decay, $\tau_{\text{AMPA}} = 5$ ms), NMDA (slow decay and voltage-dependent, $\tau_{\text{NMDA}} = 150$ ms), GABAa (fast decay, $\tau_{\text{GABA}_a} = 6$ ms), and GABAb (slow decay, $\tau_{\text{GABA}_a} = 150$ ms). A spike arriving at a synapse that was post-synaptically connected to an excitatory (inhibitory) neuron increased both g_{AMPA} and g_{NMDA} (g_{GABA_a} and g_{GABA_b}) with receptor-specific efficacy $\eta_{\text{AMPA}} = 1.5$ and $\eta_{\text{NMDA}} = 0.5$ ($\eta_{\text{GABA}_a} = 1.0$ and $\eta_{\text{GABA}_b} = 1.0$). Having $\eta_{\text{AMPA}} > \eta_{\text{NMDA}}$ agrees with experimental findings (Myme, Sugino, Turrigiano, & Nelson, 2003) and allowed the network to quickly react to changing sensory input. The total synaptic current i_{syn} in (1) was then given by:

$$i_{\text{syn}} = -g_{\text{AMPA}}(v - 0) - g_{\text{NMDA}}\frac{\left(\frac{v+80}{60}\right)^2}{1 + \left(\frac{v+80}{60}\right)^2}(v - 0)$$
$$- g_{\text{GABA}_a}(v + 70) - g_{\text{GABA}_b}(v + 90). \tag{5}$$

For more information on the exact implementation of the Izhikevich model, please refer to the CARLsim 2.0 release paper (Richert et al., 2011).

2.2.5. Posterior parietal cortex (PPC)

The simulated PPC (labeled "PPCl" and "PPCr" in Fig. 2) combined visual representations of goal and obstacle information to produce a steering signal. The resulting dynamics of the steering signal resembled the Balance Strategy, which is a simple control law that aims to steer away from large sources of optic flow in the visual scene. For example, honeybees use this control law to steer collision-free paths through even narrow gaps by balancing the apparent speeds of motion of the images in their eyes (Srinivasan & Zhang, 1997). Interestingly, there is also some evidence for the

Balance Strategy in humans (Kountouriotis et al., 2013). However, the present model differs in an important way from the traditional Balance Strategy, in that it tries to balance the flow generated from motion discontinuities in the visual field, which are thought to correspond to obstacles in the scene, instead of balancing a conventional optic flow field. For more information see Section "Neurophysiological evidence and model alternatives ".

The steering signal took the form of a turning rate, $\dot{\theta}$:

$$\dot{\theta} = \hat{\theta} - \theta, \tag{6}$$

which was derived from the difference between the robot's current angular orientation, θ, and an optimal angular orientation estimated by the cortical network, $\hat{\theta}$. The variable θ was based on a copy of the current PWM signal sent to the servos of the robot (efference copy), whereas the variable $\hat{\theta}$ was derived from the neural activation in PPC, as described in (7). The resulting turning rate, $\dot{\theta}$, was then directly mapped into a PWM signal sent to the servos that controlled the steering of the robot. In order not to damage the servos, we made sure that the computed instantaneous change in turning rate, $\ddot{\theta}$, never exceeded a threshold (set at roughly 10% of the full range of PWM values). Steering with the second derivative also contributed to the paths of the robot being smoother.

The cortical estimate of angular orientation was realized by separately summing optic flow information from the left (F_L) and right halves (F_R) of the visual scene according to the Balance Strategy, and combining these flow magnitudes with information about the target location (T_L and T_R) weighed with a scaling factor α (set at 0.6):

$$\hat{\theta} \sim \frac{F_L - F_R + \alpha(T_R - T_L)}{F_L + F_R + \alpha(T_L + T_R)}. \tag{7}$$

Here, the total optic flow in the left (F_L) and right (F_R) halves was computed as follows:

$$F_L = \sum_{\theta} \sum_{y=0}^{R-1} \sum_{x=0}^{C/2-1} \| r_\theta(x, y) e_\theta \|, \tag{8}$$

$$F_R = \sum_{\theta} \sum_{y=0}^{R-1} \sum_{x=C/2}^{C-1} \| r_\theta(x, y) e_\theta \|, \tag{9}$$

where R is the number of rows in the image, C is the number of columns in the image, and $\| \bullet \|$ denotes the Euclidean 2-norm. Depending on the experiment, $r_\theta(x, y)$ is the firing rate of either an MT or a V1 neuron that is selective to direction of motion θ at spatial location (x, y), and e_θ is a unit vector pointing in the direction of θ.

The target location was represented as a two-dimensional blob of activity centered on the target's center of mass (x_G, y_G). If the target was located in the left half of the image, it contributed to a term T_L:

$$T_L = \alpha A_G \sum_{y=0}^{R-1} \sum_{x=0}^{C/2-1} e^{-\frac{(x-x_G)^2 + (y-y_G)^2}{2\sigma_G^2}}, \tag{10}$$

and if it was located in the right half of the image, it contributed to a term T_R:

$$T_R = \alpha A_G \sum_{y=0}^{R-1} \sum_{x=C/2}^{C-1} e^{-\frac{(x-x_G)^2 + (y-y_G)^2}{2\sigma_G^2}}, \tag{11}$$

where $\alpha = 0.6$, $\sigma_G = 0.2C$, and A_G was the perceived area of the target, which was determined with OpenCV using color blob detection. This allowed the target contribution to increase
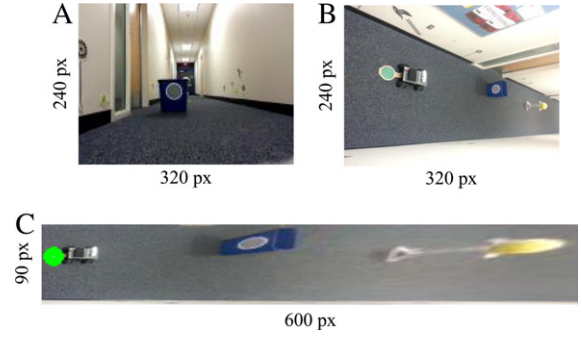

Fig. 4. Camera setup. (A) The robot's initial view of the scene from the onboard Android phone. (B) View of an overhead camera that tracked the green marker attached to the robot. (C) Birds-eye view image of the scene, obtained via perspective transformation from the image shown in (B). The robot's location in each frame was inferred from the location of the marker, which in turn was determined using color blob detection. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with target size, which we assumed to scale inversely with target distance (also compare Section "Model limitations "). Note that if the target was perfectly centered in the visual field, it contributed equally to T_L and T_R. Also, note that the contribution of the target component from the right half of the image, T_R, to θ was positive, whereas the contribution of the obstacle component F_R was negative. This led to the target component exhibiting an attractor-like quality in the steering dynamics, whereas the goal component exhibited a repeller-like quality.

3. Results

3.1. Experimental setup

In order to investigate whether the motion estimates in the simulated MT were sufficient for human-like steering performance, the model was tested on a reactive navigation task in the hallway of the University of California, Irvine Social and Behavioral Sciences Gateway, where our laboratory is located (see Fig. 1). This setup provided a relatively narrow yet highly structured environment with multiple sources of artificial lighting, thus challenging the cortical model to deal with real-world obstacles as well as to quickly react to changing illumination conditions.

Analogous to the behavioral paradigm of Fajen and Warren (2003), the robot's goal was to steer around an obstacle placed in its way (i.e., a recycle bin) in order to reach a distant goal (i.e., a yellow foam ball). The obstacle was placed between the location of the target and the robot's initial position, at three different angles (off-set by one, four, and eight degrees to the left of the robot's initial view) and different distances (2.5, 3, and 3.5 m from the robot's initial position). For each of these configurations we ran a total of five trials, during which we collected both behavioral and simulated neural data. An example setup is shown in Fig. 4(A). At the beginning of each trial, the robot was placed in the same initial position, directly facing the target. The robot would then quickly accelerate to a maximum speed of roughly 1 m/s. Speed was then held constant until the robot had navigated to the vicinity of the goal (roughly 0.5 m apart), at which point the trial was ended through manual intervention.

An over-head camera was used to monitor the robot's path throughout each trial (see Fig. 4(B)). The camera was mounted on the wall such that it overlooked the hallway, and was used to track the position of a large green marker attached to the rear of the robot (best seen in Fig. 1). Using a perspective transformation, a 320×240 pixel image (taken every 100 ms) was converted into a birds-eye view of the scene (see Fig. 4(C)). We used the
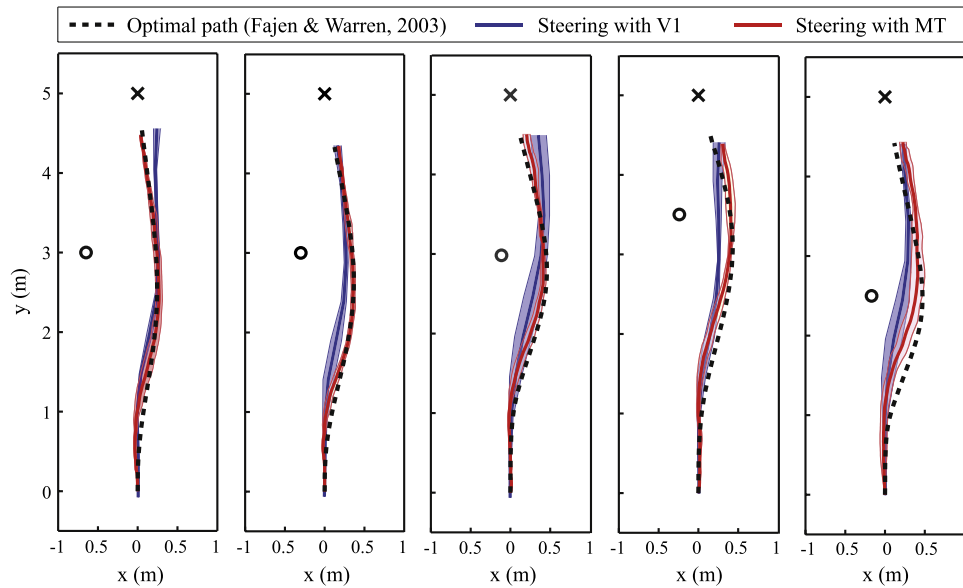
Fig. 5. Behavior paths of the robot (colored solid lines) around a single obstacle (recycle bin, 'O') toward a visually salient goal (yellow foam ball, 'X') for five different scene geometries, compared to "ground truth" (black dashed lines) obtained from the behavioral model by Fajen and Warren (2003). Results are shown for steering with V1 (blue) as well as for steering with MT (red). Solid lines are the robot's mean path averaged over five trials, and the shaded regions correspond to the standard deviation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

standard OpenCV implementation of this transformation and fine-tuned the parameters for the specific internal parameters of the camera. Although three-dimensional objects will appear distorted (as is evident in Fig. 4(C)), flat objects lying on the ground plane (i.e., the hallway floor) will be drawn to scale. This allowed us to detect and track the green marker on the robot directly in the transformed image (Fig. 4(C)), in which the size of the marker does not depend on viewing distance (as opposed to Fig. 4(B)). Color blob detection was applied to find the marker in each frame, and the marker's location was used to infer the robot's location. In order to aid tracking of the marker, we added temporal constraints between pairs of subsequent frames for outlier rejection. This procedure not only allowed us to automatically record the robot's location at any given time, but also allowed direct comparison of the robot's behavioral results with human psychophysics data.

3.2. Behavioral results

Fajen and Warren (2003) studied how humans walk around obstacles toward a stationary goal, and developed a model to explain the steering trajectories of the study's participants. Their work revealed that accurate steering trajectories can be obtained directly from the scene geometry, namely the distance and angles to the goal and obstacles. Using their behavioral model, we calculated "ground truth" paths for the scene geometry of our experimental setup, and compared them to steering trajectories generated by the robot's cortical neural network model. In order to assess the contribution of motion processing in MT we conducted two sets of experiments: one where steering commands were based solely on V1 activity, and one where steering commands were based on MT activity, by adjusting $r_\theta(xy)$ in (8) and (9).

Fig. 5 illustrates the robot's path around a single obstacle (recycle bin, 'O') toward a visually salient goal (yellow foam ball, 'X') for five different scene geometries, analogous to Fig. 10 in Fajen and Warren (2003). Results are shown for steering with V1 (blue) as well as for steering with MT (red). In the first three setups, the obstacle was placed 3 m away from the robot's initial location, off-set by eight (red), four (green), and one (blue) degrees to the left of the robot's initial view. Two other setups were tested, in which the obstacle distance was 3.5 m (yellow) and 2.5 m

(magenta) at an angle of 4°. The robot's mean path from five experimental runs is shown as a thick solid line in each panel, and the shaded region corresponds to the standard deviation. The black dashed lines correspond to the "ground truth" paths calculated from the behavioral model by Fajen and Warren (2003) using the corresponding scene geometry and standard parameter values. Each of these five obstacle courses was run five times for a total of 25 successful trials.

Steering paths were generally more accurate and more robust when steering with MT as opposed to steering with V1. Paths generated from neural activity in MT not only closely matched human behavioral data, but they also were surprisingly robust. That is, the robot successfully completed all 25 trials without hitting an obstacle, the walls of the hallway, or missing the goal. In fact, the only setup that proved difficult for the robot had the obstacle placed only 2.5 m away from the robot (rightmost panel in Fig. 5). In this scenario, "ground truth" data indicates that humans start veering to the right within the first second of a trial, which seemed not enough time for the robot. In contrast, steering from neural activity in V1 led to a total of 7 crashes, which involved the robot driving either into the obstacle or the wall, upon which the trial was aborted and all collected data discarded. We repeated the experiment until the robot successfully completed five trials per condition. When steering from V1 neural activity, the robot tended to react more strongly to behaviorally irrelevant stimuli, leading to more variability in the robot's trajectories.

For all five tested conditions we calculated the path area error (using the trapezoidal rule for approximating the region under the graph) as well as the maximum distance the robot's path deviated from the ground truth at any given time. Mean and standard deviation of these data (averaged over five trials each) are summarized in Table 1. All path errors were on the order of 10^{-1} m^2, with errors generated from steering with MT consistently smaller than errors generated from steering with V1. When the robot was steering from signals in MT, trial-by-trial variations were on the order of 10^{-2} m^2, suggesting that under these circumstances the robot reliably exhibited human-like steering behavior in all tested conditions. The maximum distance that the robot's path deviated from the human-like path was on the order of 10 cm, which is much smaller than the obstacle width (37 cm). In contrast,

Table 1
Summary of mean path errors when steering with signals in V1 vs. MT (5 trials each).

Obstacle distance (m)	Obstacle angle (deg)	Area error (m^2)		Maximum deviation (m)	
		V1	MT	V1	MT
3	−1°	0.472 ± 0.174	0.193 ± 0.080	0.237 ± 0.077	0.112 ± 0.045
3	−4°	0.407 ± 0.097	0.140 ± 0.024	0.193 ± 0.052	0.091 ± 0.013
3	−8°	0.354 ± 0.050	0.170 ± 0.055	0.165 ± 0.020	0.102 ± 0.026
3.5	−4°	0.332 ± 0.110	0.264 ± 0.100	0.169 ± 0.039	0.143 ± 0.029
2.5	−4°	0.575 ± 0.176	0.319 ± 0.093	0.288 ± 0.099	0.193 ± 0.056

steering with V1 often led to much larger deviations in some cases. The best single-trial result was 0.084 m^2 area error and 5 cm maximum deviation on the −1° condition. These results suggest that in most trials the task could still be performed even when steering with V1, but that motion processing in MT was critical for reliability and accuracy, both of which are distinct qualities of human-like steering.

Furthermore, these results are comparable to simulated data obtained with the STARS model, where the authors reported path area errors and maximum deviations on the same order of magnitude (see Table 2 in Elder et al., 2009). Their best result was 0.097 m^2 area error and 2.52 cm maximum deviation on a single trial of the −2° condition. However, the STARS model did not directly process visual input, nor was it tested in a real-world environment. The follow-up study to the STARS model (called ViSTARS) did not provide quantitative data on path errors in the same manner that would allow comparison here. Thus similar performance to the STARS model is assumed.

A movie showcasing the behavioral trajectories of the robot is available on our website: www.socsci.uci.edu/~jkrichma/ABR/lecarl_obstacle_avoidance.mp4.

3.3. Neural activity

Fig. 6 shows the activity of neurons in V1 (Panel A) and MT (Panel B) recorded in a single trial of the third setup in Fig. 5 (blue). For the sake of clarity, only every eighth neuron in the population is shown over a time course of six seconds, which included both the obstacle avoidance and the goal approach. Neurons are organized according to their direction selectivity, where neurons with ID 0–299 were most selective to rightward motion (0°), IDs 300–599 mapped to upward–rightward motion at 45°, IDs 600–899 mapped to upward motion at 90°, IDs 900–1199 mapped to 135°, IDs 1200–1499 mapped to 180°, IDs 1500–1799 mapped to 225°, IDs 1800–2099 mapped to 270°, and IDs 2100–2299 mapped to 315°. Panel A shows the spike trains generated from the rate-based activity of V1 neurons using a Poisson spike generator. The firing of V1 neurons was broad and imprecise in response to stimuli. In contrast, spiking responses of neurons in MT (Panel B) were more selective, and sharper than V1.

As the robot approached the obstacle roughly from 1000 to 2500 ms of the trial shown in Fig. 6, neural activity in both V1 and MT steadily increased in both strength and spatial extent. As activity in MT increased a repeller-like behavioral response was triggered, which in turn introduced even more apparent motion on the retina (roughly from 2200 to 3200 ms). As soon as the obstacle moved out of the robot's view due to successful avoidance (around 3200 ms), activity in MT rapidly dropped off, causing the robot's turning rate to momentarily decrease. With the goal still in view and growing in size, turning rate was now increased with opposite sign (roughly from 3500 to 4500 ms), leading to an attractor-like behavioral response that led the robot directly to the goal. As the robot navigated to the goal (within a 0.5 m distance), the trial was ended manually. Note that near the end of the trial (roughly from 4000 to 5000 ms), neurons in MT rightly perceived the goal object also as an obstacle, which is something the Fajen and Warren
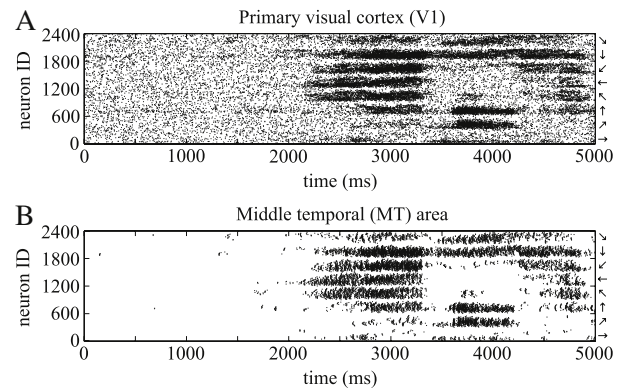

Fig. 6. Rasterplot of neuronal activity for (a) V1 and (b) MT recorded in a single trial where the obstacle was 3 m away and offset by 8° (red colored line in Fig. 5). A dot represents an action potential from a simulated neuron. For the sake of visualization, only every eighth neuron in the population is shown over the time course of six seconds that included both the obstacle avoidance and the goal approach. Neurons are organized according to their direction selectivity, where neurons with ID 0–299 were most selective to rightward motion (0°), IDs 300–599 mapped to upward–rightward motion at 45°, IDs 600–899 mapped to upward motion at 90°, IDs 900–1199 mapped to 135°, IDs 1200–1499 mapped to 180°, IDs 1500–1799 mapped to 225°, IDs 1800–2099 mapped to 270°, and IDs 2100–2299 mapped to 315°. (A) shows the spike trains generated from a Poisson distribution with mean firing rate what the linear filter response is. Average activity in the population was 18.6 ± 15.44 Hz. (B) shows the spike trains of Izhikevich neurons in MT. Average activity in the population was 5.74 ± 8.98 Hz.

model does not take into account. But, because the goal component of the steering command grew more rapidly with size than the obstacle component, the robot continued to approach the goal, rather than starting to avoid it.

Overall network activity and corresponding behavioral output is illustrated in Figs. 7 and 8. Each Panel summarizes the processing of a single visual input frame, which corresponded to a time interval of 50 ms. The indicated time intervals are aligned with the neuronal traces presented in Fig. 6. A frame received at time $t − 50$ ms led to a motor response at time t. This sensorimotor delay was due to the fact that the SNN component of the model needed to be executed for 50 ms. Neural activity during these 50 ms is illustrated as an optic flow field, overlaid on visual input, for both Poisson spike generators in V1 and Izhikevich spiking neurons in MT. Here, every arrow represents the population vector of the activity of eight neurons (selective to the eight directions of motion) coding for the same spatial location. Note that, because these neurons were maximally selective to a speed of 1.5 pixels per frame, vector length does not indicate velocity as is the case in a conventional optic flow field, but rather confidence in the direction estimate. For the sake of clarity, only every fourth pixel location is visualized. The resulting turning rate, $\dot{\theta}$, which was mapped to a PWM signal for the robot's servos, is illustrated using a sliding bar, where the position of the triangle indicates the sign and magnitude of the turning rate. The small horizontal line indicates $\dot{\theta} = 0$. Note that the turning rate depended both on an obstacle term (from optic flow) and a goal term (from color blob detection) (see Section "Posterior parietal cortex (PPC)").
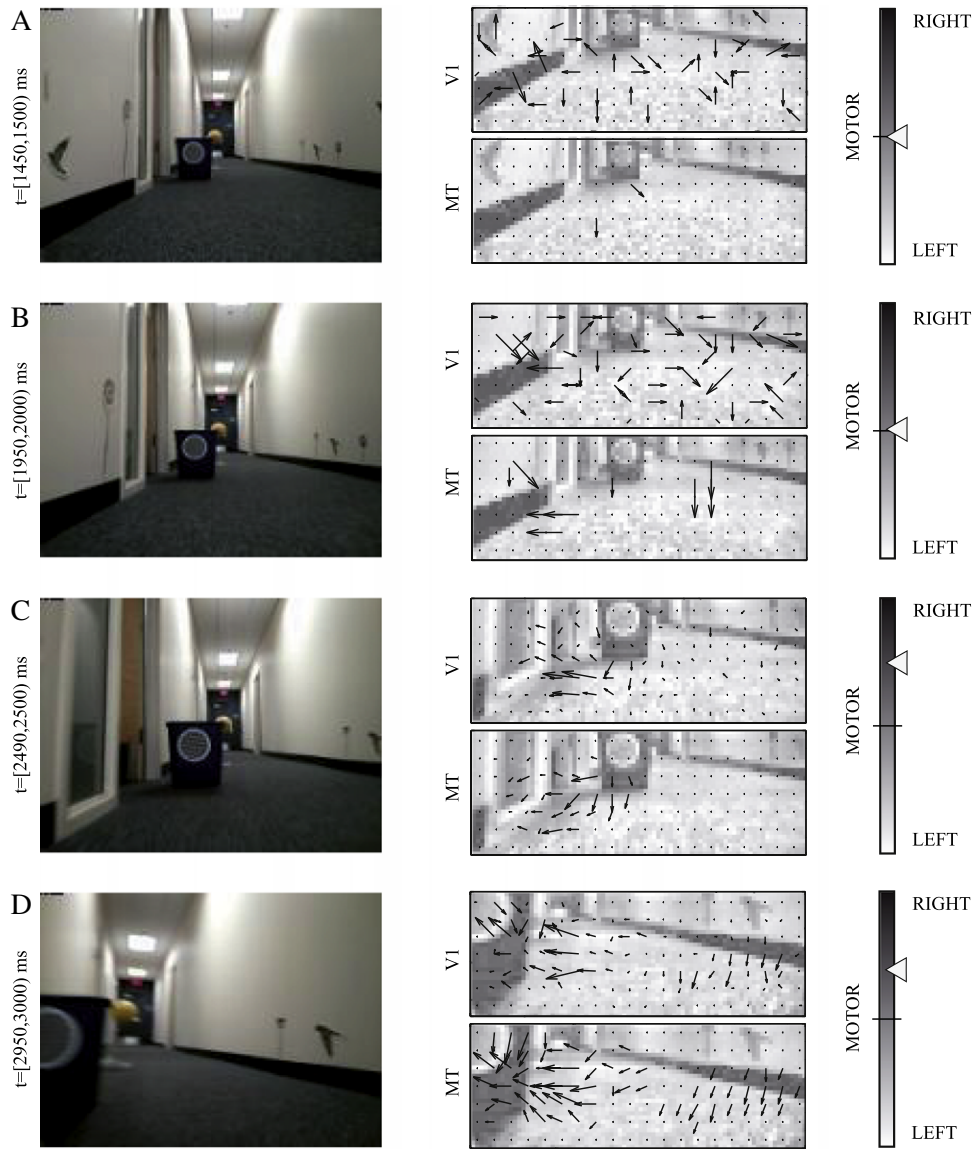
Fig. 7. Overall network activity and corresponding behavioral output during obstacle avoidance in a single trial. Each Panel summarizes the processing of a single visual input frame, which corresponded to a time interval of 50 ms. The indicated time intervals are aligned with the neuronal traces presented in Fig. 6. A frame received at time $t - 50$ ms led to a motor response at time t. Neural activity during these 50 ms is illustrated as an optic flow field, overlaid on visual input, for both Poisson spike generators in V1 and Izhikevich spiking neurons in MT (population vector). Please note that vector length does not correspond to velocity as is the case in a conventional optic flow field, but instead indicates the "confidence" of the direction estimate. For the sake of clarity, only every fourth pixel location is visualized. The resulting turning rate, $\dot{\theta}$, which was mapped to a PWM signal for the robot's servos, is illustrated using a sliding bar, where the position of the triangle indicates the sign and magnitude of the turning rate. The small horizontal line indicates $\dot{\theta} = 0$. Note that the turning rate depended both on an obstacle term (from optic flow) and a goal term (from color blob detection).

Obstacle avoidance is illustrated in Fig. 7. At the beginning of the trial (Panel A), the image of the obstacle on the retina is too small to produce significant optic flow. Instead, V1 responds to arbitrary features of high contrast in the visual scene. Because these responses were relatively low in magnitude and roughly uniformly distributed across the scene, MT neurons successfully suppressed them, effectively treating them as noise. As a result, the turning rate was near zero, informing the robot to steer straight ahead. In the following second (Panels B–D) the obstacle as well as patterned regions in its vicinity generated an increasingly uniform pattern of motion, leading to strong responses in MT and a strongly positive turning rate. In turn, the initiated robot motion introduced even more apparent motion on the retina, leading to a repeller-like behavioral response and successful obstacle avoidance. Note that it is sometimes possible for the motion signals to slightly extend from the object to neighboring contrast-rich regions, which is due to the strong motion pooling in MT.

Goal-directed steering is illustrated in Fig. 8. As soon as the obstacle moved out of the robot's view due to successful avoidance (Panel A), activity in MT rapidly dropped off, causing the robot's turning rate to momentarily decrease. With the goal still in view and growing in size (Panels B–D), turning rate was now increased with opposite sign, eventually overpowering the flow signals that would have instructed the robot to turn away from the goal (Panel D), and instead leading to an attractor-like behavioral response that drew the robot directly to the goal. As the robot navigated to the goal (within a 0.5 m distance; following Panel D), the trial was ended manually.

Overall these results demonstrate how a spiking model of MT can generate sufficient information for the detection of motion
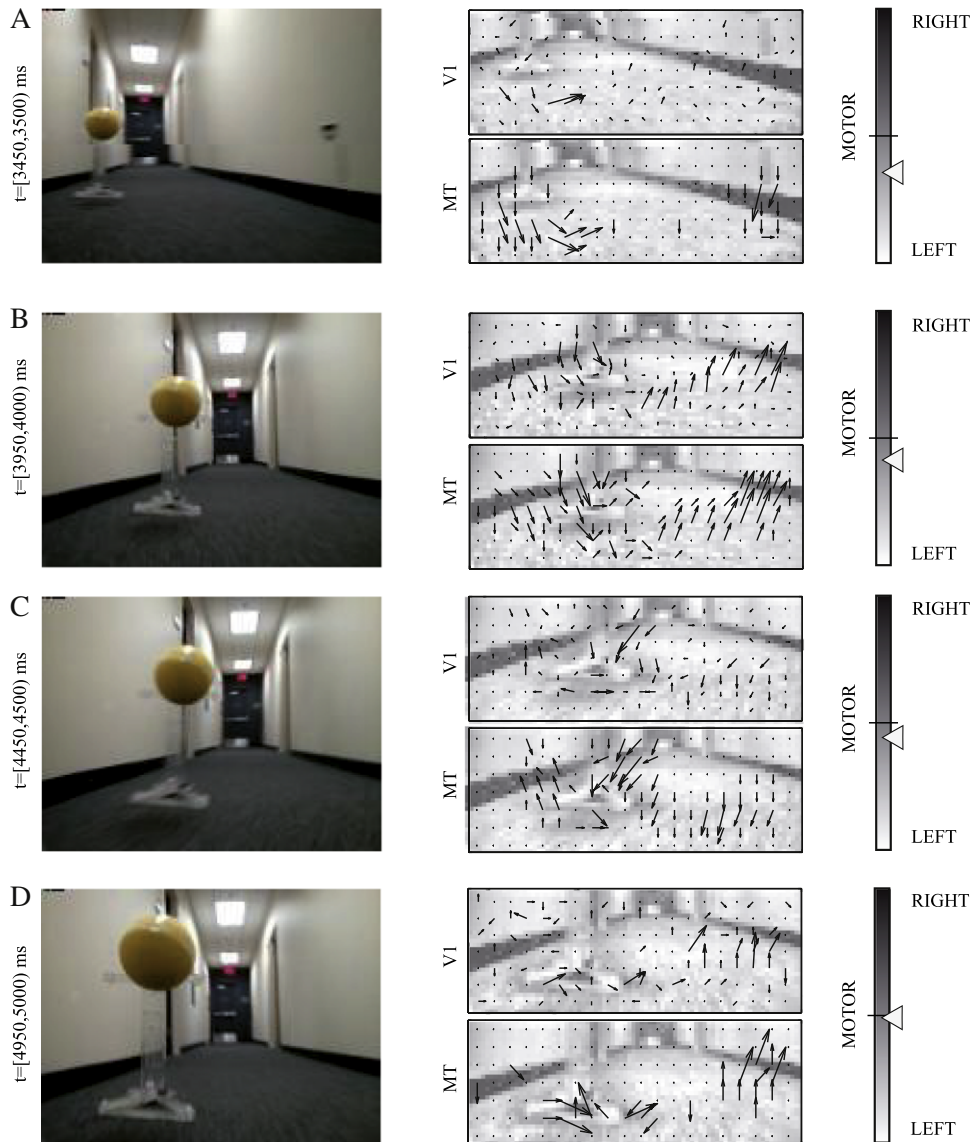
Fig. 8. Overall network activity and corresponding behavioral output during goal-directed steering in a single trial. Each Panel summarizes the processing of a single visual input frame, which corresponded to a time interval of 50 ms. The indicated time intervals are aligned with the neuronal traces presented in Fig. 6. A frame received at time $t - 50$ ms led to a motor response at time t. Neural activity during these 50 ms is illustrated as an optic flow field, overlaid on visual input, for both Poisson spike generators in V1 and Izhikevich spiking neurons in MT (population vector). Please note that vector length does not correspond to velocity as is the case in a conventional optic flow field, but instead indicates the "confidence" of the direction estimate. For the sake of clarity, only every fourth pixel location is visualized. The resulting turning rate, $\dot\theta$, which was mapped to a PWM signal for the robot's servos, is illustrated using a sliding bar, where the position of the triangle indicates the sign and magnitude of the turning rate. The small horizontal line indicates $\dot\theta = 0$. Note that the turning rate depended both on an obstacle term (from optic flow) and a goal term (from color blob detection).

boundaries, which can be used to steer a robot around obstacles in a real-world environment.

3.4. Computational performance

The complete cortical model ran in real-time on a single GPU (NVIDIA GTX 780 with 2304 CUDA cores and 3 GB of GDDR5 memory). In fact, during an average trial, most of the visual frames were processed faster than real-time. In order for the network to satisfy the real-time constraint, it must be able to process up to 20 frames per second, which was the upper bound of the ABR client's frame rate. In practice, the effective frame rate might be reduced due to UDP packet loss and network congestion (depending on the quality and workload of the WiFi connection). In other words, the cortical model must process a single frame in no more than 50 ms.

The majority of the computation time was spent on neural processing in V1 and MT. Every 50 ms, model V1 computed a total

of 201,600 filter responses (80 × 30 pixel image, 28 spatiotemporal filters at three different scales), taking up roughly 15 ms of execution time. Model MT then calculated the temporal dynamics of 40,000 Izhikevich spiking neurons and roughly 1,700,000 conductance-based synapses, which took roughly 25 ms of execution time. During this time, a different thread performed color blob detection using OpenCV, which allowed the total execution time to stay under 50 ms. Compared to these calculations, the time it took to perform CLAHE on the GPU was negligible.

4. Discussion

We presented a neural network modeled after visual motion perception areas in the mammalian neocortex that is controlling a physical robot performing a visually-guided navigation task in the real world. The system described here builds upon our previous work on visual motion perception in large-scale spiking neural networks (Beyeler et al., 2014). The prior system was tested

only on synthetic visual stimuli to demonstrate neuronal tuning curves, whereas the present work demonstrates how perception may relate to action. As this is a first step toward an embodied neuromorphic system, we wanted to ensure that the model could handle real sensory input from the real world. Constructing an embodied model also ensured that the algorithm could handle noisy sensors, imprecise motor responses, as well as sensory and motor delays. The model generated a cortical representation of dense optic flow using thousands of interconnected spiking neurons, determined the position of objects based on motion discontinuities, and combined these signals with the representation of a goal location in order to calculate motor commands that successfully steered the robot around obstacles toward to the goal. Therefore, the present study demonstrates how cortical motion signals in a model of MT might relate to active steering control, and suggests that these signals might be sufficient to generate human-like trajectories through a cluttered hallway. This emergent behavior might not only be difficult to achieve in simulation, but also strengthens the claim that MT contributes to these smooth trajectories in natural settings. This finding exemplifies the importance of embodiment, as behavior is deeply coupled not only with the underlying model of brain function, but also with the anatomical constraints of the physical body.

The behavior of the robot was contingent on the quality of a cortical representation of motion generated in a model of visual area MT. While it is generally assumed that vector-based representations of retinal flow in both V1 and MT are highly accurate, modeling work has suggested that the generation of highly accurate flow representations in complex environments is challenging due to the aperture problem (Baloch & Grossberg, 1997; Bayerl & Neumann, 2004; Chey, Grossberg, & Mingolla, 1997; Simoncelli & Heeger, 1998). As a result, there is a degree of uncertainty in all retinal flow estimations. Nevertheless, the present model is able to generate optic flow fields of sufficient quality to steer a physical robot on human-like trajectories through a cluttered hallway. Paramount to this competence are the receptive fields of neurons in model MT, which are able to vastly reduce ambiguities in the V1 motion estimates (see Figs. 7 and 8) and enhance motion discontinuities through spatial pooling and directional opponent inhibition. As a result, the robot is able to correctly infer the relative angle and position of nearby obstacles, implicitly encoded by the spatial arrangement and overall magnitude of optic flow, and produce steering commands that lead to successful avoidance.

It is interesting to consider whether there is a benefit in using spiking neurons instead of rate based neurons. Under the present experimental conditions, the steering network may have worked just as well with a model based on mean-firing rate neurons. However, there is evidence that humans adjust their steering in response not only to spatial but also to temporal asymmetries in the optic flow field (Duchon & Warren, 2002; Kountouriotis et al., 2013). Therefore, having at our exposal a tested, embodied model of brain function that respects the detailed temporal dynamics of neuronal and synaptic integration is likely to benefit follow-up studies that aim to quantitatively investigate how such spatiotemporal behavior can arise from neural circuitry. In addition, because the main functional components of this system were event-driven spiking neurons, the model has the potential to run on neuromorphic hardware, such as HRL (Srinivasa & Cruz-Albrecht, 2012), IBM True North (Cassidy et al., 2014), NeuroGrid (Boahen, 2006), SpiNNaker (Khan et al., 2008), and BrainScaleS (Schemmel et al., 2010). Some of these platforms are now capable of emulating neural circuits that contain more than a million neurons, at rates that are significantly faster than real time. In addition, because spiking neurons communicate via the AER protocol, the model could also be interfaced with an event-based, neuromorphic vision sensor as a

sensory front-end (Lichtsteiner et al., 2008). Thus, developing complex spiking networks that display cognitive functions or learn behavioral abilities through autonomous interaction may also represent an important initial step into realizing functional large-scale networks on neuromorphic hardware.

Additionally, we have developed software to further extend the functionality of the ABR platform (Oros & Krichmar, 2013b). A set of tools and interfaces exists that provides a standard interface for non-experts to bring their models to the platform, ready for exploration of networked computation principles and applications. The fact that our platform is open source, extensible, and affordable makes ABR highly attractive for embodiment of brain-inspired models. For more information and software downloads, see our website: www.socsci.uci.edu/~jkrichma/ABR.

4.1. Neurophysiological evidence and model alternatives

Human behavioral data suggests that visually-guided navigation might be based on several possible perceptual variables, which can be flexibly selected and weighted depending on the environmental constraints (Kountouriotis et al., 2013; Morice, Francois, Jacobs, & Montagne, 2010). In the case of steering to a stationary goal, evidence suggests that human subjects rely on both optic flow, such that one aligns the heading specified by optic flow with the visual target (Gibson, 1958; Warren, Kay, Zosh, Duchon, & Sahuc, 2001), and egocentric direction, such that one aligns the locomotor axis with the egocentric direction of the goal (Rushton, Harris, Lloyd, & Wann, 1998). Optic flow tends to dominate when there is sufficient visual surface structure in the scene (Li & Warren, 2000; Warren et al., 2001; Wilkie & Wann, 2003), whereas egocentric direction dominates when visual structure is reduced (Rushton et al., 1998; Rushton, Wen, & Allison, 2002). Interestingly, optic flow asymmetries are able to systematically bias the steering of human subjects, even in the presence of explicit path information (Kountouriotis et al., 2013). This finding may hint at the possibility of a cortical analog to the Balance Strategy (Duchon & Warren, 2002; Srinivasan & Zhang, 1997), which suggests that humans adjust their steering in response to both spatial and temporal asymmetries in the optic flow field. However, the present model differs in an important way from the traditional Balance Strategy, in that it does not try to balance a conventional optic flow field. Instead, the model considers only signals generated from motion discontinuities (due to motion processing in MT) in the balance equation, which seems to have similar functional implications as the steering potential function from Huang, Fajen, Fink, and Warren (2006), generating a repeller signal that gets stronger the closer the robot gets to the obstacle.

Visual self-motion is processed in a number of brain areas located in the intraparietal (IPS) and cingulate sulci, including MST, the ventral intraparietal (VIP) region, and the cingulate sulcus visual region (CSv) (Wall & Smith, 2008). Lesions to the IPS lead to navigational impairments when retracting a journey shown from an egocentric viewpoint (Seubert, Humphreys, Muller, & Gramann, 2008). There is evidence that object position and velocity might be encoded by a pathway that includes center–surround cells in MT and cells in the ventral region of the medial superior temporal (MSTv) area (Berezovskii & Born, 2000; Duffy & Wurtz, 1991b; Tanaka et al., 1993). This object information might then be combined with information about self-motion gathered in MST and VIP from cues such as optic flow, eye rotations, and head movements (Andersen, Essick, & Siegel, 1985; Bradley, Maxwell, Andersen, Banks, & Shenoy, 1996; Colby, Duhamel, & Goldberg, 1993; Duffy & Wurtz, 1991a). How neural processing in these regions could relate to steering control has been modeled in detail elsewhere (Browning et al., 2009; Elder et al., 2009).

The present study demonstrates that it might not be necessary to explicitly model these areas in order to explain human psychophysics data about visually-guided steering and obstacle avoidance. In the case of a robot with a single stationary camera, there is no need to calculate eye or head rotations. As a result of this, there is no need for coordinate transformation, because retinotopic coordinates are equivalent to craniotopic coordinates. These simplified anatomical constraints allow for gross simplification of the underlying neural network model without restricting behavioral performance, at least under the present experimental conditions.

4.2. Model limitations

Although the present model is able to generate human-like steering paths for a variety of experimental setups, it does not attempt to explain how the direction of travel is extracted from optic flow, how these signals are made independent of eye rotations and head movements, and how these signals can be converted into a craniotopic reference frame. However, these processes would need to be modeled if one were to consider a robot that could freely move around its head (or its eye). A suitable future study could investigate how these computations and transformations could be achieved for a robot that could freely move around its head (or its eye), for example via a pan/tilt unit. The resulting cortical model might not only reproduce the present behavioral results, but also lead to robust behavior in more complicated experimental setups.

As mentioned above, the behavioral model by Fajen and Warren (2003) computes a steering trajectory using third-person information obtained directly from the scene geometry, namely the distance and angles to the goal and obstacle. In contrast, our model steers from a first-person view using active local sensing, which cannot directly compute the distance and angle to the goal and obstacle. Instead, distance is implicitly represented by neural activity, that is, the overall magnitude of the goal and activity related to obstacles. This is based on the assumption that perceived object size scales inversely with distance. While this assumption is generally true, it might not necessarily enable the model to generalize to arbitrarily complex settings. For example, the model might try to avoid a large but far away obstacle with the same dynamics as it would a small but close-by one. In this case, an accurate estimate of object distance would be imperative. Also, in the current setup it is sometimes possible that motion signals caused by the obstacle can extend into contrast-rich neighboring regions, due to strong motion pooling in MT. A future study could alleviate this issue by giving the robot a means to establish clearer boundaries between objects and their surroundings, perhaps by modeling motion-form interactions (Baloch & Grossberg, 1997).

Binocular vision is an important source of depth perception that influences object segmentation in depth (Xiao, Raiguel, Marcar, & Orban, 1997) as well as the pre-planning and on-line control of movement (Patla & Vickers, 1997). Therefore it is possible that binocular vision might significantly benefit the task at hand (Pauwels, Kruger, Lappe, Worgotter, & Van Hulle, 2010). A suitable follow-up study would thus be to quantify the contribution of depth perception to the quality of stimulus responses in areas MT as well as to the behavioral performance of the robotic agent. It is interesting to note that most MT cells in the macaque receive balanced input from both eyes (Maunsell & Van Essen, 1983), but it turns out that neurons in (at least) macaque MT are not strongly tuned to motion in depth; i.e., no units are truly activated for stimuli changing disparity, which would simulate trajectories with components of motion toward or away from the animal (Maunsell & Van Essen, 1983). Therefore, extending the single camera version of the model by adding areas important for segmenting a scene and recognizing objects (Baloch & Grossberg, 1997), would also improve planning trajectories through space.

Because vergence is not a factor for distances on the order of meters, scene contributions from brain areas such as V4, MST, and parietal cortex are probably important for long-range trajectory planning. Moreover, these areas strongly interact with area MT.

Also, evidence suggests that humans use a different behavioral strategy when it comes to intercepting moving targets (Fajen & Warren, 2007). Having an agent steer toward a moving target according to the present behavioral dynamics is likely to result in pursuit behavior (as opposed to interception behavior), in which the agent always lags behind the target. Thus we do not expect the model to generalize to moving targets. However, it might be possible to extend the model to account for these scenarios. The superior parietal lobe (SPL) might be involved in encoding future path information, such as the location of targets and obstacles, which are indicative of impending changes in heading and using this for the purpose of accurately timing motor responses (Billington et al., 2013; Field et al., 2007). Computing and updating of object locations in space, which might be relevant especially for moving targets, might be done by a network involving the precuneus and dorsal precentral gyrus (Leichnetz, 2001; Parvizi, Van Hoesen, Buckwalter, & Damasio, 2006). However, not much is known about the neuronal representation of these signals.

4.3. Practical implications

The present work might be of interest to the neuroscientist, neurorobotics, and neuromorphic engineering communities for the following reasons.

First, we have shown that the present approach is practical for studying the link between neural circuitry and behavioral performance. The real-world consequences of neural activity can be observed in real-time through visualization software and analyzed off-line. The behavioral responses were comparable to psychophysical data and the neuronal responses were comparable to neurophysiological data. We believe this to be a powerful approach to studying models of neural circuitry in real-world environments.

Second, we have shown that the present system can handle neural network models of non-trivial size. By making use of the CUDA programming framework, we were able to accelerate computationally intensive parts of the model on a GPU. Every 50 ms, model V1 computed a total of 201,600 filter responses (80×30 pixel image, 28 spatiotemporal filters at three different scales), and model MT calculated the temporal dynamics of 40,000 Izhikevich spiking neurons and roughly 1,700,000 conductance-based synapses. If necessary, the execution of the model could be sped up further, for example by parallelizing processing in the goal and obstacle components of the model.

Third, implementing the entire model using spiking neurons would make the model amenable to emulation on neuromorphic hardware (Cassidy et al., 2014; Schemmel et al., 2010; Srinivasa & Cruz-Albrecht, 2012). This could enable the development of a self-contained, fully autonomous neurorobotic platform that combines the algorithmic advantages of the present model with the speed, efficiency, and scalability of neuromorphic hardware. In addition, since spiking neural networks communicate via the AER protocol, the model could be interfaced with an event-based, neuromorphic camera as a sensory front-end (Lichtsteiner et al., 2008). This would make it possible for the cortical network to react more quickly (at least on the millisecond scale, but theoretically even on the microsecond scale) to temporal changes in optic flow. It would also open the door for a complete neuromorphic vision system that operates with low power and rapid responses.

Fourth, the system presented here is open source, extensible, and affordable. The complete ABR source code is hosted on GitHub (www.github.com/UCI-ABR), and our CARLsim simulator

can be freely obtained from our website (www.socsci.uci.edu/~jkrichma/CARLsim). The ABR framework can be combined with a variety of R/C based vehicles, sensors, actuators, and C/C++ based software components. For example, it is straightforward to plug additional software components (such as image transformation with OpenCV) into the client–server loop, or mount IR sensors on the robot and read out the sensory values directly with the ABR client software. We also provide online instructions and video tutorials to assemble the ABR platform, which has an estimated cost of $200 (excluding the phone). Because of this, we believe ABR to be an attractive platform for students and researchers alike that will simplify both the development of neurorobotics platforms as well as the study of how neural machinery can be used to realize cognitive function.

Acknowledgments

References

Allman, J., Miezin, F., & McGuinness, E. (1985). Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception, 14*, 105–126.

Andersen, R. A., Essick, G. K., & Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science, 230*, 456–458.

Baloch, A. A., & Grossberg, S. (1997). A neural model of high-level motion processing: line motion and formotion dynamics. *Vision Research, 37*, 3037–3059.

Bayerl, P., & Neumann, H. (2004). Disambiguating visual motion through contextual feedback modulation. *Neural Computation, 16*, 2041–2066.

Berezovskii, V. K., & Born, R. T. (2000). Specificity of projections from wide-field and local motion-processing regions within the middle temporal visual area of the owl monkey. *Journal of Neuroscience, 20*, 1157–1169.

Beyeler, M., Carlson, K.D., Shuo-Chou, T., Dutt, N., & Krichmar, J.L. (2015). CARLsim 3: A user-friendly and highly optimized library for the creation of neurobiologically detailed spiking neural networks. In *IEEE international joint conference on neural networks*. Killarney, Ireland.

Beyeler, M., Richert, M., Dutt, N. D., & Krichmar, J. L. (2014). Efficient spiking neural network model of pattern motion selectivity in visual cortex. *Neuroinformatics, 12*, 435–454.

Billington, J., Wilkie, R. M., & Wann, J. P. (2013). Obstacle avoidance and smooth trajectory control: neural areas highlighted during improved locomotor performance. *Frontiers in Behavioral Neuroscience, 7*, 9.

Boahen, K. (2006). Neurogrid: emulating a million neurons in the cortex. In *Conf. proc. IEEE eng. med. biol. soc., suppl.* (p. 6702).

Bonin-Font, F., Ortiz, A., & Oliver, G. (2008). Visual navigation for mobile robots: A survey. *Journal of Intelligent & Robotic Systems, 53*, 263–296.

Born, R. T. (2000). Center–surround interactions in the middle temporal visual area of the owl monkey. *Journal of Neurophysiology, 84*, 2658–2669.

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience, 28*, 157–189.

Bradley, D. C., & Andersen, R. A. (1998). Center–surround antagonism based on disparity in primate area MT. *Journal of Neuroscience, 18*, 7552–7565.

Bradley, D. C., Maxwell, M., Andersen, R. A., Banks, M. S., & Shenoy, K. V. (1996). Mechanisms of heading perception in primate visual cortex. *Science, 273*, 1544–1547.

Britten, K. H., & van Wezel, R. J. (1998). Electrical microstimulation of cortical area MST biases heading perception in monkeys. *Nature Neuroscience, 1*, 59–63.

Browning, N. A., Grossberg, S., & Mingolla, E. (2009). Cortical dynamics of navigation and steering in natural scenes: Motion-based object segmentation, heading, and obstacle avoidance. *Neural Networks, 22*, 1383–1398.

Cassidy, A. S., Alvarez-Icaza, R., Akopyan, F., Sawada, J., Arthur, J. V., Merolla, P. A., & Modha, D. S. (2014). Real-time scalable cortical computing at 46 giga-synaptic OPS/watt with ∼100× speedup in time-to-solution and ∼100,000× reduction in energy-to-solution. In *Proceedings of the international conference for high performance computing, networking, storage and analysis* (pp. 27–38). New Orleans, Louisana: IEEE Press.

Chey, J., Grossberg, S., & Mingolla, E. (1997). Neural dynamics of motion grouping: from aperture ambiguity to object speed and direction. *Journal of the Optical Society of America A-Optics Image Science and Vision, 14*, 2570–2594.

Colby, C. L., Duhamel, J. R., & Goldberg, M. E. (1993). Ventral intraparietal area of the macaque: anatomic location and visual response properties. *Journal of Neurophysiology, 69*, 902–914.

Duchon, A. P., & Warren, W. H., Jr. (2002). A visual equalization strategy for locomotor control: of honeybees, robots, and humans. *Psychological Science, 13*, 272–278.

Duffy, C. J., & Wurtz, R. H. (1991a). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology, 65*, 1329–1345.

Duffy, C. J., & Wurtz, R. H. (1991b). Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli. *Journal of Neurophysiology, 65*, 1346–1359.

Duffy, C. J., & Wurtz, R. H. (1997). Planar directional contributions to optic flow responses in MST neurons. *Journal of Neurophysiology, 77*, 782–796.

Eifuku, S., & Wurtz, R. H. (1998). Response to motion in extrastriate area MSTl: center–surround interactions. *Journal of Neurophysiology, 80*, 282–296.

Elder, D. M., Grossberg, S., & Mingolla, E. (2009). A neural model of visually guided steering, obstacle avoidance, and route selection. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 1501–1531.

Fajen, B. R., & Warren, W. H. (2003). Behavioral dynamics of steering, obstacle avoidance, and route selection. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 343–362.

Fajen, B. R., & Warren, W. H. (2007). Behavioral dynamics of intercepting a moving target. *Experimental Brain Research, 180*, 303–319.

Field, D. T., Wilkie, R. M., & Wann, J. P. (2007). Neural systems in the visual control of steering. *Journal of Neuroscience, 27*, 8002–8010.

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience, 14*, 1195–1201.

Gibson, J. J. (1958). Visually controlled locomotion and visual orientation in animals. *British Journal of Psychology, 49*, 182–194.

Gu, Y., Watkins, P. V., Angelaki, D. E., & DeAngelis, G. C. (2006). Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. *Journal of Neuroscience, 26*, 73–85.

Huang, W. H., Fajen, B. R., Fink, J. R., & Warren, W. H. (2006). Visual navigation and obstacle avoidance using a steering potential function. *Robotics and Autonomous Systems, 54*, 288–299.

Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks, 14*, 1569–1572.

Khan, M., Lester, D., Plana, L., Rast, A., Jin, X., & Painkras, E. (2008). SpiNNaker: Mapping neural networks onto a massively-parallel chip multiprocessor. In *IEEE international joint conference on neural networks* (pp. 2849–2856).

Kountouriotis, G. K., Shire, K. A., Mole, C. D., Gardner, P. H., Merat, N., & Wilkie, R. M. (2013). Optic flow asymmetries bias high-speed steering along roads. *Journal of Vision, 13*, 23.

Krichmar, J. L., & Edelman, G. M. (2006). Principles underlying the construction of brain-based devices. In T. Kovacs, & J. A. R. Marshall (Eds.), *Adaptation in artificial and biological systems. Vol. 6* (pp. 37–42). Bristol, UK: Society for the Study of Artificial Intelligence and the Simulation of Behaviour.

Leichnetz, G. R. (2001). Connections of the medial posterior parietal cortex (area 7m) in the monkey. *The Anatomical Record, 263*, 215–236.

Li, L., & Warren, W. H., Jr. (2000). Perception of heading during rotation: sufficiency of dense motion parallax and reference objects. *Vision Research, 40*, 3873–3894.

Lichtsteiner, P., Posch, C., & Delbruck, T. (2008). A 128× 128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid-State Circuits, 43*, 566–576.

Liu, S.C., van Schaik, A., Minch, B.A., & Delbruck, T. (2010). Event-based 64-channel binaural silicon cochlea with Q enhancement mechanisms. In *2010 IEEE international symposium on circuits and systems* (pp. 2027–2030).

Maunsell, J. H., & Van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology, 49*, 1148–1167.

Morice, A. H., Francois, M., Jacobs, D. M., & Montagne, G. (2010). Environmental constraints modify the way an interceptive action is controlled. *Experimental Brain Research, 202*, 397–411.

Myme, C. I., Sugino, K., Turrigiano, G. G., & Nelson, S. B. (2003). The NMDA-to-AMPA ratio at synapses onto layer 2/3 pyramidal neurons is conserved across prefrontal and visual cortices. *Journal of Neurophysiology, 90*, 771–779.

Nageswaran, J. M., Dutt, N., Krichmar, J. L., Nicolau, A., & Veidenbaum, A. V. (2009). A configurable simulation environment for the efficient simulation of large-scale spiking neural networks on graphics processors. *Neural Networks, 22*, 791–800.

Oros, N., & Krichmar, J.L. (2012). Neuromodulation, attention and localization using a novel ANDROID robotic platform. In *IEEE international conference on development and learning and epigenetic robotics. San Diego, CA*.

Oros, N., & Krichmar, J.L. (2013a). Android™ based robotics: Powerful, flexible and inexpensive robots for hobbyists, educators, students and researcher. In. Cognitive Anteater Robotics Laboratory: University of California, Irvine.

Oros, N., & Krichmar, J.L. (2013b). Smartphone based robotics: Powerful, flexible and inexpensive robots for hobbyists, educators, students and researchers. In CECS technical report. Center for Embedded Computer Systems: University of California, Irvine (pp. 13–16).

Parvizi, J., Van Hoesen, G. W., Buckwalter, J., & Damasio, A. (2006). Neural connections of the posteromedial cortex in the macaque. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 1563–1568.

Patla, A. E., & Vickers, J. N. (1997). Where and when do we look as we approach and step over an obstacle in the travel path? *NeuroReport, 8*, 3661–3665.

Pauwels, K., Kruger, N., Lappe, M., Worgotter, F., & Van Hulle, M. M. (2010). A cortical architecture on parallel hardware for motion processing in real time. *Journal of Vision, 10*, 18.

Raiguel, S., Van Hulle, M. M., Xiao, D. K., Marcar, V. L., & Orban, G. A. (1995). Shape and spatial distribution of receptive fields and antagonistic motion surrounds in the middle temporal area (V5) of the macaque. *European Journal of Neuroscience, 7*, 2064–2082.

Richert, M., Nageswaran, J. M., Dutt, N., & Krichmar, J. L. (2011). An efficient simulation environment for modeling large-scale cortical processing. *Frontiers in Neuroinformatics, 5*, 19.

Rushton, S. K., Harris, J. M., Lloyd, M. R., & Wann, J. P. (1998). Guidance of locomotion on foot uses perceived target location rather than optic flow. *Current Biology, 8*, 1191–1194.

Rushton, S.K., Wen, J., & Allison, R.S. (2002). Egocentric direction and the visual guidance of robot locomotion background, theory and implementation. In *Biologically motivated computer vision, proceedings. Vol. 2525* (pp. 576–591).

Schemmel, J., Bruderle, D., Grubl, A., Hock, M., Meier, K., & Millner, S. (2010). A wafer-scale neuromorphic hardware system for large-scale neural modeling. In *Circuits and systems (ISCAS), Proceedings of 2010 IEEE international symposium on* (pp. 1947–1950).

Seubert, J., Humphreys, G. W., Muller, H. J., & Gramann, K. (2008). Straight after the turn: the role of the parietal lobes in egocentric space processing. *Neurocase, 14*, 204–219.

Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Research, 38*, 743–761.

Srinivasa, N., & Cruz-Albrecht, J. M. (2012). Neuromorphic adaptive plastic scalable electronics analog learning systems. *IEEE Pulse, 3*, 51–56.

Srinivasan, M. V., & Zhang, S. W. (1997). Visual control of honeybee flight. *EXS, 84*, 95–113.

Tanaka, K., Sugita, Y., Moriya, M., & Saito, H. (1993). Analysis of object motion in the ventral part of the medial superior temporal area of the macaque visual cortex. *Journal of Neurophysiology, 69*, 128–142.

Wall, M. B., & Smith, A. T. (2008). The representation of egomotion in the human brain. *Current Biology, 18*, 191–194.

Warren, W. H., Jr., Kay, B. A., Zosh, W. D., Duchon, A. P., & Sahuc, S. (2001). Optic flow is used to control human walking. *Nature Neuroscience, 4*, 213–216.

Wen, B., & Boahen, K. (2009). A silicon cochlea with active coupling. *IEEE Transaction on Biomedical Circuits and Systems, 3*, 444–455.

Wilkie, R., & Wann, J. (2003). Controlling steering and judging heading: retinal flow, visual direction, and extraretinal information. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 363–378.

Xiao, D. K., Raiguel, S., Marcar, V., & Orban, G. A. (1997). The spatial distribution of the antagonistic surround of MT/V5 neurons. *Cerebral Cortex, 7*, 662–677.