

# Proactive Inhibitory Control and Attractor Dynamics in Countermanding Action: A Spiking Neural Circuit Model

Chung-Chuan Lo,<sup>1,2</sup> Leanne Boucher,<sup>3</sup> Martin Paré,<sup>4</sup> Jeffrey D. Schall,<sup>3</sup> and Xiao-Jing Wang<sup>1</sup>

<sup>1</sup>Department of Neurobiology and Kavli Institute for Neuroscience, Yale University, New Haven, Connecticut 06510, <sup>2</sup>Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu 30013, Taiwan, <sup>3</sup>Center for Integrative & Cognitive Neuroscience, Vanderbilt Vision Research Center, Department of Psychology, Vanderbilt University, Nashville, Tennessee 37240, and <sup>4</sup>Centre for Neuroscience Studies and Departments of Physiology and Psychology, Queen's University, Kingston, Ontario K7L 3N6, Canada

Flexible behavior depends on the brain's ability to suppress a habitual response or to cancel a planned movement whenever needed. Such inhibitory control has been studied using the countermanding paradigm in which subjects are required to withhold an imminent movement when a stop signal appears infrequently in a fraction of trials. To elucidate the circuit mechanism of inhibitory control of action, we developed a recurrent network model consisting of spiking movement (GO) neurons and fixation (STOP) neurons, based on neurophysiological observations in the frontal eye field and superior colliculus of behaving monkeys. The model places a premium on the network dynamics before the onset of a stop signal, especially the experimentally observed high baseline activity of fixation neurons, which is assumed to be modulated by a persistent top-down control signal, and their synaptic interaction with movement neurons. The model simulated observed neural activity and fit behavioral performance quantitatively. In contrast to a race model in which the STOP process is initiated at the onset of a stop signal, in our model whether a movement will eventually be canceled is determined largely by the proactive top-down control and the stochastic network dynamics, even before the appearance of the stop signal. A prediction about the correlation between the fixation neural activity and the behavioral outcome was verified in the neurophysiological data recorded from behaving monkeys. The proposed mechanism for adjusting control through tonically active neurons that inhibit movement-producing neurons has significant implications for exploring the basis of impulsivity associated with psychiatric disorders.

## Introduction

Cognitively controlled behavior relies on our ability to suppress prepotent responses to external stimuli whenever demanded (Logan and Cowan, 1984; Verbruggen and Logan, 2009a). With insufficient inhibitory control, a condition often associated with psychiatric disorders such as attention-deficit hyperactivity disorder (ADHD) (Schachar and Logan, 1990; Armstrong and Munoz, 2003; Hanisch et al., 2006) and schizophrenia (Boudet et al., 2005; Donohoe et al., 2006), prepotent responses tend to dominate at the expense of cognitively controlled deliberate actions (Aron et al., 2003). Inhibitory control can be triggered by a response to a stimulus signaling stop, but can also be performed by an endogenous top-down signal (Brass and Haggard, 2007), which presumably originates from frontal cortex (Naito and Matsumura, 1994; Konishi et al., 1998; Tinsley and Everling,

2002; Ridderinkhof et al., 2004; Johnston and Everling, 2006; Stuphorn and Schall, 2006). In contrast to passively responding to an external stop signal, a top-down control signal is likely to be constantly present, especially when the need for withholding movements is anticipated. Imagine that you are in a city known for its drivers' tendency to run a red light. To be safe while crossing a street, a sound strategy is to hold your steps for a short while after the pedestrian light turns green. Indeed, various behavioral tasks of inhibitory control showed that response times to a go signal are longer when a stop signal is expected compared with when it is not (Hanisch et al., 2006; Stuphorn and Schall, 2006; Verbruggen et al., 2006; Verbruggen and Logan, 2009b). This observation poses fundamental questions regarding the neural circuit mechanisms of inhibitory control: How does this proactive top-down inhibitory control interact with the bottom-up drive for movement at the neuronal levels before a stop signal appears, and how does this interaction affect a subject's response to the stop signal?

The countermanding (or stop-signal) task has been used for decades to investigate normal and disordered control of thought and action (Logan and Cowan, 1984; Verbruggen and Logan, 2009a). In the task, a subject has to withhold the response to a go signal when an infrequent and delayed stop signal appears (see Materials and Methods). The success of the countermanding task is derived from the effectiveness of a formal mathematical model (independent race model) that accounts for behavioral performance as the outcome of a race between independent GO and

Received Dec. 27, 2008; revised June 4, 2009; accepted June 7, 2009.

This work was supported by National Institutes of Health Grant 2-R01-MH062349 and the Swartz Foundation (C.C.L. and X.J.W.), by Robin and Richard Patton through the E. Bronson Ingram Chair in Neuroscience and National Science Foundation Grants BCS0218507, F32-EY016679, and R01-MH55806 (J.D.S. and L.B.), and by the Canadian Institutes of Health Research and the EJLB Foundation (M.P.). We thank Dr. Gordon D. Logan for valuable comments. C.C.L. thanks Dr. S. Fusi for the early work in the neural simulator used in this study and the National Center for High-Performance Computing, Taiwan for providing computational resources.

Correspondence should be addressed to Xiao-Jing Wang, Department of Neurobiology, Yale University, New Haven, CT 06510. E-mail: xjwang@yale.edu.

C.-C. Lo's present address: Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu 30013, Taiwan.

DOI:10.1523/JNEUROSCI.6164-08.2009

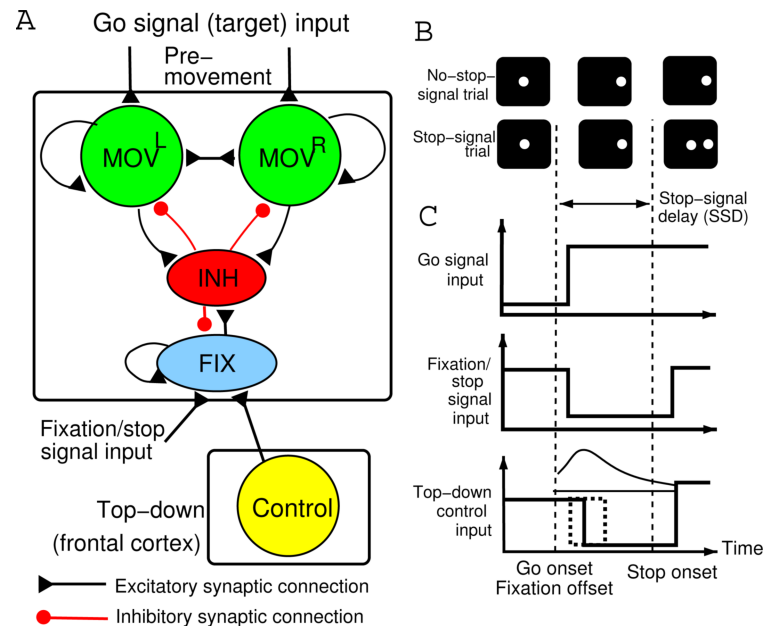
Copyright © 2009 Society for Neuroscience 0270-6474/09/29059-13\$15.00/0

STOP processes (Logan and Cowan, 1984). Although the model provides a measure of the duration of the covert STOP process, it did not specify just how the STOP process prevented the GO process from finishing. Neurophysiological recordings from the frontal eye field and the superior colliculus of rhesus monkeys suggested that the GO and STOP processes may be instantiated by movement and fixation neurons, respectively, and that the antagonistic interplay between them may be responsible for the interruption of the GO process (Hanes et al., 1998; Paré and Hanes, 2003). To address this question explicitly, an interactive race model (Boucher et al., 2007) with mutually inhibitory GO and STOP units was developed. The model showed how delayed and potent inhibition of the GO unit by the STOP unit could account for the behavioral performance as well as the profile of neural modulation of movement neurons (GO unit) and fixation neurons (STOP unit) after the onset of the stop signal.

However, neurophysiological observations have revealed a strong tonic activity of fixation neurons as well as reciprocity between activity of movement and fixation neurons before the stop signal appears, which was not accounted for in the interactive race model (where the STOP unit or process is completely inactive before the appearance of an external stop cue). Since the movement neurons are influenced by the fixation neurons even in the absence of an external stop stimulus, the dynamic interplay between the two cell types before the stop signal may be critical for our mechanistic understanding of inhibitory control in countermanding saccades. This study aims at testing this hypothesis using a biophysically realistic model endowed with a top-down signal, which is persistent internally and may originate from the prefrontal cortex (Funahashi et al., 1989, 1993; Wang, 2001). Part of the result of this work was published in an abstract form (Lo and Wang, 2007).

## Materials and Methods

**Behavioral task.** The proposed model of inhibitory control was designed for explaining neuronal and behavioral observations in the saccade countermanding task. In every trial, the task started with a fixation signal located on the center of the screen. The fixation signal was then turned off while a peripheral target (go signal) was turned on at the same time. In a small fraction (typically 1/4 to 1/3) of randomly interleaved trials, after a variable delay [stop signal delay (SSD)] after the onset of the target, the center fixation signal was turned on again serving as the stop signal. A subject's task was to make a saccadic eye movement into the target as quickly as possible but withhold the movement when a stop signal appeared. In the present study, trials with or without the stop signal presented are referred to as stop-signal trials or no-stop-signal trials, respectively. To be consistent with previous saccade countermanding experiments with nonhuman primates (Hanes et al., 1998; Paré and Hanes, 2003), we used four SSD values: 69, 117, 169, and 217 ms when we compared the simulated behavior to the experiment. In the model, initiation of a saccadic eye movement was defined as the crossing of a fixed threshold (70 spike per second, sp/s) in the population firing rate of



**Figure 1.** A spiking neural network model of inhibitory control for the countermanding task. **A**, The schematic model architecture. The pre-movement module consists of two populations of MOV<sup>L</sup> and MOV<sup>R</sup>, a population of INH, and a population of FIX. The control module consists of only one population of excitatory neurons. **B**, The saccade countermanding task. In both no-stop-signal and stop-signal trials a central fixation signal is presented for a period of time, then the fixation signal is turned off when a peripheral target for a saccade (go signal) is turned on. In stop-signal trials, after a variable delay after the go signal, the central fixation spot is turned back on, serving as the stop signal. The two types of trials are randomly interleaved. **C**, Time courses of the input in stop-signal trials. The left and right dashed lines mark the go and stop signal onsets respectively. After the fixation signal offset, the top-down control remains for a variable interval (holding period). The duration of the holding period follows a Gaussian distribution illustrated above the panel. The holding period mimics a subject's endogenous tendency to delay the saccade response.

movement neurons plus a 10 ms ballistic period of movement production. A stop-signal trial was counted as failed (noncanceled trial) if a saccade was initiated within 700 ms after the onset of the go signal. Otherwise, the trial was successful (canceled trial). The fixation/stop signal was modeled by an input to the fixation neurons while the target signal was modeled by an input to the movement neurons.

**Subjects and experimental data.** In the presented study we fit the model to behavioral data from one rhesus monkey (*Macaca mulatta*) and compared the model simulations to neural data from four rhesus monkeys (two for frontal eye field and two for superior colliculus). All experimental data were previously reported: the neural data of superior colliculus were derived from data reported in the study by Paré and Hanes (2003), and the neural data of frontal eye field and behavioral data were reported in the study by Hanes et al. (1998).

**Neural network model.** The neural network model consists of a pre-movement module that directly controls initiation of movements and a control module that provides a top-down inhibitory control over the pre-movement module (Fig. 1A). For the sake of simplicity, we assumed that the pre-movement module encodes two target stimuli (left or right go signals) and the fixation/stop signal stimulus. To capture observations in various experiments, we further assume that each stimulus activates a small subpopulation of neurons in the pre-movement module and the remaining neurons do not respond to any of the three stimuli. Based on these assumptions, the pre-movement module consists of five neural populations: excitatory gaze-shifting movement neurons (MOV<sup>R</sup>, MOV<sup>L</sup>), inhibitory interneurons (INH), excitatory gaze-holding fixation neurons (FIX), and non-selective excitatory neurons (NSE). NSE is included to mimic the large number of neurons that do not respond to the target stimuli and fixation/stop signal stimulus but may selective to other stimuli that are irrelevant to the present study. For example, in situations where there are multiple response options (e.g., three or four possible target positions), NSE neurons would become activated. In the current model, NSE neurons only receive the background spike input and main-

tain a baseline activity (several sp/s) in the task. The control module contains only one population of excitatory neurons.  $MOV^R$  and  $MOV^L$  neurons receive inputs driven by the right and left saccade targets, respectively (Fig. 1B, C) and form mutual inhibition through INH neurons. The recurrent excitation of  $MOV^R$  and  $MOV^L$  neurons and slow synaptic dynamics mediated by NMDA receptors produce a ramping activity and, together with inhibition, give rise to winner-take-all dynamics (Wang, 2002; Lo and Wang, 2006), an activity pattern consists with neural responses observed in various brain areas during saccade-related behavioral tasks (Hanes and Schall, 1996; Dorris et al., 1997; Hanes et al., 1998; Paré and Hanes, 2003).

Specifically, the premovement module models neural activity observed in primate frontal eye field and superior colliculus (Hanes et al., 1998; Paré and Hanes, 2003). In response to a target input, the activity of the MOV population ramps up over time and triggers a saccade when a fixed threshold is crossed. The FIX neurons receive two inputs: a visual input driven by the fixation/stop signal (Fig. 1B, C) and an excitatory projection from neurons in the control module. The time course of each input is shown in Figure 1C. Note that because the control module projects to the FIX neurons, the excitatory connection produces an inhibitory effect on MOV neurons. We further assume that in each trial after the offset of the fixation signal, the activity of the control module persists for a variable interval (a holding period) (Fig. 1C, bottom panel), corresponding to the subject's endogenous tendency to hold fixation.

Each population consists of spiking neurons simulated using the leaky integrate-and-fire model. The number of neurons in populations  $MOV^R$ ,  $MOV^L$ , INH, FIX, NSE, and control are 240, 240, 400, 240, 1120, and 120, respectively. Each neuron receives external spike inputs with Poisson statistics serving as the background noise. For all excitatory neurons in the premovement module, the background noise is a mixture of AMPA receptor mediated excitatory and  $GABA_A$  receptor mediated inhibitory inputs, but the net effect is excitatory. For the AMPA inputs, the maximum synaptic conductance was 2.1 nS for excitatory neurons, 1.62 nS for inhibitory interneurons. The overall AMPA input spike rate (in sp/s) was 2900 for each neuron in  $MOV^R$ ,  $MOV^L$ , and NSE, 2400 for each neuron in INH, 2304 for each neuron in FIX, and 1840 for each neuron in control. For the GABA inputs, the overall spike rate (in sp/s) was 675 for each neuron in  $MOV^R$ ,  $MOV^L$  and NSE.

Stimuli and the top-down control signal were modeled as spike inputs with Poisson statistics via AMPA receptors, with the maximum synaptic conductance of 2.1 nS. The spike rate (in sp/s) was 560 for the go signal (input to each MOV neuron), 256 for the fixation and stop signals (input to each FIX neuron), 296 for the inhibitory control during the fixation and holding periods, and 360 for the inhibitory control in the stop period (input to each control neuron).

All synaptic connections between neural populations and within a neural population (recurrent connections) are all-to-all, i.e., every neurons in the source population make synaptic connections to every neurons in the target population. The values of synaptic efficacy  $g$  (in nS) are as follows (for excitatory connections, values are given as AMPA/NMDA):  $g^{MOV^R-MOV^R} = g^{MOV^L-MOV^L} = 0.165/0.1823$ ,  $g^{MOV^R-INH} = g^{MOV^L-INH} = 0.08/0.08705$ ,  $g^{MOV^R-NSE} = g^{MOV^L-NSE} = 0.1/0.11048$ ,  $g^{MOV^R-MOV^L} = g^{MOV^L-MOV^R} = 0.08765/0.096838$ ,  $g^{NSE-MOV^R} = g^{NSE-MOV^L} = 0.08765/0.096838$ ,  $g^{NSE-INH} = 0.08/0.08705$ ,  $g^{NSE-NSE} = 0.1/0.11048$ ,  $g^{FIX-FIX} = 0.066/0.072919$ ,  $g^{FIX-INH} = 0.04/0.043524$ ,  $g^{INH-INH} = 0.9625$ ,  $g^{INH-FIX} = 0.3$ ,  $g^{INH-MOV^R} = g^{INH-MOV^L} = g^{INH-NSE} = 1.25125$ ,  $g^{CNT-FIX} = 0.18/0.198864$ .

We note that we used a higher AMPA/NMDA ratio here than in the cortical decision network model (Wang, 2002; Lo and Wang, 2006). This is because we need a much higher activity ( $\sim 100$  sp/s) in movement neurons than in the previous model (Wang, 2002; Lo and Wang, 2006), and the recurrent excitatory drive saturates at higher neural firing rates when it is dominated by the fast AMPA receptors than by the slow NMDA receptors. We note that with a lower AMPA/NMDA ratio as in the previous model (Wang, 2002; Lo and Wang, 2006), the proposed model can still reproduce behaviors and neural activity patterns observed in the countermanding task, except that the overall firing rate of movement neurons is lower than that observed (Hanes et al., 1998; Paré and Hanes, 2003).

The time interval between the onset of the go signal and the actual sensory input to the model network was set to be 8 ms. This resulted in a neural latency of  $\sim 35$  ms, which is consistent with observation (Hanes et al., 1998; Boucher et al., 2007). The time interval between the onset of the stop signal and the actual input to the model network was determined to be 62 ms (see below, Parameter tuning). For the sake of simplicity, in the present study we did not explicitly model the neural circuit that generates the persistent activity in the control module. Instead, we used a tonic Poisson spike input to drive the control neurons. In each trial a value for the duration of the tonic Poisson input to control neurons (holding period) was randomly drawn from a Gaussian distribution (mean = 113 ms and SD = 95 ms; see below, Parameter tuning), but only with values larger than zero for the obvious reason that the offset of the top-down control should only happen after the offset of the fixation signal.

*Single neuron model.* Each neuron in the model is simulated by the leaky integrate-and-fire model. The membrane potential  $V(t)$  for each neuron obeys the following equation:

$$C_m \frac{dV(t)}{dt} = -g_L(V(t) - V_L) - I_{syn}(t), \quad (1)$$

where  $C_m$  is the capacitance,  $g_L$  is the leak conductance,  $V_L$  is the resting potential, and  $I_{syn}$  is the total synaptic current in the cell.

When the membrane potential  $V(t)$  of each neuron reaches a threshold  $V_{threshold} = -50$  mV, a spike is emitted and  $V(t)$  is set to the reset potential  $V_{reset} = -55$  mV for a refractory period  $T_r = 2$  ms. For inhibitory neurons, we used the following parameters:  $C_m = 0.2$  nF,  $g_L = 20$  nS and  $V_L = -70$  mV. For excitatory neurons, we used  $C_m = 0.5$  nF,  $g_L = 25$  nS, and  $V_L = -70$  mV.

The synaptic current  $I_{syn}(t)$  is described by the following:

$$I_{syn}(t) = g_{AMPA} s_{AMPA}(t)(V(t) - V_E) + \frac{g_{NMDA} s_{NMDA}(t)(V(t) - V_E)}{1 + [Mg^{2+}]e^{-0.062V(t)}/3.57} + g_{GABA} s_{GABA}(t)(V(t) - V_I), \quad (2)$$

where  $V_E (=0$  mV) and  $V_I (= -70$  mV) are the reversal potentials,  $[Mg^{2+}] (=1.0$  mM) is the extracellular magnesium concentration,  $g$  is the synaptic efficacy, and  $s$  is the gating variable. Subscripts in  $g$  and  $s$  denote the receptor type. The gating variable obeys the following:

$$\frac{ds(t)}{dt} = \sum_k \delta(t - t^k) - \frac{s}{\tau} \quad (3)$$

for AMPA and  $GABA_A$  receptors and the following:

$$\frac{ds(t)}{dt} = \alpha(1 - s(t)) \sum_k \delta(t - t^k) - \frac{s}{\tau} \quad (4)$$

for NMDA receptors with  $\alpha = 0.63$ . The decay constant  $\tau$  is 2 ms for AMPA, 5 ms for  $GABA_A$  and 100 ms for NMDA.  $\delta(t - t^k)$  is the  $\delta$  function and  $t^k$  is the time of the  $k$ th presynaptic spike.

*Parameter tuning.* We adopted a procedure to reduce the parameters that are needed for fitting to the data. As a result, we only used three parameters to fit the model to the behavioral data. For cellular parameters such as membrane time constants, membrane conductances etc., we simply used the typical values used in the cortical decision network model (Wang, 2002; Lo and Wang, 2006). Synaptic strengths within and between MOV, INH, and NSE also follow the values used in Wang (2002) and Lo and Wang (2006) except that we used a higher AMPA/NMDA ratio in excitatory connections. Synaptic connections in the rest of the network and the external inputs to MOV, FIX, and control neurons were tuned for the model to reproduce neuronal activity observed in the countermanding task (Hanes et al., 1998).

We aimed to reproduce the following neural behaviors: (1) The activity of movement neurons should reach 100 sp/s when they are fully activated by the visual target input (making a saccade). (2) The baseline activity of fixation neurons during the fixation period is  $\sim 80$  sp/s. (3)

After the offset of the fixation signal, the activity of fixation neurons drops to  $\sim 40$ – $50$  sp/s at the presence of the top-down control. (4) When the fixation neurons are fully activated by the stop signal and a saccade is canceled, the activity of fixation neurons should reach at least 80 sp/s. The behavior (1) can be reproduced by tuning the strength of the target input (go signal) to MOV neurons. The behavior (2) can be reproduced by tuning the total input to FIX neurons during the fixation period. Note that the input to FIX neurons comes from two sources: an input driven by the fixation/stop signal and a top-down signal from control neurons. How much input each of the sources contributes can be determined by considering behaviors (3) and (4). After the offset of the fixation signal, there is no visual input to FIX, so the only input that supports the activity of FIX neurons is the input from control neurons. We set the strength of the input to the control neurons so that they fire at a typical value  $\sim 15$  sp/s and then set the synaptic strength accordingly to generate a desired input to FIX neurons, which should fire at the observed 40–50 sp/s. The obtained input strength is for the fixation and holding periods, and subjects may apply stronger top-down control in response to the stop signal onset, because the subjects need to suppress the activity of MOV neurons that may have already been activated. Assuming that the fixation and stop visual inputs to FIX neurons are the same because of the identical stimuli, we could use behavior (4) to determine the necessary input to control neurons which drives FIX neurons to fire at the observed level ( $>80$  sp/s) after the onset of the stop signal in canceled trials.

Next, we considered the connection strengths between FIX and INH. During the countermanding task, after the onset of the target and offset of the fixation signal, the top-down control alone should still be able to maintain the fixation until the subjects decide to commit a saccade response and turn off the top-down control. This requires a strong inhibition from FIX to MOV neurons in the model, thus sets a minimum value for the FIX-to-INH synaptic strength and a maximum value for the INH-to-FIX synaptic strength. In some noncanceled trials, subjects make eye movements toward the target even at the presence of the stop signal. This requires a strong inhibition from MOV to FIX in the model, thus sets a minimum value for the INH-to-FIX synaptic strength and a maximum value for the FIX-to-INH synaptic strength. These two considerations confine the INH-to-FIX and FIX-to-INH synaptic strengths to narrow ranges. We chose a smaller value for the FIX-to-INH synaptic connection from the range because a stronger FIX-to-INH synaptic connection will completely silence MOV neurons during the holding period, whereas a slow ramping activity is observed for movement neurons right after the onset of the target in frontal eye field (Hanes et al., 1998).

Now we have only a handful parameters left to be determined: the mean and SD of the holding period, the latency of the target signal input [ $D_{go}$ , as in the study by Boucher et al. (2007)] and the latency of the stop signal input [ $D_{stop}$ , as in the study by Boucher et al. (2007)]. These parameters cannot be inconsistent with neural activity but were determined by behavioral data. We decided to further constrain the latency of the target signal input. By setting the latency to 8 ms, the activity of MOV neurons increases significantly  $\sim 30$ – $40$  ms after the onset of the target stimulus, a value observed previously for the same animal we used in the presented study (Boucher et al., 2007). The duration of the holding period affects the saccade reaction times. Therefore we can determine the mean and the SD of the duration of the holding period by finding the values that produce a simulated reaction time distribution that best fits to experimental observations. Finally, we have the stop signal input latency left to be determined. This parameter describes the time interval from the stop signal onset until the actual input into FIX neurons and affects the probability of response cancelling in stop-signal trials. To determine this latency, we find the value of the latency that produces simulated inhibition function and reaction time distributions for noncanceled trials that best fit to the experimental observation.

**Phase plan plot.** We used phase plan plots to analyze the time evolution of neural activity in stop-signal trials by plotting the population firing rate of MOV neurons  $r_{MOV}(t)$  against that of FIX neurons  $r_{FIX}(t)$ . This two-dimensional phase plan is sufficient to describe the dynamics of the proposed six-population network for the following reasons: (1) NSE and INH neurons do not receive external inputs, but are only driven by background noise and by MOV and FIX populations. As a result, the

activities of NSE and INH neurons are completely determined by the activities of MOV and FIX neurons. (2) MOV and FIX neurons receive task-related inputs and their interactions determine whether a saccade is triggered. (3) The control module only serves as a relay of the top-down control input with a stereotyped activity and does not participate in the neural interactions in the premovement module. To help analyze the dynamics of the network, we need to find out the stable states where  $r_{MOV}(t)$  and  $r_{FIX}(t)$  will converge to. To this end, we plotted nullclines (Wilson, 1999) for MOV neurons and FIX neurons. Because the activity of MOV depends on its three main inputs: external stimulus ( $I_{MOV}$ ), recurrent excitation from itself and input from FIX, we can express the time derivative of  $r_{MOV}(t)$  as the following:

$$\frac{dr_{MOV}(t)}{dt} = f_{MOV}(I_{MOV}, \bar{r}_{MOV}, \bar{r}_{FIX}). \quad (5)$$

The MOV nullcline is defined as the steady state  $\bar{r}_{MOV}$  (given by  $dr_{MOV}/dt = 0$ ) as a function of the FIX activity ( $r_{FIX}$ ), for a given external input  $I_{MOV}$ . That is,

$$\frac{dr_{MOV}}{dt} = 0 = f_{MOV}(I_{MOV}, \bar{r}_{MOV}, r_{FIX}). \quad (6)$$

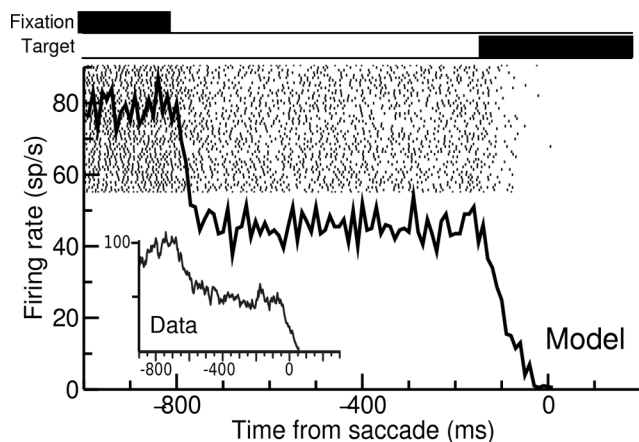
Therefore, the nullcline represents a curve on the phase plan that depicts the relation between  $\bar{r}_{MOV}$  and  $r_{FIX}$ . In addition, the nullcline varies with the external input  $I_{MOV}$ . The nullcline of the FIX neurons is defined similarly as the following:

$$f_{FIX}(I_{FIX}, \bar{r}_{FIX}, r_{MOV}) = 0. \quad (7)$$

Now, an intersection of the MOV and FIX nullclines satisfies both  $dr_{FIX}/dt = 0$  and  $dr_{MOV}/dt = 0$ , hence represents a steady state (equilibrium point) for both populations under the given inputs ( $I_{MOV}$  and  $I_{FIX}$ ). Depending on how the two nullclines intersect each other, an equilibrium point can be stable or unstable. A stable equilibrium point forms an attractor to which the system state may eventually converge. For a complex system such as the proposed spiking neural network model, there is no simple equations describing  $r_{MOV}$  and  $r_{FIX}$ . Therefore one cannot find analytical solutions for the nullclines without simplifying the system. To circumvent this problem, we solve for nullclines numerically using the following procedure: to compute the MOV nullcline, we removed all synaptic inputs to FIX neurons but kept their synaptic projections to other populations unchanged. For a given external input to MOV neurons, we used a Poisson input to drive FIX neurons at a fixed firing rate level, and then recorded the steady state firing rate of MOV neurons. We obtained the MOV nullcline by repeating this procedure with various fixed activity of FIX neurons and plotting the steady state activity of MOV neurons as a function of FIX activity. Similarly, FIX nullclines for different external inputs were obtained by removing all synaptic input to MOV neurons and recording the steady state firing rates of FIX neurons as functions of MOV activity.

## Results

As schematically shown in Figure 1, the core of our model circuit consists of MOV neurons and FIX neurons, instantiating GO and STOP processes respectively. Each population contains spiking neurons that are interconnected by recurrent excitatory synapses. The MOV and FIX neural populations compete against with each other through a shared pool of INH neurons. Furthermore, the model assumes that fixation neurons exhibit an elevated baseline activity, as observed in fixation neurons of the frontal eye field and superior colliculus of behaving monkeys. MOV neural activity cannot accumulate to trigger a behavioral response, unless it is disinhibited from FIX neurons. In our model, tonic firing of FIX neurons is sustained in part by a top-down signal from a control module. At the offset of an external fixation input, this internal control signal lingers for a variable period while MOV neurons remain suppressed by FIX neurons. Such an endogenous holding process is evident in a gap saccade task, in which subjects must



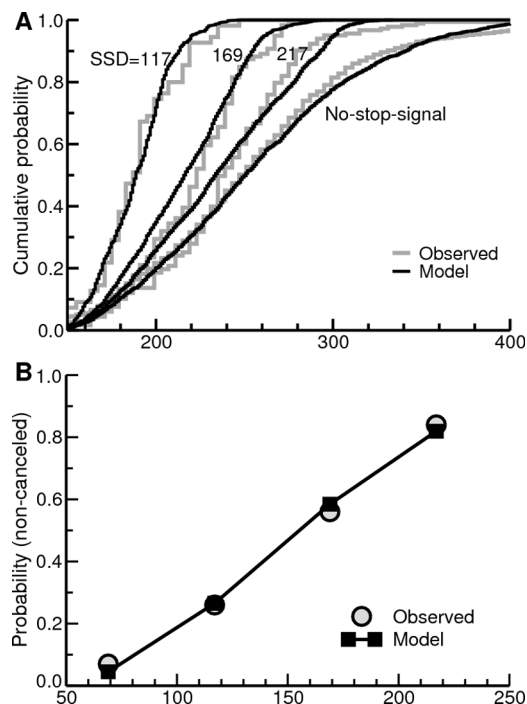
**Figure 2.** Simulated top-down modulation in the absence of sensory input. Spike rastergram (black dots) on top of the average firing rate (black curve) (200 trials) from a FIX neuron in a gap saccade task. Each row in the rastergram represents the spike train from one trial. In the task a subject has to hold saccade after the offset of the fixation signal until the onset of the target signal which appears after a delay (gap period). In the proposed model, although the visual stimulus (fixation signal) is off, the top-down control remains during the gap period until the onset of the target, resulting in a lower but persistent activity in fixation neurons during the gap period. The simulated activity resembles the observed activity of fixation neurons in the frontal eye field of a behaving monkey as shown in the inset (Hanes et al., 1998).

withhold a planned saccade during a temporal delay between the fixation point offset and the target onset (Dorris et al., 1997; Hanes et al., 1998). In our model, persistent top-down modulation is able to support FIX neurons to fire at a medium rate without the sensory stimulation during the temporal gap (Fig. 2), similar to the observed activity pattern of fixation neurons in the frontal eye field of a behaving monkey in the gap saccade task (Fig. 2, inset).

### Behavioral performance

The proposed model can quantitatively fit to the observed inhibition function and the saccade reaction times of no-stop-signal trials and noncanceled stop-signal trials. The inhibition function is defined as the probability of noncanceled response in stop-signal trials as a function of SSD. With most of the model parameters constrained by neurophysiological observations, the fitting to the behavioral data was performed using only three parameters (see Materials and Methods). We fit the model to the cumulative distribution of saccade reaction times of the no-stop-signal trials from a rhesus monkey [same data as in the study by Boucher et al. (2007)] by tuning the mean and SD of the holding period. We then fit the model to the inhibition function and saccade reaction times of noncanceled stop-signal trials by tuning the neural latency of the stop signal input (the time interval between the onset of the stop signal and its actual input to the network). The proposed spiking neural circuit model quantitatively accounts for the behavioral data obtained in this task (Fig. 3*A, B*) as well as the interactive and independent race models [ $\chi^2 = 64.92$ , compared with  $\chi^2 = 50.64$  for the interactive race model and  $\chi^2 = 57.24$  for the independent race model (Boucher et al. (2007))].

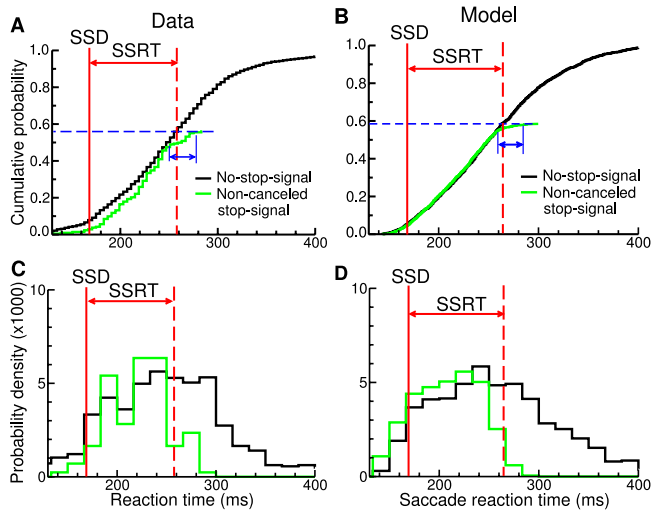
In our model, we can use the behavioral data to estimate the stop-signal reaction time (SSRT), a hidden variable that measures how long it takes for a STOP process to inhibit a go response after the presentation of a stop signal. There are several ways to measure SSRT and here we used the integration method formulated by Logan and Cowan (1984). The method is based on the assumption that after the presentation of the stop signal, the STOP



**Figure 3.** The model quantitatively fits to the behavioral data from a rhesus monkey (Boucher et al., 2007). **A**, Cumulative reaction time distributions of no-stop-signal trials (right most) and noncanceled stop-signal trials for different stop signal delays (from left to right: SSD = 117, 169, and 217 ms). **B**, Inhibition function shows the probability of noncanceled (signal response) in stop-signal trials as a function of SSD.

process takes time to finish and does not affect the GO process before the finish time. As a result, movements are only canceled if their reaction times are longer than the STOP process finish time, which is SSD + SSRT, on that trial. Furthermore, because of the independence between the two processes, the reaction times in the noncanceled stop-signal trials match those in no-stop-signal trials that are shorter than the STOP process finish time. Hence, one can estimate SSRT using the reaction time distribution of no-stop-signal trials and the probability of noncanceled response in stop-signal trials (Fig. 4). Although in our model MOV neurons (GO process) interact with the FIX neurons (STOP process) during an entire trial, the dynamics is the same in no-stop-signal trials and in stop-signal trials before the stop signal onset. Moreover, we assumed that the response of fixation neurons to a stop signal is powerful and immediately suppresses the activity of movement neurons. Under these conditions, we found that SSRT calculated using the described method remains valid (Fig. 4) (see Discussion).

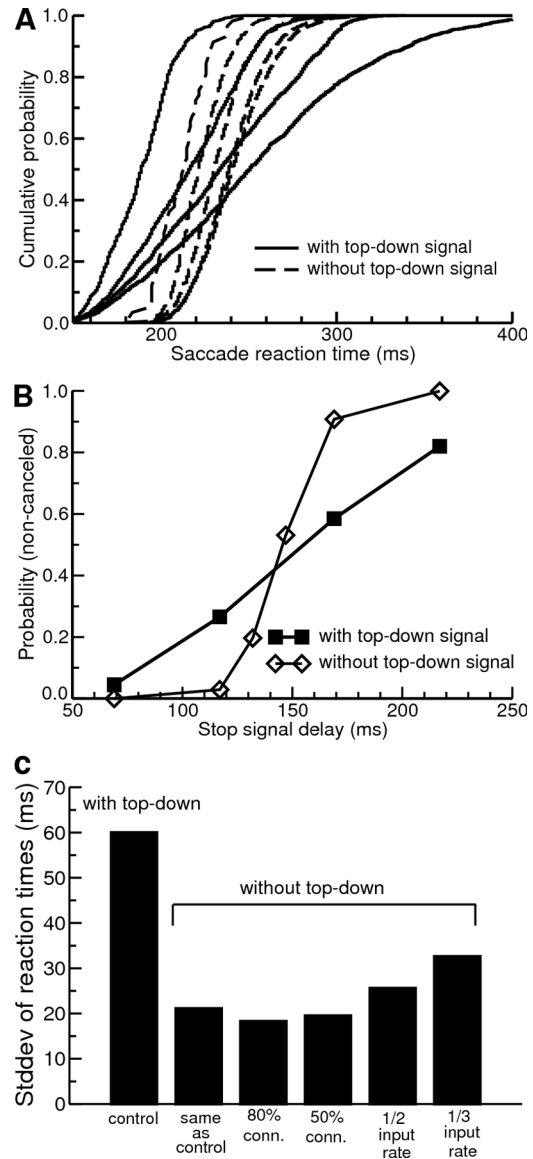
How important are our model assumptions about the top-down modulation and its variable holding period? To address this question, we tested our model by immediately turning off the top-down control module after the onset of the target, instead of maintaining the activity of the top-down control module for a variable period of time (the holding period). Without the holding period, the activity of MOV neurons increased too fast, resulting in very short mean reaction times. This can be counteracted with a reduction of the strength of the go signal input to the MOV neurons, so that the mean reaction time is the same as in the control case and in the monkey experiment. In this case, without the top-down modulation after the target onset, the model exhibits a much narrower distribution of reaction times and a much steeper inhibition function compared with the original model



**Figure 4.** Reaction time distributions from a monkey experiment and neural circuit model. **A, B**, To demonstrate how saccade reaction times are matched between no-stop-signal trials and noncanceled stop-signal trials, we plotted the cumulative reaction time distributions for no-stop-signal trials (black curves) and for noncanceled stop-signal trials (SSD = 169 ms, green curves) for data (**A**) and model (**B**). We also plotted corresponding probability density functions in **C** and **D**, respectively. For easy comparison between the two types of trials, the curves of noncanceled stop-signal trials are rescaled vertically by the probability of noncanceled (blue dashed lines, 56% for the data and 58% for the model). The result shows that for both the model and the data, the reaction times of noncanceled stop-signal trials approximately correspond to those of no-stop-signal trials until the time point that defines SSRT, estimated by the intersections between the blue dashed lines and the back curves (SSRT = 90 ms for the data and 95 ms for the model). Blue arrows indicate the ranges of STOP process variability.

(Fig. 5*A, B*). The result shows that without stochastic variability of the internal top-down control, the intrinsic dynamics of the premovement module cannot produce the observed magnitude of behavioral variability.

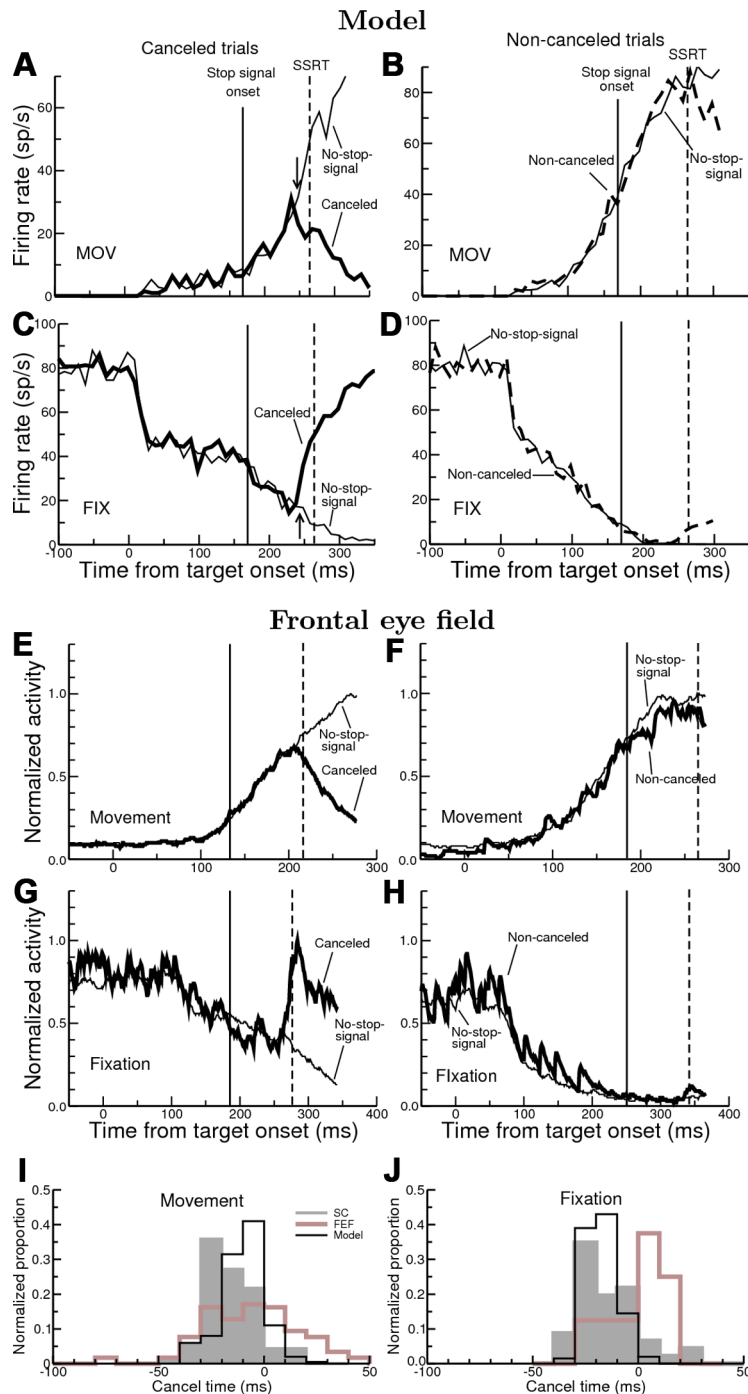
One may argue that even without the top-down module, the model could produce enough reaction time variability if other model parameters are properly tuned. In the proposed model, the reaction time variability results from the stochastic ramping activity of MOV neurons, which is affected by their recurrent excitation. Furthermore, without the top-down control, the main source of the MOV stochasticity is contributed by the intrinsic noise in the Poisson spike input representing the go signal. Therefore, we tested the model by decreasing the connectivity (probability of random connection) of the recurrent excitation of MOV populations as well as by decreasing the spike rate of the target input to MOV neurons in the absence of the top-down control (Fig. 5*C*). The result shows that changing the spike rate of the go signal input can slightly affect the reaction time variability (fifth and sixth bars from the left), but is still far less than that of the model with the top-down modulation (first bar from the left). Although we can use an extremely low go signal input rate to achieve desired reaction time variability, this low input rate is not biologically realistic. Note that to compare the SD of reaction times between different model settings, the mean reaction times must correspond. To this end, we need to keep the total synaptic input to MOV neurons the same when changing the parameters. In the case of decreasing recurrent connectivity, we increase the recurrent synaptic efficacy accordingly, whereas in the case of decreasing go signal input rate, we increase the input synaptic efficacy accordingly.



**Figure 5.** Model behavior in the absence of the top-down modulation. To test the contribution of the top-down modulation in the behavioral performance of the model, we turned off the top-down signal immediately after the target onset. **A, B**, The model exhibits a much narrower distribution of reaction times (**A**) and a much steeper inhibition function (**B**). **C**, We further test, by adjusting model parameters, whether the model without the top-down signal (second bar from the left) can produce enough reaction time variability exhibited by the one with the top-down signal (first bar from the left). To this end, we changed the connectivity of the recurrent excitation of MOV populations from 100% to 80% and 50% (third and fourth bars from the left, respectively). We also changed the spike rate of the go signal input to 1/2 and 1/3 of the original values (fifth and sixth bars from the left, respectively).

### Neuronal activity

The model goes beyond reproducing behavioral observations by exhibiting patterns of neural activation that resemble the activity of movement and fixation neurons in frontal eye field (Hanes et al., 1998) and superior colliculus (Paré and Hanes, 2003). We compared the firing activity in canceled and noncanceled stop-signal trials with latency-matched no-stop-signal trials (Fig. 6*A–D*). No-stop-signal trials with latencies matched to noncanceled stop-signal trials are a subset of no-stop-signal trials with latencies (saccade reaction times) shorter than SSRT + SSD (see Materials and Methods) (Fig. 4), whereas the rest of no-stop-signal trials have latencies longer than SSRT + SSD and are matched to



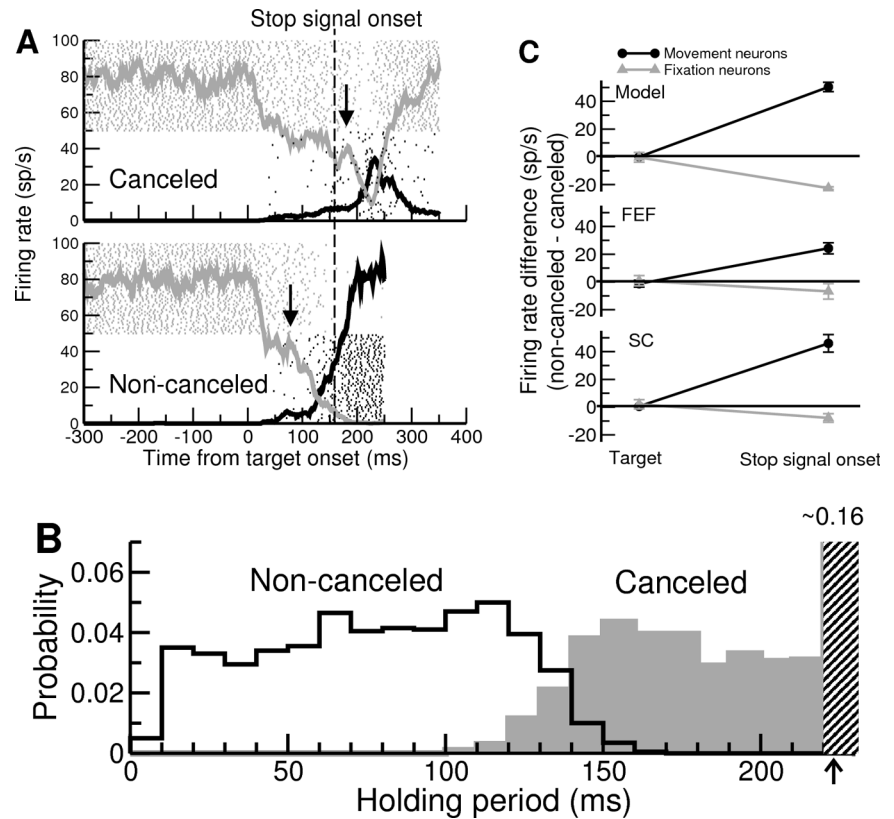
**Figure 6.** Comparison of mean firing rates between stop-signal trials (SSD = 169 ms) and latency-matched no-stop-signal trials averaged over 200 trials. **A, B**, Mean firing rates of stop-signal (canceled: thick solid curve; noncanceled: thick dashed curve) and latency-matched no-stop-signal trials (thin curves) from representative MOV neurons in the model for canceled and noncanceled trials, respectively. **C, D**, Same as **A, B**, respectively, for FIX neurons. Solid vertical lines indicate the onset of the stop signal and dashed vertical lines indicate the estimated stop signal response time (SSRT). Arrows indicate the cancel times. **E–H**, same as **A–D**, respectively, for movement and fixation neurons observed in frontal eye field of a rhesus monkey [modified from the study by Boucher et al. (2007), their Fig. 4]. Note that we did not model how the fixation neurons are reset after they cancel a movement because this is not part of the countermanding action. Hence simulated FIX activity does not go down in **C** as data shown in **G** after the movement is canceled. **I, J**, Distributions of cancel times of movement neurons for the model, frontal eye field (FEF) and superior colliculus (SC). **J**, Same as **I** for fixation neurons.

latencies in canceled stop-signal trials. The neural activity of MOV and FIX neurons in the model resembles the observed neural activity of movement and fixation neurons, respectively, in frontal eye field of a rhesus monkey (Fig. 6E–H). The model

shows that FIX neurons maintain a tonic activity during the fixation period and gradually ramp down after the offset of the fixation signal. After the onset of the stop signal, activity of FIX neurons ramps up drastically and shuts down movement neurons in canceled trials. The mean activity of stop-signal trials and latency-matched no-stop-signal trials are identical for both FIX and MOV neurons in the early part of the trial because the stimuli are identical until the onset of stop signal in stop-signal trials. The time point when the activity in stop-signal trials and latency-matched no-stop-signal trials becomes different marks the moment when the neural circuit starts to respond to the stop signal; this is referred to as the cancel time and is measured relative to SSRT (mean SSRT = 93.3 ms for data and 95.7 ms for the model). To replicate the sampling size in the physiology experiments (Hanes et al., 1998; Paré and Hanes, 2003), we randomly selected 50 MOV neurons and 50 FIX neurons and calculated the mean cancel time across 15–35 trials for each selected neuron. This resulted in distributions of cancel times with a mean of  $-9.2$  ms for MOV neurons and a mean of  $-17.0$  ms for FIX neurons (Fig. 6I, J). The results are comparable to observations in the frontal eye field and superior colliculus (Hanes et al., 1998; Paré and Hanes, 2003). It appears that the model produces cancel times that most closely approximate those measured in superior colliculus, especially the fixation neurons. We believe the model is describing events transpiring within the entire premotor network including superior colliculus and frontal eye field as well as certain basal ganglia, thalamus, and brainstem neurons. Also, we believe the quantitative differences in the distributions of cancel times between the frontal eye field data and the superior colliculus data are incidental effects of sampling and the properties of frontal eye field and superior colliculus neurons. Superior colliculus movement-related neurons tend to have lower baseline activity and higher peak discharge rates than counterparts in frontal eye field. This makes measurements of modulation times more reliable. Also, in the frontal eye field data, not all neurons provided enough data in all of the trial types needed to measure cancel time. Finally, a recent reexamination of the modulation of neural activity recorded in the

frontal eye field during the saccade countermanding task indicates that the movement-related neurons with stronger visual responses do not exhibit the systematic modulation of magnitude of activity that is characteristic of the GO units (Ray et al., 2008).

We further investigated how the neural activity correlates with the behavioral outcome. In the model, an early offset of the top-down control (shorter holding period) results in a faster ramping down activity of FIX neurons, allowing a faster ramping up in the activity of MOV neurons and increasing the probability of triggering a saccade in stop-signal trials (Fig. 7A). This observation is further verified by examining the relationship between the behavioral outcome and the holding period (Fig. 7B): We found that noncanceled trials are associated with short holding periods, whereas canceled trials are associated with long holding periods. Interestingly, if the top-down control is turned off between 100 ms and 160 ms after the target onset, the behavioral outcome is probabilistic; i.e., with a fixed value of holding period, the motor response is stochastically canceled on some trials, but not on the other trials. Although the holding period is not directly measurable in experiments, it affects the ramping-down activity of FIX neurons. Therefore, if we measure the neural activity between target onset and stop signal onset, the model predicts that the firing rate of FIX neurons will be higher in canceled trials than in noncanceled trials, whereas the firing rate of MOV neurons will be lower in canceled trials than in noncanceled trials. The prediction of differential neural activity of fixation neurons has never been tested before and we tested it by comparing the model simulations with observations in the frontal eye field and superior colliculus from primate countermanding tasks (Hanes et al., 1998; Pare and Hanes, 2003). Specifically, we calculated the difference between mean firing rates in canceled and noncanceled trials in 50 ms time windows around the target onset and the stop signal onset for each neuron. To maximize the number of canceled and noncanceled trials, we analyzed stop-signal trials with an intermediate stop signal delay for both frontal eye field data (mean SSD = 161 ms from 16 movement neurons and 164 ms from 5 fixation neurons) and superior colliculus (mean SSD = 157 ms from 26 movement neurons and 158 ms from 10 fixation neurons). To be consistent with the experimental setting, we performed model simulation for stop-signal trials with SSD = 160 ms and analyzed data from 15 MOV and 15 FIX neurons. The group mean difference between the canceled and noncanceled trials during the 50 ms epoch around the target onset is not significant for the model (paired Student's  $t$  test MOV:  $t_{(14)} = 0.0$ ,  $p = 0.5$ ; FIX:  $t_{(14)} = -0.10$ ,  $p = 0.46$ ) as well as for the frontal eye field (movement:  $t_{(15)} = -0.23$ ,  $p = 0.41$ ; fixation:  $t_{(4)} = -1.66$ ,  $p = 0.086$ ) and superior colliculus (movement:  $t_{(25)} = 1.37$ ,  $p = 0.09$ ; fixation:  $t_{(9)} = 0.61$ ,  $p = 0.28$ ), whereas the difference is significant during the 50 ms epoch starting at the stop signal onset for all data (MOV:  $t_{(14)} = 14.9$ ,  $p < 0.00001$ ; FIX:  $t_{(14)} = -24.1$ ,  $p < 0.00001$  for model; movement:  $t_{(15)} = 3.31$ ,  $p = 0.002$ ; fixation:



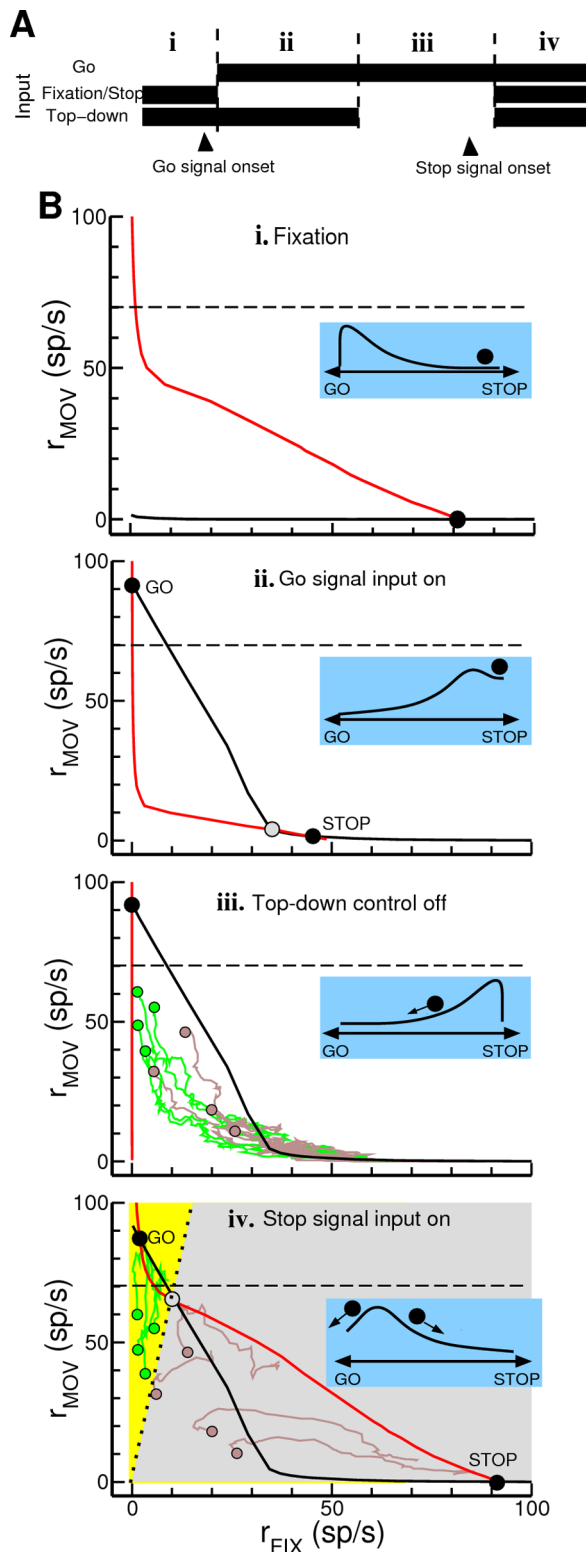
**Figure 7.** Difference in the neural activity affecting behavioral outcome. **A**, Simulated population firing rates of MOV (black lines) and FIX (gray lines) neurons on top of corresponding spike rasters (color matched dots) from two trials of model simulation (top, a canceled trial; bottom, a noncanceled trial). Each row in the rastergrams represents the spike train from one neuron. The vertical dashed line indicates the onset of the stop signal (SSD = 160 ms). Arrows indicate the offset of the top-down control. The model suggests that a quicker offset of the top-down control results in a faster ramping-down activity of fixation neurons and a faster ramping-up activity of movement neurons, thus increases the probability of making a saccade to the target. **B**, Probability distributions of holding time periods in cancel and noncancel trials for SSD = 160 ms. In ~16% trials the top-down control was not turned off before the stop signal onset (arrow). These trials are displayed in the shaded region. **C**, Mean firing rate difference (noncanceled trials – canceled trials) for movement and fixation neurons in a 50 ms epochs around the target onset and a 50 ms epochs starting at the stop signal onset. The model simulations, data from frontal eye field (FEF) and superior colliculus (SC) all show similar trends: at the time around the target onset there is no activity difference between canceled and noncanceled trials. At the time of the stop signal onset, fixation neurons have higher activity in canceled trials than in noncanceled trials, whereas movement neurons show the opposite.

$t_{(4)} = -2.12$ ,  $p = 0.05$  for frontal eye field and movement;  $t_{(25)} = 7.3$ ,  $p = 0$ , fixation:  $t_{(9)} = -2.53$ ,  $p = 0.016$  for superior colliculus) (Fig. 7C). We note that because of the 62 ms stop signal latency, the assessed activity in the 50 ms window starting at the onset of stop signal corresponds to that immediately before the arrival of the stop signal to the measured neurons.

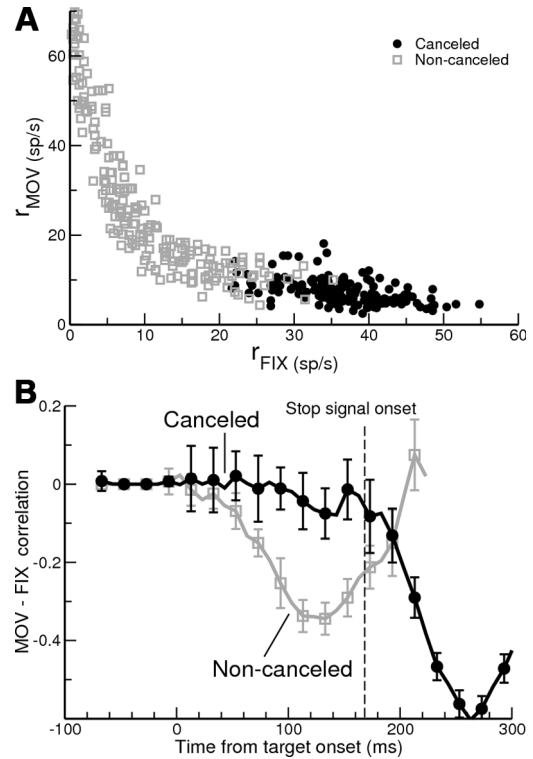
#### Attractor dynamics

The proposed model suggests that the probability of canceling a saccade is affected by the interaction between MOV and FIX neurons before the onset of the stop signal. To analyze the time evolution of action selection in stop-signal trials, we plotted the population firing rate of MOV neurons ( $r_{\text{MOV}}$ ) against the population firing rate of FIX neurons ( $r_{\text{FIX}}$ ) on a phase plane (Fig. 8). We show the network activity in four epochs of the task: during the fixation period (Fig. 8*Ai*), after the start of the go signal input and before the offset of the top-down control input (*Aii*), after the top-down control input offset and before the start of the stop signal input (*Aiii*), and after the start of the stop signal input (*Aiv*). To help analyze the dynamics of the network, we plotted nullclines for MOV neurons and FIX neurons (see Materials and





**Figure 8.** Phase plane plots demonstrate the attractor dynamics of the model network during different epochs of the task. **A**, Four epochs (i–iv) in a stop-signal trial defined by time windows separated by three events: start of the go signal input, offset of the control module input and start of the stop signal input (SSD = 169 ms). **B**, Phase plane plots (the population firing rate of MOV neurons,  $r_{MOV}$ , vs the population firing rate of FIX neurons,  $r_{FIX}$ ) for the four epochs of task. Black and red curves represent the nullclines for the MOV and FIX neurons, respectively. A nullcline is defined by the steady states of  $r_{MOV}$  (or  $r_{FIX}$ ) as a function of  $r_{FIX}$  (or  $r_{MOV}$ ) (see Materials and Methods). The intersections between the nullclines determine the equilibrium points (black circles: stable; gray circles: unstable) of the network. The brown curves (canceled trials) and green curves (noncanceled trials) depict the trajectories of eight trials in



**Figure 9.** Movement and fixation neurons develop negative correlation between their firing activity with a faster time course in noncanceled trials than in canceled trials. **A**, Snapshots of population firing rate of MOV ( $r_{MOV}$ ) versus that of FIX ( $r_{FIX}$ ) neurons at the onset of the stop signal (SSD = 169 ms) from ~400 trials show marked difference between canceled trials (black circles) and noncanceled trials (gray squares). Each mark represents one trial. The activity of the network can be well differentiated between the canceled and noncanceled trials at this time point already. **B**, The coefficient of spike count correlations between 10 pairs of randomly chosen MOV and FIX neurons as a function of time. Black and gray curves represent the average across the 10 pairs of neurons for canceled and noncanceled trials, respectively. Error bars indicate the SD. The noncanceled trials are associated with earlier onset of interaction between the MOV and FIX neurons, whereas the interaction in canceled trials starts later. The vertical dashed line marks the onset of the stop signal.

Methods) (Wilson, 1999). During the fixation period, a strong input to FIX neurons and no input to MOV neurons result in a stable equilibrium point at a large  $r_{FIX}$  and a small  $r_{MOV}$  (Fig. 8*Bi*). After the start of the go signal input, because of the strong input to MOV neurons and the reduced input to FIX neurons, the MOV neuron nullcline shifts up and the FIX neuron nullcline shifts down slightly, resulting in two stable (black circles) and one unstable (gray circles) equilibrium point (Fig. 8*Bii*). The stable equilibrium point at the upper left corner of the phase plane with a large  $r_{MOV}$  and small  $r_{FIX}$  is defined as the GO attractor and the stable equilibrium point at the lower right corner with a small  $r_{MOV}$  and large  $r_{FIX}$  is defined as the STOP attractor. Since the STOP attractor is the one closest to the original stable equilibrium point during the fixation period, the system simply moves into the STOP attractor after the start of the go signal input and maintains fixation. After the top-down control from the control

←  
the corresponding epochs. Depending on the state of the system at the moment when the stop signal input starts (indicated by the color matched circles), the network may continue converging into the GO attractor and trigger a response, or may turn back into the STOP attractor and cancels the response. Dashed lines mark the threshold for the saccade response. Insets, Schematic plots illustrating how the one-dimensional “action landscape” changes with stimulus inputs.

module switches off, the STOP attractor disappears and the system starts to converge toward the GO attractor, resulting in decreasing  $r_{\text{FIX}}$  and increasing  $r_{\text{MOV}}$  (Fig. 8*Biii*). When the stop signal input is turned on, the STOP attractor and the unstable equilibrium points reemerge (Fig. 8*Biv*). The unstable manifold (dotted line) that passes through the unstable equilibrium point divides the phase plane into two basins of attraction. Depending on the basin the system state is in (to the right or left of the unstable manifold) at the moment when the stop signal input starts, the system may continue moving toward the GO attractor and initiate a saccade, or may turn back toward the STOP attractor and maintain the fixation.

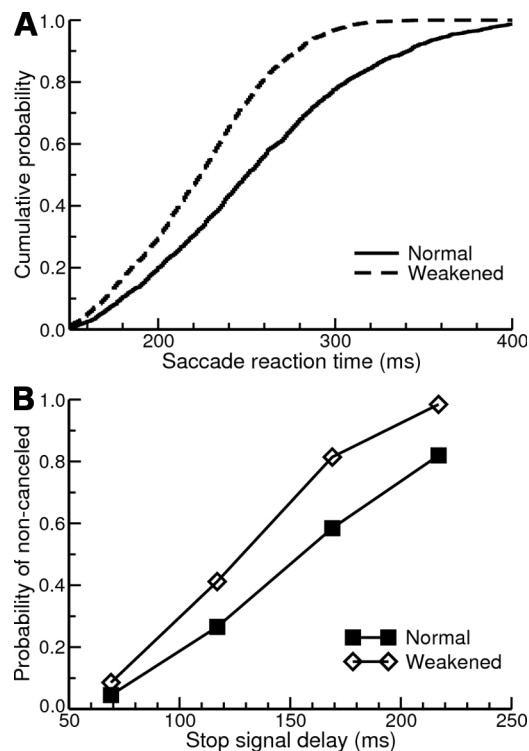
The state of the system at the moment of the start of the stop signal input is determined by the history of the neural interaction of the system before that moment as well as by the proactive top-down control. If the top-down control is switched off early in a trial, the system has more chance to move into the region to the left of the unstable manifold when the stop signal is turned on (Fig. 8*Biii, iv*, green trajectories). In contrast, if the top-down control remains for a longer time, the system is likely to be still in the region to the right of the unstable manifold when the stop signal is turned on (Fig. 8*Biii, iv*, brown trajectories).

#### Neural correlate of top-down bias

The mutual inhibition between the FIX and MOV neurons implies a strong negative correlation between the two populations even before the stop signal appears. Figure 9*A* shows snapshots of the state of the system ( $r_{\text{MOV}}$  versus  $r_{\text{FIX}}$ ) at the onset of the stop signal (SSD = 169 ms) from ~400 trials. Trials in which the activity of MOV neurons and FIX neurons start to develop negative correlation ( $r_{\text{MOV}}$  increasing versus  $r_{\text{FIX}}$  decreasing) early tend to become noncanceled trials (gray: noncanceled trials, black: canceled trials). We further investigated the time evolution of the correlation between the activity of MOV and FIX neurons. We calculated the coefficient of spike-count correlation, following the procedure stated in the study by Zohary et al. (1994), from 500 trials for 10 randomly chosen pairs of individual MOV and FIX neurons in a sliding window of 20 ms. We then averaged the coefficients across the 10 pairs of neurons and plotted the average coefficient as a function of time (Fig. 9*B*). The result shows that the spike activity of MOV and FIX neurons become negatively correlated earlier in noncanceled trials than in canceled trials. Note that an observed negative correlation could either reflect negatively correlated external inputs to the two populations, or mutually inhibitory interactions between them, or a combination of both. In the countermanding task, the stimuli in the canceled and noncanceled trials are identical, so the difference in the spike count correlation reflects the mutually inhibitory interactions between MOV and FIX neurons which are modulated by the endogenous top-down control. The top-down control holds longer in canceled trials, causing very little MOV neuron activity that remains uncorrelated with FIX neurons, hence results in a smaller spike-count correlation in canceled versus noncanceled trials in the early part of a trial.

#### Effect of weakened top-down control

The persistent top-down inhibitory control is a unique feature in our model. It is interesting to test how a weakened top-down control at the neuronal level affects the behavioral outcome of the countermanding task. To this end, we tested our model by reducing the input to neurons in the control module, which resulted in a reduced input to the FIX neurons. This change produced ~16% and 8% decreases in the firing rates of control and FIX neurons,



**Figure 10.** Difference in the behavior between weakened top-down control and normal control in the model. *A, B*, By reducing the activity of neurons in the control module, we show that the simulated weak top-down control causes faster responses as indicated by cumulative reaction time distributions of no-stop-signal trials (*A*) and a reduced performance as indicated by the inhibition functions (*B*).

respectively. Our simulation results show that the mean reaction time for no-stop-signal trials became shorter and the probability of noncanceled responses increased for stop-signal trials (Fig. 10). The result demonstrated that the baseline activity of fixation neurons plays a critical role in inhibitory control. A slight decrease in the neural activity of fixation neurons results in a substantial impact to the behavioral performance in the countermanding task. Interestingly, the deficit of the inhibitory control is not significantly reflected in the average SSRT, which is 95.7 ms for the normal condition and 95.2 ms for the weakened inhibitory control.

#### Discussion

We demonstrate a recurrent spiking neural network model of inhibitory control for the saccade countermanding task. The original race model was framed only by the finish times of GO and STOP processes (Logan and Cowan, 1984). The recent interactive race model showed formally how a STOP unit can interrupt a GO unit (Boucher et al., 2007). Neither model accounted for the ongoing balance of activity of movement and fixation neurons, which reveals new insights into how performance is affected by the state of the network before the stop signal appears. The model emphasizes the interaction between fixation and movement neurons recorded in the frontal eye field and superior colliculus, but it may also describe the inhibitory control that substantia nigra pars reticulata in basal ganglia exercises on superior colliculus (Wickens, 1997; Hikosaka et al., 2000; Frank et al., 2004; Aron and Poldrack, 2006; Frank and Claus, 2006; Aron et al., 2007; Eagle et al., 2008; Isoda and Hikosaka, 2008).

Crucially, the model emphasizes the proactive, top-down control driving the activity of fixation neurons and their interac-

tion with movement neurons before the stop signal onset. This top-down signal probably originates in dorsolateral prefrontal cortex, supplementary eye field, and anterior cingulate cortex where neurons signal error, reward, and conflict (Stuphorn et al., 2000; Ito et al., 2003; Stuphorn and Schall, 2006; Isoda and Hikosaka, 2007; Johnston et al., 2007). In fact, intracortical microstimulation of the supplementary eye field modulates performance of the saccade countermanding task (Stuphorn and Schall, 2006). One may argue that, instead of applying the top-down control on fixation neurons, the same effect can be achieved by applying a “reversed” top-down control on movement neurons: the top-down control sends an excitatory input to movement neurons after the target onset and then removes the input after the stop-signal onset. However, according to this scenario, the fixation neurons would be turned off immediately after the fixation signal offset and there would be no persistent activity in fixation neurons in the gap saccade task, which is inconsistent with the experimental observations.

The current model resembles the neural network model for two-alternative choices (Wang, 2002; Wong and Wang, 2006). However, the difference in the input time course results in fundamentally different dynamics in the two models. The decision-making model network starts from a neutral state and the relative strength of the two simultaneous inputs determines which attractor the network will converge into. The motor control model network starts from one of the two attractors (STOP), and then is driven by the target input into the other attractor (GO), whereas the time course of the top-down signal stochastically determines whether this transition will be canceled when the stop signal appears.

The model replicates the probability of responding and reaction times on correct and error trials and also the patterns of neuronal activity observed in frontal eye field and superior colliculus (Hanes et al., 1998; Paré and Hanes, 2003). We found that, before the stop signal, the firing rate of fixation and movement neurons, respectively, decreases and increases faster in noncanceled compared with canceled stop-signal trials. The model made a new prediction: the activity of movement and fixation neurons becomes negatively correlated earlier in noncanceled trials than in canceled trials. We demonstrated that adjusting the strength of the top-down drive provides an effective mechanism for modulating the amount of inhibitory control. The model offers a novel view of inhibitory control based on nonlinear attractor dynamics. Whether a response is triggered or canceled is mainly determined by bistable dynamics and is relatively insensitive to the precise value of the saccade threshold. However, saccade initiation time is given by the time taken for the stochastic rise of activity to reach that threshold. In other words, the choice (go or stop) can be specified within the circuit before the behavioral response is overtly triggered. This model may be applicable to inhibitory control of motor movements in other neural systems and tasks (Johnston and Everling, 2006; Narayanan and Laubach, 2006; Isoda and Hikosaka, 2007).

The attractor network reproduces two key observations that have been interpreted as evidence for the independence of the GO and STOP process finish times. First, the reaction times of noncanceled stop-signal trials do not exceed  $SSD + SSRT$  (Fig. 4). Second, the neural activity leading to noncanceled movements in stop-signal trials corresponds to that leading to movements in trials with no stop signal (Fig. 6*B, D*). The model reproduces these observations although the fixation and movement units interact synaptically throughout a trial because the fixation units potently inhibit movement units only  $\sim 80$  ms after the stop

signal. Before this instant, the activity of movement and fixation neurons is identical between no-stop-signal and stop-signal trials, giving the appearance of independence.

The model assumes that top-down control is exerted a variable duration after fixation signal offset, consistent with the activity patterns of candidate neurons possibly carrying this signal in dorsolateral prefrontal cortex (Suzuki and Azuma, 1977; Mikami et al., 1982; Tinsley and Everling, 2002) and dorsomedial prefrontal cortex (Schlag et al., 1992; Bon and Lucchetti, 1992) as well as of fixation neurons in superior colliculus and frontal eye field (Munoz and Wurtz, 1993; Dorris et al., 1997; Hanes et al., 1998). The duration of this control of the activity of fixation neurons can explain why reaction times in the stop-signal task are longer than typical visually guided saccades (Hanisch et al., 2006; Stuphorn and Schall, 2006; Verbruggen et al., 2006) and may be a means of adapting to trial history (Emeric et al., 2007). Critically, the delay and variability of reaction time observed in saccade countermanding could not be produced by the model network without the top-down modulation of the fixation neurons, even with random connectivity of the network or with lower Poisson spike rate for the go signal input to the movement neurons. However, it remains to be seen, in future research, whether the required amount of variability can be realized with modifications of network architecture and other circuit and neuronal properties. The consequence of the top-down modulation is that the activity of fixation neurons before the stop signal occurs will predict the behavioral outcome; this was confirmed in the neurophysiological data (Fig. 7).

Furthermore, weaker simulated top-down control resulted in shorter mean reaction times, more erroneous saccades in stop-signal trials but unchanged SSRT. This result is consistent with performance of the saccade countermanding task by subjects diagnosed with ADHD (Armstrong and Munoz, 2003; Hanisch et al., 2006), but manual countermanding may be different (Schar and Logan, 1990). The lack of sensitivity of SSRT to the precise level of the top-down input signal occurs because the synaptic interactions between fixation and movement neurons are highly asymmetrical, predominated by inhibition of movement neurons by fixation neurons. A slightly reduced top-down control is still strong enough so that when the stop signal is turned on, the fixation neurons can immediately suppress movement neurons, resulting in an unchanged SSRT. It is conceivable that if the dynamic interaction between movement and fixation neurons is more symmetrical, then a change in the strength of the top-down signal would lead to a modification of behavioral SSRT.

The attractor network model can be tested and extended in several ways. First, it would be interesting to assess whether optimization of performance can be achieved by varying the initial activation level of movement and/or fixation neurons by adjusting the top-down control signal or by changing the duration of the holding period (Wilimzig et al., 2006; Wong et al., 2007). Second, it will be informative to separately manipulate the stimulus-driven input and the internal top-down input to fixation neurons and assess their effects on inhibitory control, for instance by using a peripheral visual stop signal, in which its sensory input goes into movement neurons rather than the fixation neurons (Asrress and Carpenter, 2001; Armstrong and Munoz, 2003). One can also investigate the effect of truly endogenous inhibitory control by removing the sensory stop signal and using a self-controlled stopping paradigm (Brass and Haggard, 2007). Third, the neural circuit of the control module endowed with persistent activity must be modeled. Finally, the model can

be extended to include the effect of trial history on the saccade reaction times (Dorris et al., 2000; Emeric et al., 2007) by implementing plastic synapses in the circuit of the control module that are updated in a manner that depends on the outcome of the current trial (Soltani and Wang, 2006; Fusi et al., 2007).

In conclusion, this framework for understanding inhibitory control in terms of attractor dynamics of a recurrent spiking neural network driven by a proactive top-down signal is notable in several respects. First, the nature of the spiking neural modeling enables us to address issues such as neural firing activity and synaptic mechanisms underlying inhibitory control. Second, the model indicates that inhibitory control of action includes not only a reactive inhibitory signal driven by the presentation of the stop signal but also a pro-active persistent inhibitory signal preceding the stop signal. Third, the recurrent network model is endowed with attractor dynamics. The attractor mechanism underlying competition between choice alternatives is similar to a model for perceptual decision making (Wang, 2002; Wong and Wang, 2006), except that here the decision is about whether to act or not. Fourth, the original formal race model and the proposed spiking network model describe inhibitory control at very different levels. Reproducing behavioral performance and firing properties of individual neurons at the same time, this model represents an important step in translating between the cognitive theory and the neurophysiology of countermanding action.

## References

- Armstrong IT, Munoz DP (2003) Inhibitory control of eye movements during oculomotor countermanding in adults with attention-deficit hyperactivity disorder. *Exp Brain Res* 152:444–452.
- Aron AR, Poldrack RA (2006) Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus. *J Neurosci* 26:2424–2433.
- Aron AR, Fletcher PC, Bullmore ET, Sahakian BJ, Robbins TW (2003) Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nat Neurosci* 6:115–116.
- Aron AR, Durston S, Eagle DM, Logan GD, Stinear CM, Stuphorn V (2007) Converging evidence for a fronto-basal-ganglia network for inhibitory control of action and cognition. *J Neurosci* 27:11860–11864.
- Assrress KN, Carpenter RH (2001) Saccadic countermanding: a comparison of central and peripheral stop signals. *Vision Res* 41:2645–2651.
- Bon L, Lucchetti C (1992) The dorsomedial frontal cortex of the macaque monkey: fixation and saccade-related activity. *Exp Brain Res* 89:571–580.
- Boucher L, Palmeri TJ, Logan GD, Schall JD (2007) Inhibitory control in mind and brain: an interactive race model of countermanding saccades. *Psychol Rev* 114:376–397.
- Boudet C, Stiemerling N, Delva N, Paré M (2005) Countermanding saccades in patients with schizophrenia. *Soc Neurosci Abstr* 31:443.3.
- Brass M, Haggard P (2007) To do or not to do: the neural signature of self-control. *J Neurosci* 27:9141–9145.
- Donohoe G, Reilly R, Clarke S, Meredith S, Green B, Morris D, Gill M, Corvin A, Garavan H, Robertson IH (2006) Do antisaccade deficits in schizophrenia provide evidence of a specific inhibitory function? *J Int Neuropsychol Soc* 12:901–906.
- Dorris MC, Paré M, Munoz DP (1997) Neuronal activity in monkey superior colliculus related to the initiation of saccadic eye movements. *J Neurosci* 17:8566–8579.
- Dorris MC, Paré M, Munoz DP (2000) Immediate neural plasticity shapes motor performance. *J Neurosci* 20:52.
- Eagle DM, Baunez C, Hutcheson DM, Lehmann O, Shah AP, Robbins TW (2008) Stop-signal reaction-time task performance: role of prefrontal cortex and subthalamic nucleus. *Cereb Cortex* 18:178–188.
- Emeric EE, Brown JW, Boucher L, Carpenter RH, Hanes DP, Harris R, Logan GD, Mashru RN, Paré M, Pouget P, Stuphorn V, Taylor TL, Schall JD (2007) Influence of history on saccade countermanding performance in humans and macaque monkeys. *Vision Res* 47:35–49.
- Frank MJ, Claus ED (2006) Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychol Rev* 113:300–326.
- Frank MJ, Seeberger LC, O'Reilly RC (2004) By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science* 306:1940–1943.
- Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J Neurophysiol* 61:331–349.
- Funahashi S, Chafee MV, Goldman-Rakic PS (1993) Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade task. *Nature* 365:753–756.
- Fusi S, Asaad WF, Miller EK, Wang XJ (2007) A neural circuit model of flexible sensorimotor mapping: learning and forgetting on multiple time-scales. *Neuron* 54:319–333.
- Hanes DP, Schall JD (1996) Neural control of voluntary movement initiation. *Science* 274:427–430.
- Hanes DP, Patterson WF 2nd, Schall JD (1998) Role of frontal eye fields in countermanding saccades: visual, movement, and fixation activity. *J Neurophysiol* 79:817–834.
- Hanisch C, Radach R, Holtkamp K, Herpertz-Dahlmann B, Konrad K (2006) Oculomotor inhibition in children with and without attention-deficit hyperactivity disorder (ADHD). *J Neural Transm* 113:671–684.
- Hikosaka O, Takikawa Y, Kawagoe R (2000) Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiol Rev* 80:953–978.
- Isoda M, Hikosaka O (2007) Switching from automatic to controlled action by monkey medial frontal cortex. *Nat Neurosci* 10:240–248.
- Isoda M, Hikosaka O (2008) Role for subthalamic nucleus neurons in switching from automatic to controlled eye movement. *J Neurosci* 28:7209–7218.
- Ito S, Stuphorn V, Brown JW, Schall JD (2003) Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science* 302:120–122.
- Johnston K, Everling S (2006) Monkey dorsolateral prefrontal cortex sends task-selective signals directly to the superior colliculus. *J Neurosci* 26:12471–12478.
- Johnston K, Levin HM, Koval MJ, Everling S (2007) Top-down control-signal dynamics in anterior cingulate and prefrontal cortex neurons following task switching. *Neuron* 53:453–462.
- Konishi S, Nakajima K, Uchida I, Sekihara K, Miyashita Y (1998) No-go dominant brain activity in human inferior prefrontal cortex revealed by functional magnetic resonance imaging. *Eur J Neurosci* 10:1209–1213.
- Lo CC, Wang XJ (2006) Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nat Neurosci* 9:956–963.
- Lo CC, Wang XJ (2007) A spiking neural network model of control of action in the countermanding task. *Soc Neurosci Abstr* 33:636.10.
- Logan GD, Cowan WB (1984) On the ability to inhibit thought and action: a theory of an act of control. *Psychol Rev* 91:295–327.
- Mikami A, Ito S, Kubota K (1982) Visual response properties of dorsolateral prefrontal neurons during visual fixation task. *J Neurophysiol* 47:593–605.
- Munoz DP, Wurtz RH (1993) Fixation cells in monkey superior colliculus. 1. characteristics of cell discharge. *J Neurophysiol* 70:559–575.
- Naito E, Matsumura M (1994) Movement-related potentials associated with motor inhibition as determined by use of a stop signal paradigm in humans. *Brain Res Cogn Brain Res* 2:139–146.
- Narayanan NS, Laubach M (2006) Top-down control of motor cortex ensembles by dorsomedial prefrontal cortex. *Neuron* 52:921–931.
- Paré M, Hanes DP (2003) Controlled movement processing: superior colliculus activity associated with countermanded saccades. *J Neurosci* 23:6480–6489.
- Ray S, Pouget J, Schall JD (2008) Relation of visuomotor activity in macaque frontal eye field to response time during saccade countermanding. *Soc Neurosci Abstr* 34:855.14.
- Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S (2004) The role of the medial frontal cortex in cognitive control. *Science* 306:443–447.
- Schachar R, Logan GD (1990) Impulsivity and inhibitory control in normal development and childhood psychopathology. *Dev Psychol* 26:710–720.
- Schlag J, Schlag-Rey M, Pigarev I (1992) Supplementary eye field: influence of eye position on neural signals of fixation. *Exp Brain Res* 90:302–306.
- Soltani A, Wang XJ (2006) A biophysically based neural model of matching law behavior: melioration by stochastic synapses. *J Neurosci* 26:3731–3744.
- Stuphorn V, Schall JD (2006) Executive control of countermanding saccades by the supplementary eye field. *Nat Neurosci* 9:925–931.

- Stuphorn V, Taylor TL, Schall JD (2000) Performance monitoring by the supplementary eye field. *Nature* 408:857–860.
- Suzuki H, Azuma M (1977) Prefrontal neuronal activity during gazing at a light spot in the monkey. *Brain Res* 126:497–508.
- Tinsley CJ, Everling S (2002) Contribution of the primate prefrontal cortex to the gap effect. *Prog Brain Res* 140:61–72.
- Verbruggen F, Logan GD (2009a) Models of response inhibition in the stop signal and stop change paradigms. *Neurosci Biobehav Rev* 33:647–661.
- Verbruggen F, Logan GD (2009b) Proactive adjustments of response strategies in the stop-signal paradigm. *J Exp Psychol Hum Percept Perform* 35:835–854.
- Verbruggen F, Liefoghe B, Vandierendonck A (2006) The effect of interference in the early processing stages on response inhibition in the stop signal task. *Q J Exp Psychol (Colchester)* 59:190–203.
- Wang XJ (2001) Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci* 24:455–463.
- Wang XJ (2002) Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* 36:955–968.
- Wickens J (1997) Basal ganglia: structure and computations. *Network-Comp Neural* 8:77–109.
- Wilimzig C, Schneider S, Schöner G (2006) The time course of saccadic decision making: dynamic field theory. *Neural Netw* 19:1059–1074.
- Wilson HR (1999) *Spikes, decisions and actions: the dynamical foundations of neuroscience*. Oxford: Oxford UP.
- Wong KF, Wang XJ (2006) A recurrent network mechanism of time integration in perceptual decisions. *J Neurosci* 26:1314–1328.
- Wong KF, Eckhoff P, Holmes P, Cohen JD (2007) Analysis of a biologically realistic model for saccade-countermanding tasks. *Soc Neurosci Abstr* 33:719.3.
- Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370:140–143.