LETTER Communicated by Mark van Rossum

Neural Information Processing with Feedback Modulations

Wenhao Zhang
whzhang@ion.ac.cn
Institute of Neuroscience, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China, and State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China

Si Wu
wusi@bnu.edu.cn
State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China, and Institute for Neuroscience, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

Descending feedback connections, together with ascending feedforward ones, are the indispensable parts of the sensory pathways in the central nervous system. This study investigates the potential roles of feedback interactions in neural information processing. We consider a two-layer continuous attractor neural network (CANN), in which neurons in the first layer receive feedback inputs from those in the second one. By utilizing the intrinsic property of a CANN, we use a projection method to reduce the dimensionality of the network dynamics significantly. The simplified dynamics allows us to elucidate the effects of feedback modulation analytically. We find that positive feedback enhances the stability of the network state, leading to an improved population decoding performance, whereas negative feedback increases the mobility of the network state, inducing spontaneously moving bumps. For strong, negative feedback interaction, the network response to a moving stimulus can lead the actual stimulus position, achieving an anticipative behavior. The biological implications of these findings are discussed. The simulation results agree well with our theoretical analysis.

1 Introduction

External information is processed layer by layer in the sensory pathways of the central nervous system. Anatomical data have revealed that apart from the ascending ones, there exist abundant descending connections between layers, whose number is even larger than that of the former (Sillito, Cudeiro, & Jones, 2006). For instance, about 30% of the synaptic inputs received by

 © 2012 Massachusetts Institute of Technology

lateral geniculate nucleus (LGN) neurons is from the feedback connections of layer 6 neurons in V1 and only 10% from the feedforward connections of retina ganglion cells (Van Horn, Erişir, & Sherman, 2000); for layer 4 neurons in V1, the feedback connections from layer 6 neurons comprise about 45% of the total synaptic inputs, whereas the feedforward connections from LGN neurons contribute only 6% to 9% (Latawiec, Martin, & Meskenaite, 2000). Thus, understanding the role of feedback interaction in neural information processing is critical for elucidating brain functions.

A general perspective on the role of feedback interaction is that it conveys the information in higher layers and modulates neural responses in descendant ones, so that the efficiency and accuracy of neural systems extracting external stimuli are improved (Lee & Mumford, 2003). This view has been supported by experimental findings. For instance, the feedback input from V1 can enhance the center surround antagonism of an LGN neuron, improving its response sharpness to salient features embedded in noisy backgrounds (Murphy & Sillito, 1987). The V1 feedback input can also switch the firing pattern of an LGN neuron between burst and tonic, controlling the features of the information flow relayed to the cortex (McClurkin, Optican, & Richmond, 1994). Furthermore, the feedback inputs from V1 to LGN, and from MT to V1, can enhance motion information processing. It was found that the feedback input from V1 strengthens the LGN response to moving stimuli, enabling it to track faster-moving objects (Gulyas, Lagae, Eysel, & Orban, 1990). The feedback input from MT to V1 conveys the motion information of objects in a wide visual field not available to V1 neurons and helps solve the aperture problem (Pack & Born, 2001). Theoretical studies, based on optimal inference, also suggest that feedback interaction serves to estimate the input information recursively, achieving efficient predictive coding (Rao & Ballard, 1999; Lee & Mumford, 2003).

Although the importance of feedback interaction has been widely recognized in the field, there are few modeling studies elucidating its detailed computational role. Our knowledge on the structure of feedback interactions is very limited, and a dynamical system with feedback connections is often extremely difficult to analyze. In this study, based on experimental data and a simple model, we explore the potential computational role of feedback interaction on modulating the network response properties to external inputs. In particular, we consider a two-layer network with neurons reciprocally connected between layers. This agrees with the experimental finding that neurons tend to feed back their activities to those from which they receive the ascending inputs (Lund, Angelucci, & Bressloff, 2003). Neurons in both layers are tuned to the same continuous stimulus, and they are connected recurrently in the same layer, forming a continuous attractor neural network (CANN). CANNs have been successfully applied to describe the encoding of continuous stimuli in neural systems, including orientation (Ben-Yishai, Bar-Or, & Sompolinsky, 1995), head direction (Zhang, 1996), moving direction (Georgopoulos, Taira, & Lukashin, 1993),

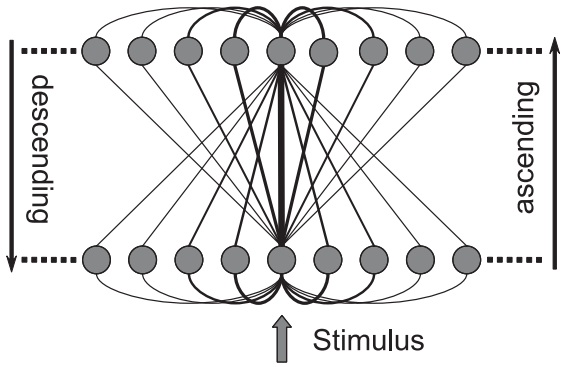

Figure 1: The structure of the network. It consists of two layers of neurons. Neurons are aligned according to their preferred stimuli and connected with each other in the same layer and between layers. Only two neurons' connection patterns are shown, with the line width indicating the connection strength.

and spatial location of objects (Samsonovich & McNaughton, 1997). If the preferred stimulus of neurons is associated with the orientation, our model can be regarded as a simplified network formed by layers 4 and 6 neurons in V1 or formed by neurons in V1 and MT (or V2) that are retinotopically connected. Our model, however, is not directly applicable to the interaction between LGN and V1 neurons, since no evidence shows that the LGN network can be described as a CANN.

Experimental data have revealed that feedback interaction can be positive or negative depending on the relationship between neurons' preferred stimuli, and the modulation can display either a push or a pull effect or both (Wang, Jones, Andolina, Salt, & Sillito, 2006). In this study, we consider both positive and negative feedback interactions and compare their effects on network dynamics. To overcome the high dimensionality of network dynamics, we use a projection method to simplify it significantly: we approximate the network dynamics by considering only its dominating motion models. With the simplified dynamics, we systematically investigate the modulations of feedback interaction on the network response properties and obtain a number of interesting results that may have important implications in neural information processing.

2 The Model

We consider a two-layer network with neurons reciprocally connected between layers (see Figure 1). Neurons in both layers are tuned to the same one-dimensional stimulus x. We consider the case that the range of all neurons' preferred stimulus is much larger than the range of neuronal

interactions (the neuronal interaction has a gaussian form, and the width of the gaussian function defines the range of neuronal interactions; see equation 2.4). We can thus effectively take $x \in (-\infty, \infty)$ in our analysis. In simulations, however, we set the stimulus range to be $-L/2 < x \leq L/2$ and impose a periodic condition.

Let $U_i(x,t)$ be the synaptic input at time t to the neurons in the ith layer whose preferred stimulus is x. The dynamics of $U_i(x,t)$ is determined by the recurrent input from other neurons in the same layer, the feedback (for the first layer) or feedforward (for the second layer) input from the other layer, the external input $I_{ext}(x,t)$ (only for the first layer), and its own decay. The dynamical equations for $U_i(x,t)$ are written as

$$\tau \frac{\partial U_1(x,t)}{\partial t} = -U_1(x,t) + \rho \int_{x'} W(x,x') r_1(x',t) dx' + \rho \int_{x'} W_{FB}(x,x') r_2(x',t) dx' + I_{ext}(x,t), \tag{2.1}$$

$$\tau \frac{\partial U_2(x,t)}{\partial t} = -U_2(x,t) + \rho \int_{x'} W(x,x') r_2(x',t) dx' + \rho \int_{x'} W_{FF}(x,x') r_1(x',t) dx', \tag{2.2}$$

where τ is the time constant for the synaptic input, which is typically on the order of 2 to 5 ms. ρ is the neural density. $r_i(x,t)$ is the firing rate of neurons in the ith layer having the preferred stimulus x. $r_i(x,t)$ increases with the synaptic input but saturates in the presence of global inhibition. A solvable model capturing these features is given by divisive normalization (Deneve, Latham, & Pouget, 1999; Wu, Amari, & Nakahara, 2002),

$$r_i(x,t) = \frac{\left[U_i(x,t)\right]_+^2}{1 + k\rho \int_{x'} U_i^2(x',t) dx'}, \tag{2.3}$$

where the symbol $[x]_+$ denotes a half-rectifying function: $[x]_+ = 0$, for $x \leq 0$ and $[x]_+ = x$, for $x > 0$, and k the global inhibition strength.

It has been suggested that divisive normalization can be achieved by a shunting inhibition (Heeger, 1992). Here we sketch a possible way to achieve this. In a recent experimental study, Hao, Wang, Dan, Poo, and Zhang (2009) showed that if an inhibitory input is on the path for an excitatory current propagating to the soma of a neuron (called the on-path configuration), then the total synaptic current generated at the soma can be written as $U_i = U_i^E + U_i^I + kU_i^E U_i^I$, where $U_i^E > 0$ and $U_i^I < 0$ denote, respectively, the excitatory and inhibitory currents, and k is a constant. Here, the multiplicative term, $kU_i^E U_i^I$, represents the contribution of shunting inhibition. For simplicity,

let us assume that due to noise, the firing rate of a neuron has a linear relationship with its synaptic input, $r_i \sim U_i$. Consider a network structure in which all excitatory neurons are connected to a group of inhibitory neurons, and the inhibitory neurons send feedback inputs to the excitatory ones. Then the magnitude of the inhibitory input received by an excitatory neuron is proportional to the total activities of all excitatory neurons: $U_i^I \sim -\sum_j r_j^E$. Furthermore, consider that the inhibitory input to an excitatory neuron is on path with respect to all excitatory inputs from other neurons (called global shunting); then the total synaptic input received by this excitatory neuron is $U_i \sim U_i^E - \sum_j r_j^E - kU_i^E \sum_j r_j^E$, and the firing rate of the neuron is $r_i^E \sim U_i^E - \sum_j U_j^E - kU_i^E \sum_j U_j^E$. Therefore, in this network, the firing rate of an excitatory neuron increases first with its synaptic input U_i^E but saturates when U_i^E is sufficiently large, displaying the divisive normalization effect. By choosing the parameters properly, we get the normalization form in equation 2.3.

$W(x, x')$ denotes the neuronal recurrent interaction in the same layer, which is of the gaussian form with strength J_0 and range a,

$$W(x, x') = \frac{J_0}{\sqrt{2\pi}a} \exp\left[-\frac{(x-x')^2}{2a^2}\right]. \tag{2.4}$$

$W_{FF}(x, x')$ and $W_{FB}(x, x')$ are, respectively, the feedforward and feedback interactions between layers, which are chosen to be

$$W_{FF}(x, x') = \frac{J_{ff}}{\sqrt{2\pi}b_1} \exp\left[-\frac{(x-x')^2}{2b_1^2}\right], \tag{2.5}$$

$$W_{FB}(x, x') = \frac{J_{fb}}{\sqrt{2\pi}b_2} \exp\left[-\frac{(x-x')^2}{2b_2^2}\right], \tag{2.6}$$

where the parameters J_{ff} and b_1 control, respectively, the strength and range of feedforward interaction, and so do J_{fb} and b_2 for the feedback interaction. For $J_{fb} > 0$, the feedback interaction is positive, representing the push effect from the second- to the first-layer neurons, whereas for $J_{fb} < 0$, the feedback interaction has a pull effect. Note that in practice, the direct feedback connection between layers of neurons is typically excitatory, and the negative modulation is achieved by interneurons (Sillito et al., 2006).

In our model, the neuronal interactions, including the recurrent, feedforward, and feedback ones, are translationally invariant with respect to the preferred stimuli of neurons: they are decaying functions of the difference between neuronal preferred stimuli $(x - x')$. This implies that the network can hold a continuous family of localized stationary states if the parameters are properly chosen (see Figure 2). In the continuum limit, these stationary

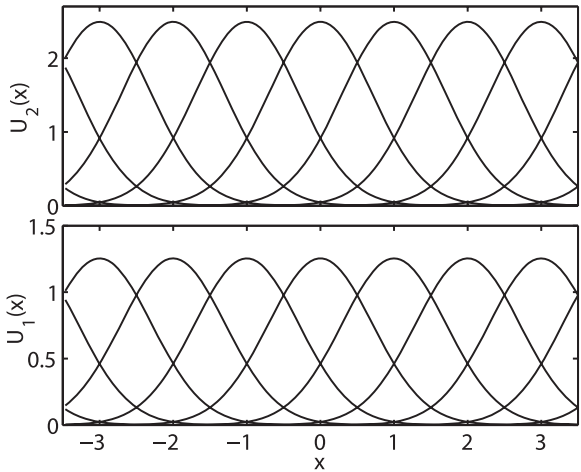

Figure 2: The stationary states of the network when no external input is applied. They are the solutions of equations 2.1 and 2.2. In both layers, the profile of neural population activity displays a bump shape and can be approximated as the gaussian function when the feedforward and feedback interactions are weak. The network holds a continuous family of such bump-shaped stationary states. Parameters: $N = 200$, $b_1 = b_2 = a = 0.5$, $J_0 = \sqrt{2\pi}a$, $J_{ff} = 1$, $J_{fb} = -0.1$.

states form a continuous manifold in which the network is neutrally stable: that is, the network state can translate easily when the external stimulus changes continuously (Amari, 1977; Wu et al., 2002). This property endows the network with the capacities of reading out the stimulus information easily (Deneve, Latham, & Pouget, 1999) and tracking time-varying stimuli smoothly (Samsonovich & McNaughton, 1997; Wu & Amari, 2005). This type of network, called a continuous attractor neural network (CANN), has been successfully applied to describe the encoding of continuous stimuli in neural systems, including orientation (Ben-Yishai et al., 1995), head direction (Zhang, 1996), moving direction (Georgopoulos et al., 1993), and spatial locations of objects (Samsonovich & McNaughton, 1997). Compared with other models in the literature, the network model considered here includes two layers of neurons.

It is instructive to first review the network dynamics when no feedforward or feedback interactions exist. In this case, the two layers of the network are independent of each other, and each of them can support a continuous family of gaussian-shaped stationary states when the global inhibition is below a critical value (Fung, Wong, & Wu, 2010; Wu, Amari, & Nakahara, 2002). These steady states without the external inputs are given

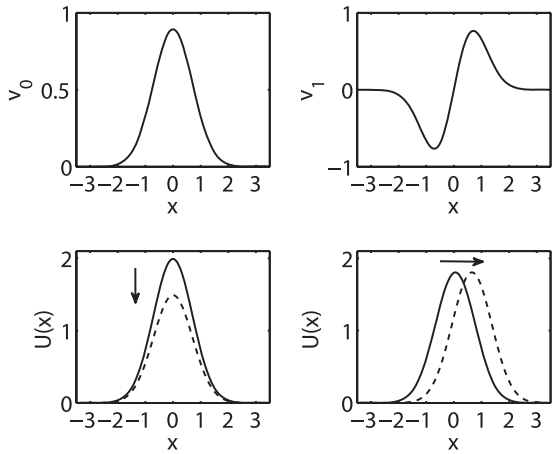

Figure 3: The two dominant motion modes. The first two wave functions of quantum harmonic oscillators and their corresponding physical meanings to the bump dynamics. (Left) The bump height change. (Right) The bump position change.

by (only the case for the first layer is shown)

$$\widetilde{U}_1(x) = \widetilde{U}_1 \exp\left[-\frac{(x-z_1)^2}{4a^2}\right], \tag{2.7}$$

where z_1 is a free parameter, representing the peak position of the bump, and $\widetilde{U}_1 = [1 + (1-k_c)^{1/2}]J_0/(4ak\sqrt{\pi})$. They exist for $0 < k < k_c$, with $k_c = \rho J_0^2/(8a\sqrt{2\pi})$, the critical global inhibition strength above which only silent states with $\widetilde{U}_1 = 0$ exist.

3 Simplifying the Network Dynamics

The key to analyzing the dynamics of a large network is to reduce its dimensionality. A recent approach to solve the dynamics of a CANN analytically (Wu, Hamaguchi, & Amari, 2008; Fung et al., 2010) involves using the wave functions of the quantum harmonic oscillators as the motion modes of a bump state. These modes have clear physical meanings, corresponding to distortions in the height, position, and other higher-order features of the gaussian bump (see Figure 3). The neutral stability of a CANN implies that its dynamics is dominated by a few motion modes. Therefore, by projecting the network dynamics onto these dominant modes, one can simplify the network dynamics significantly.

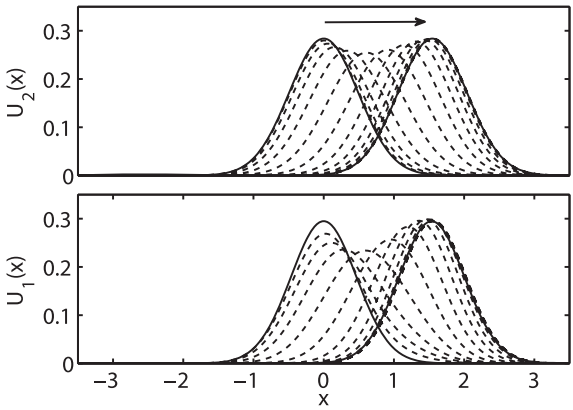

Figure 4: The changes of the network state in response to an abrupt change input. (Top) The bump states in the first layer. (Bottom) The bump states in the second layer.

It turns out that in the case of weak feedforward and feedback interactions, the network stationary states can be well approximated to be of the gaussian form (see Figure 2). Furthermore, for weak external inputs, it is sufficient to include only the first two motion modes, the changes in the bump position and height, to describe the network dynamics. An example displaying the network state changes in a tracking task is presented in Figure 4. We see that during the tracking process, the main changes in the bump state are its position and height, with other higher-order distortions in the shape negligible. Thus, we propose the following gaussian ansatz to approximate the network states:

$$\widetilde{U}_1(x,t) \approx A_1(t)\exp\left\{-\frac{\left[x-z_1(t)\right]^2}{4a_1^2}\right\}, \tag{3.1}$$

$$\widetilde{U}_2(x,t) \approx A_2(t)\exp\left\{-\frac{\left[x-z_2(t)\right]^2}{4a_2^2}\right\}, \tag{3.2}$$

where $A_i(t)$ and $z_i(t)$ are variables describing the height and position changes of the bumps and a_1 and a_2 are the width of the bumps in two layers.

The wave functions for the two dominating motion modes in the first layer are given by (see Figure 3)

$$\text{height}: v_0(x|z) = \exp\left[-\frac{(x-z)^2}{4a_1^2}\right], \tag{3.3}$$

$$\text{position}: v_1(x|z) = \left(\frac{x-z}{a_1}\right)\exp\left[-\frac{(x-z)^2}{4a_1^2}\right]. \tag{3.4}$$

The two dominant motion modes in the second layer have the similar forms, except that a_1 is replaced by a_2. By projecting a function $f(x)$ on a motion mode $u(x|z)$, we mean to compute the quantity, $\int_x f(x)u(x|z)/\int_x u(x|z)$ (see appendix A).

In the following study, we choose the external input to be

$$I_{ext}(x,t) = \alpha \exp\left\{-\frac{\left[x - z_0 + \sigma\xi(t)\right]^2}{4a_1^2}\right\}, \tag{3.5}$$

where z_0 denotes the stimulus value, I_{ext} drives the network state to be stable at z_0 when no noise exists, and α is the input strength. $\xi(t)$ is gaussian white noise of zero mean and unit variance, and σ the noise strength. The exact form of I_{ext} is not critical once it has a unimodal form.

Substituting equations 3.1, 3.2, and 3.5 into the network dynamics, equations 2.1 and 2.2, and projecting them on the two modes, equations 3.3 and 3.4, we obtain the dynamics for the height and position of the bumps in two layers (see appendix A),

$$\tau\frac{dA_1}{dt} = -A_1 + \frac{\sqrt{2}\rho J_0 a_1}{\left(3a_1^2 + a^2\right)^{1/2}}B_1 + \frac{\sqrt{2}\rho J_{fb} a_2}{\left(2a_1^2 + a_2^2 + b_2^2\right)^{1/2}}B_2 + \alpha e^{-(z_0 - z_1)^2/8a_1^2}, \tag{3.6}$$

$$\frac{\tau A_1}{2a_1}\frac{dz_1}{dt} = -\frac{2\sqrt{2}\rho J_{fb} B_2 a_1 a_2}{\left(2a_1^2 + a_2^2 + b_2^2\right)^{3/2}}\left(z_1 - z_2\right) + \frac{\alpha}{2a_1}(z_0 - z_1 + \sigma\xi), \tag{3.7}$$

$$\tau\frac{dA_2}{dt} = -A_2 + \frac{\sqrt{2}\rho J_0 a_2}{\left(3a_2^2 + a^2\right)^{1/2}}B_2 + \frac{\sqrt{2}\rho J_{ff} a_1}{\left(2a_2^2 + a_1^2 + b_1^2\right)^{1/2}}B_1, \tag{3.8}$$

$$\frac{\tau A_2}{2a_2}\frac{dz_2}{dt} = -\frac{2\sqrt{2}\rho J_{ff} B_1 a_1 a_2}{\left(2a_2^2 + a_1^2 + b_1^2\right)^{3/2}}\left(z_2 - z_1\right), \tag{3.9}$$

where $B_i = [A_i]_+^2/(1 + \sqrt{2\pi}k\rho a_i A_i^2)$.

The physical meanings of the above equations are straightforwardly understandable. Consider the dynamics of A_1. The first term on the right-hand side of equation 3.6 corresponds to the decay of neural activity, the second one to the effect of recurrent interaction, the third one to the modulation of feedback interaction from the second layer, and the last one to the contribution of the external input. For the dynamics of z_1, the first term on the right-hand side of equation 3.7 corresponds to the modulation of feedback interaction, and the second one to the driving force of the external input.

Notably, for $J_{fb} > 0$, the positive feedback modulation tends to reduce the separation between two bumps, that is, it decreases the absolute value

of $(z_1 - z_2)$ ($z_2 > z_1$, the first term in the right-hand side of equation 3.7, is positive and contributes to the increase of z_1; for $z_2 < z_1$, the effect is the opposite), whereas for $J_{fb} < 0$, the negative feedback modulation has the opposite effect.

Reorganizing the above equations into the vector and matrix forms, we get

$$\frac{d\mathbf{Z}}{dt} = \mathbf{M}_Z\mathbf{Z} + \mathbf{I}_Z + \boldsymbol{\beta}\xi(t), \tag{3.10}$$

$$\frac{d\mathbf{A}}{dt} = -\frac{1}{\tau}\mathbf{A} + \mathbf{M}_A\mathbf{B} + \mathbf{I}_A, \tag{3.11}$$

where $\mathbf{Z} = (z_1, z_2)^{\mathrm{T}}$, $\mathbf{I}_Z = \left(\alpha z_0/\tau A_1, 0\right)^{\mathrm{T}}$, $\beta = \left(\alpha\sigma/\tau A_1, 0\right)^{\mathrm{T}}$, $\mathbf{A} = (A_1, A_2)^{\mathrm{T}}$, $\mathbf{B} = (B_1, B_2)^{\mathrm{T}}$, $\mathbf{I}_A = (\alpha/\tau e^{-(z_0-z_1)^2/8a_1^2}, 0)^{\mathrm{T}}$, and

$$\mathbf{M}_Z = \begin{bmatrix} -\dfrac{4\sqrt{2}\rho J_{fb}B_2a_1^2a_2}{\tau A_1\left(2a_1^2+a_2^2+b_2^2\right)^{3/2}} - \dfrac{\alpha}{\tau A_1} & \dfrac{4\sqrt{2}\rho J_{fb}B_2a_1^2a_2}{\tau A_1\left(2a_1^2+a_2^2+b_2^2\right)^{3/2}} \\ \dfrac{4\sqrt{2}\rho J_{ff}B_1a_1a_2^2}{\tau A_2\left(2a_2^2+a_1^2+b_1^2\right)^{3/2}} & -\dfrac{4\sqrt{2}\rho J_{ff}B_1a_1a_2^2}{\tau A_2\left(2a_2^2+a_1^2+b_1^2\right)^{3/2}} \end{bmatrix}, \tag{3.12}$$

$$\mathbf{M}_A = \begin{bmatrix} \dfrac{\sqrt{2}\rho J_0a_1}{\tau\left(3a_1^2+a^2\right)^{1/2}} & \dfrac{\sqrt{2}\rho J_{fb}a_2}{\tau\left(2a_1^2+a_2^2+b_2^2\right)^{1/2}} \\ \dfrac{\sqrt{2}\rho J_{ff}a_1}{\tau\left(2a_2^2+a_1^2+b_1^2\right)^{1/2}} & \dfrac{\sqrt{2}\rho J_0a_2}{\tau\left(3a_2^2+a^2\right)^{1/2}} \end{bmatrix}. \tag{3.13}$$

4 Spontaneous Moving Bumps States

We first investigate the stationary states of the network when no external input is applied (by setting $\alpha = 0$ in equation 3.5). With the feedback modulation, the network exhibits new dynamical behaviors. Apart from the static bumps, the network also holds spontaneous moving bump solutions.

The distinguishing difference between static and moving bump states is that in the latter, the bump position is unstable. Hence, by analyzing the stability of bump positions, we can obtain the phase boundary between two states. Suppose z_1^* and z_2^* are the bump positions in the first and second layers. Denote $\delta z_1 = z_1 - z_1^*$ and $\delta z_2 = z_2 - z_2^*$ the fluctuations of the bump positions. Since the bump heights are constants in both states, by linearizing

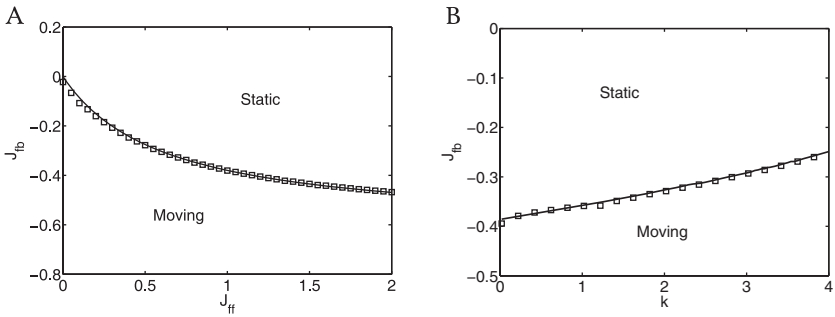

Figure 5: Phase diagram of the intrinsic states of the first-layer CANN in the presence of feedback. Line: theory result. Square: simulation result. The boundaries between the moving and static states are obtained by equation 4.4. Parameters: (A) $k = 0.1$, (B) $J_{ff} = 1$; the others are the same as in Figure 2.

equations 3.7 and 3.9, we get

$$\frac{d}{dt}\begin{pmatrix}\delta z_1\\ \delta z_2\end{pmatrix} = \mathbf{M}_Z\begin{pmatrix}\delta z_1\\ \delta z_2\end{pmatrix}. \tag{4.1}$$

The stability of bump positions is determined by the eigenvalues of matrix $\mathbf{M}_Z$, which is calculated to be $\lambda_\pm = (T_Z \pm \sqrt{(T_Z)^2 - 4D_Z})/2$, with T_Z and D_Z, respectively, the trace and determinant of $\mathbf{M}_Z$. They are calculated to be

$$T_Z = -\frac{4\sqrt{2}\rho J_{fb} B_2 a_1^2 a_2}{\tau A_1\left(2a_1^2 + a_2^2 + b_2^2\right)^{3/2}} - \frac{4\sqrt{2}\rho J_{ff} B_1 a_1 a_2^2}{\tau A_2\left(2a_2^2 + a_1^2 + b_1^2\right)^{3/2}}, \tag{4.2}$$

$$D_Z = \frac{4\sqrt{2}\alpha\rho J_{ff} B_1 a_1 a_2^2}{\tau^2 A_1 A_2\left(2a_2^2 + a_1^2 + b_1^2\right)^{3/2}}. \tag{4.3}$$

The bump positions are stable if the real parts of all eigenvalues are smaller than zero. Hence, the boundary between static and moving bumps is determined by the fact that the real part of λ_+ equals zero, that is,

$$\mathrm{Re}\{\lambda_+\} = \mathrm{Re}\{(T_Z + \sqrt{(T_Z)^2 - 4D_Z})\} = 0 \tag{4.4}$$

The phase diagram for network states in the first layer is summarized in Figure 5. The theoretical predictions based on the simplified network dynamics agree well with the simulations. When the feedback interaction is negative, the network can hold spontaneous moving bump states when

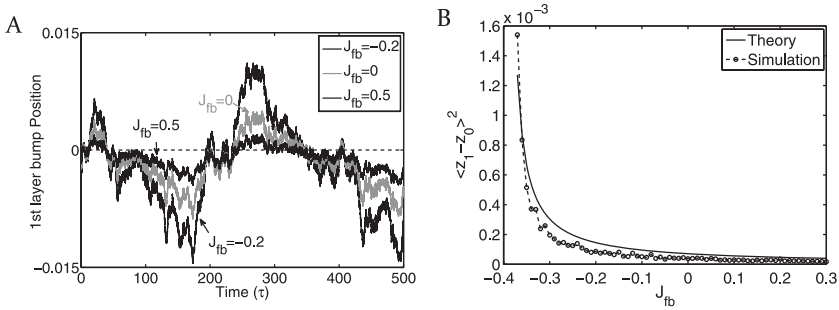

Figure 6: Decoding performance of the network. (A) The typical response behaviors of the network to noisy inputs. The first-layer bump position fluctuates around the true stimulus value $z_0 = 0$. The fluctuations are smaller for positive than negative feedback. (B) The statistical result of the decoding error, which decreases with the positive feedback strength. The parameter regime in which the network holds a static bump is considered. Parameters: $\alpha = 0.1$, $J_{ff} = 1$, $k = 0.1$, $\sigma = 0.1$; the others are the same as in Figure 2.

$|J_{fb}|$ is sufficiently large. Thus, negative feedback has the role of enhancing the mobility of network states.

5 Performances of Population Decoding

A CANN can be interpreted as an efficient framework for implementing population decoding (Deneve et al., 1999; Wu et al., 2002). In response to a transient noisy input, the network bump will evolve to a location having the maximum overlap with it, and the corresponding bump position is the estimated stimulus value. Here, we investigate how the feedback interaction modulates the accuracy of population decoding.

We consider that the network receives an external input in the form of equation 3.5. Due to noise, the bump position in the first layer fluctuates around the true stimulus value z_0 and is modulated by the feedback interaction from the second layer (see Figure 6A). We measure the decoding error of the network by the variance of the bump position after the network reaches a stationary state. By using the simplified dynamics, equations 3.7 and 3.9, we can also analytically calculate the decoding error (see appendix B), which is given by

$$\langle (z_1 - z_0)^2 \rangle = \frac{\alpha\sigma^2}{2\tau A_1} \frac{\sqrt{2}\alpha A_2 + \rho J_{ff} A_1 B_1}{\rho (J_{ff} A_1 B_1 + J_{fb} A_2 B_2) + \sqrt{2}\alpha A_2}. \tag{5.1}$$

From the denominator of the above equation, we see that positive feedback ($J_{fb} > 0$) decreases the error, whereas negative feedback ($J_{fb} < 0$) increases the error. The error diverges for strong negative feedback, since in this parameter regime, the network state becomes a spontaneously moving bump. The simulation results agrees with our theoretical predictions very well (see Figure 6B).

The fact that positive feedback improves population decoding can be intuitively understood as follows. The neural response in the second layer is the integrated result of neural activities in the first layer, and it serves as if holding a memory trace of external inputs. When this information is fed back, the bump position in the first layer is determined by both the instant value and the history of the external inputs. Consequently, temporal noise in the external inputs is largely averaged out, leading to an improved population decoding result.

6 Tracking Performances of the Network

We further investigate how feedback interaction modulates network responses to time-varying stimuli. Two computational tasks are considered: the network catches up to an abrupt change in external stimuli, and it tracks continuously moving stimuli. In the former, we compute the reaction time for the network catching up to the change; in the latter, we measure the maximal speed of a moving stimulus the network can track and quantify the lagging or leading behavior of the bump with respect to the moving stimulus.

6.1 Reaction Time. Consider the value of the external stimulus as it jumps abruptly from $z_0 = 0$ to $z_0 = h$ at time $t = 0$. Because of neutral stability, the network can catch up to this change. The reaction time is defined as the bump position reaches a threshold distance θ to the target location h. Depending on the parameter values, the tracking behavior of the network exhibits two forms: under- and overdamping situations (see Figure 7). The underdamping case is characterized by the bump position's overshooting the target location, which occurs for large negative feedback. For the convenience of presenting results, we introduce the notations conventionally used in control theory (Ogata, 2002), which are

$$\omega_n^2 = D_Z; \quad 2\zeta\omega_n = -T_Z; \quad \omega_d = \omega_n\sqrt{1-\zeta^2}, \tag{6.1}$$

where T_Z and D_Z are given by equations 4.2 and 4.3. The variables ω_n, ω_d, and ζ are called undamped natural frequency, damped natural frequency, and damping ratio, respectively. With these notations, the eigenvalues of the matrix $\mathbf{M}_Z$ are written as $\lambda_\pm = (\zeta \pm \sqrt{\zeta^2 - 1})\omega_n$.

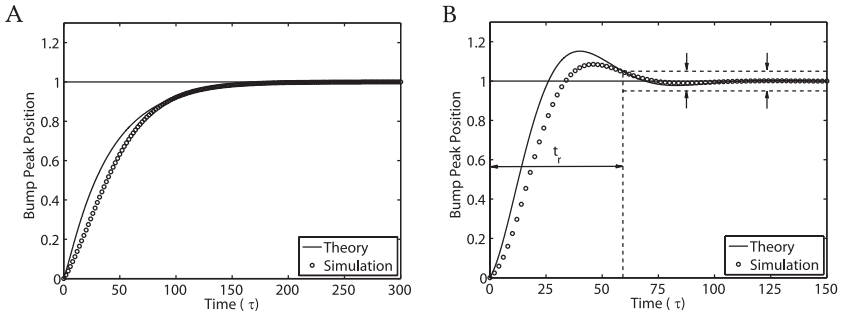

Figure 7: The trace of the bump position in response to an abrupt change in stimuli. $z_0 = 0$ for $t < 0$, and $z_0 = 1$ for $t \geq 0$. (A) The overdamping case. (B) The underdamping case. Reaction time t_r is defined as the time that the bump position falls into a region within a threshold distance to the target. Parameters: $\alpha = 0.1, k = 0.1$. (A) $J_{ff} = 0.8, J_{fb} = -0.15$. (B) $J_{ff} = 0.8, J_{fb} = -0.3$. Other parameters are the same as in Figure 2.

The bump position oscillates during tracking if the damping ratio $0 < \zeta < 1$, corresponding to the imaginary part of the eignevalue, is smaller than 0, that is, $\mathrm{Im}\{\lambda_{\pm}\} < 0$. If $\zeta \geq 1$, the bump position approaches the target from below without overshooting, corresponding to the overdamping situation.

With the new denotations, the dynamics of bump positions in equations 3.7 and 3.9 are rewritten as

$$z_1'' + 2\zeta\omega_n z_1' + \omega_n^2 z_1 = a_0(z_0' - m_{22}z_0), \tag{6.2}$$

$$z_2'' + 2\zeta\omega_n z_2' + \omega_n^2 z_2 = a_0 m_{21} z_0, \tag{6.3}$$

where m_{ij} is the element of the matrix $\mathbf{M}_Z$, and $a_0 = \alpha/\tau A_1$

The solutions of the above equations are presented below. The detailed derivation is given in appendix C:

- Underdamping case. In the underdamping situation ($\zeta < 1$), the dynamics of the bump position is solved to be

$$z_1(t) = h + he^{-\zeta\omega_n t}\left[\frac{a_0 - \zeta\omega_n}{\omega_d}\sin(w_d t) - \cos(\omega_d t)\right]. \tag{6.4}$$

- Overdamping case. In the overdamping situation ($\zeta \geq 1$), the dynamics of the bump position is solved to be

$$z_1(t) \approx h - he^{-(\zeta - \sqrt{\zeta^2 - 1})\omega_n t}\left[\left(\frac{a_0}{\omega_n} - \zeta + \sqrt{\zeta^2 - 1}\right)(\zeta - \sqrt{\zeta^2 - 1})\right]. \tag{6.5}$$

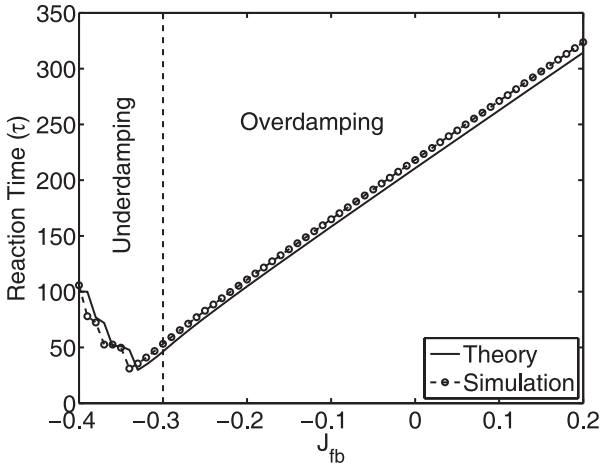

Figure 8: The reaction time of the network to an abrupt change in stimuli. The dashed line is the boundary between under- and overdamping cases. The minimal reaction time is obtained when the bump position oscillates a little at $\zeta = 0.7$. Parameters: $\alpha = 0.1$, $k = 0.1$, $J_{ff} = 1.2$, $h = 0.5$, $\theta = 5\%$; the others are the same as in Figure 2.

Figure 8 shows the dependence of reaction time on the feedback strength. The simulation results agree well with our theoretic analysis. We observe the following interesting properties:

- Compared with positive feedback, negative feedback tends to decrease the reaction time. This agrees with our previous analysis that negative feedback increases the mobility of the network state.
- The minimal reaction time is obtained when the bump position oscillates a bit (the damping ratio is about 0.7), suggesting that a little overshooting is helpful for tracking.
- The reaction time does not increase smoothly with the feedback strength when the bump position oscillates. This is because a small change in ζ induces a dramatic change in the reaction time, agreeing with the control theory (Ogata, 2002).

6.2 Effect of the Range of Feedback Interaction. We also investigate the effect of the feedback range on network performance. For the convenience of analysis, we consider the widths of the feedforward, the feedback, and the recurrent interactions to be same ($b_1 = b_2 = a$). This ensures that the gaussian ansatz equations 3.1 and 3.2, holds well, and the network behaviors can be well approximated by equations 5.1, 6.4, and 6.5. When these widths are not equal, the network performances are still qualitatively similar

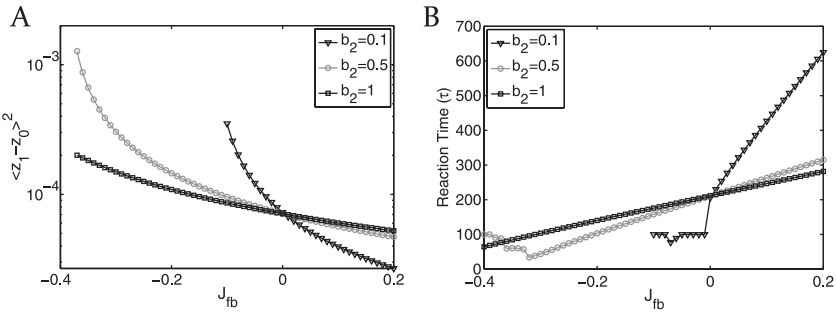

Figure 9: Network performance in three feedback ranges. The parameter regime in which the network holds a static bump is considered. (A) Decoding error of the network. The results are obtained by equation 5.1. (B) Reaction time of the network in response to an abrupt change in stimuli. The results are obtained by equations 6.4 and 6.5. The other parameters are the same as in Figures 6 and 8.

(data not shown), but we need to include high-order distortions in the bump state to get good theoretical predictions.

Figure 9 shows the network decoding performances to noisy inputs and the network reaction times to an abrupt change in inputs in three different feedback widths (for clarity, only the theoretical results are shown). We find that the feedback range tends to have a trade-off effect on the decoding error and the reaction time of the network and that this effect is opposite for positive and negative feedback interactions: for negative feedback, narrow feedback (with small b_2) increases the decoding error but decreases the reaction time of the network, and for positive feedback, narrow feedback decreases the decoding error but increases the reaction time of the network.

6.3 Tracking Continuously Moving Stimuli. We explore the tracking performances of the network to continuously moving stimuli. Without loss of generality, we consider that the stimulus value increases with time at a constant positive speed, $z_0 = vt$, with $v > 0$. We investigate how the feedback interaction modulates the tracking performance of the network.

Let us denote the discrepancies between the bump positions and the stimulus value to be $s_i = z_0 - z_i$, for $i = 1, 2$. $s_i > 0$ ($s_i < 0$) means the bump position is lagging (leading) the true stimulus value. Substituting s_i in the simplified dynamics, equations 3.7 and 3.9, we obtain

$$v - \frac{ds_1}{dt} = -m_{12}(s_2 - s_1)e^{-(s_2-s_1)^2/8a^2} + \alpha s_1 e^{-s_1^2/8a^2}/(\tau A_1), \tag{6.6}$$

$$v - \frac{ds_2}{dt} = m_{21}(s_2 - s_1)e^{-(s_2-s_1)^2/8a^2}. \tag{6.7}$$

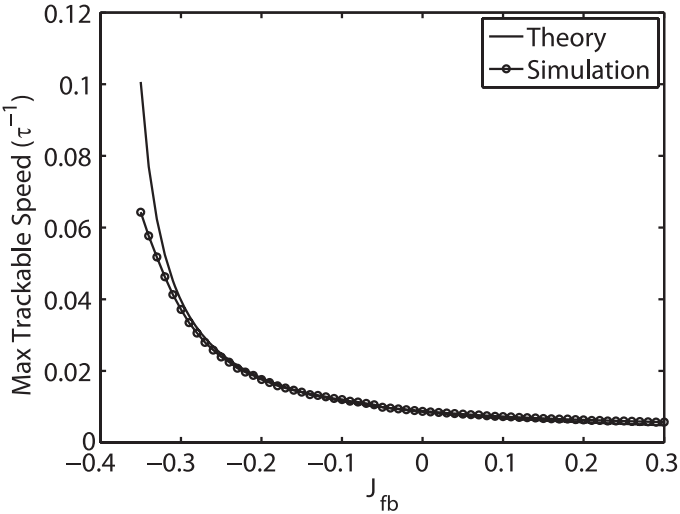

Figure 10: The maximal speed of a moving stimulus the network can track. It increases with the strength of negative feedback. Parameters: $N = 400$, $k = 0.1$, $J_{ff} = 1$, $\alpha = 0.1$; the others are the same as in Figure 2.

In a steady state, s_i, for $i = 1, 2$, become constants. From the above equations, we get

$$v = \frac{a_0 m_{21}}{m_{12} + m_{21}} s_1 e^{-s_1^2/8a^2}. \tag{6.8}$$

6.3.1 The Maximum Trackable Speed. From equation 6.8, we can estimate the maximal speed of the moving stimulus the network can track, which is the largest value of the function on the right-hand side of the equation, that is,

$$\begin{aligned} v_{\max} &= \mathbf{max}_{s_1} \left[\frac{a_0 m_{21}}{m_{12} + m_{21}} s_1 e^{-s_1^2/8a^2} \right], \\ &= \frac{2a\alpha J_{ff} B_1}{\tau\sqrt{e}(J_{ff} A_1 B_1 + J_{fb} A_2 B_2)}. \end{aligned} \tag{6.9}$$

Above $v_{\max}$, the network is unable to catch up with the stimulus. The relationship between the maximal trackable speed and the feedback strength is shown in Figure 10. The simulation results agree well with our analysis. We see that the maximal speed the network can track increases with the magnitude of negative feedback.

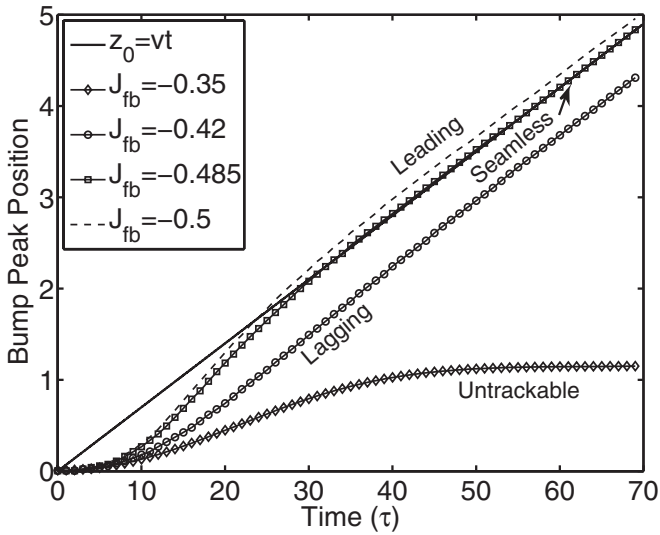

Figure 11: The tracking behaviors of the network to a moving stimulus. $z_0 = t$. Parameters: $N = 400, k = 0.1, J_{ff} = 1.5, V = 0.07, \alpha = 0.1$; the others are the same as in Figure 2.

6.3.2 Lagging or Leading of the Bump with Respect to the Moving Stimulus. Since $v \geq 0$ as predefined, equations 3.12 and 6.8 tell us that

$$(m_{12} + m_{21})s_1 \geq 0, \tag{6.10}$$

where $(m_{12} + m_{21}) = -T_Z$ with T_Z the trace of the matrix $\mathbf{M}_Z$ given by equation 3.12. We make the following observations (see Figure 11):

- When $T_z < 0$, we have $s_1 > 0$. The bump is lagging behind the moving stimulus. In the parameter regime for $T_z < 0$, the network holds static bumps (see Figure 5).
- When $T_z > 0$, we have $s_1 < 0$. The bump is leading the moving stimulus. In the parameter regime for $T_z < 0$, the network holds spontaneous moving bumps (see Figure 5).
- When $T_z = 0$, we have $s_1 = 0$. The bump tracks the moving stimulus seamlessly. The parameter values in this case are on the boundary between static and moving bumps (see Figure 5).

7 Conclusion and Discussion

In this study, we have explored the potential roles of feedback interactions in neural information processing. We consider a two-layer network model

in which neurons in the first layer receive feedback inputs from those in the second layer. The neuronal connections are structured so that the network can hold a continuous family of bump states, mimicking, for instance, the orientation tuning in the visual system. By utilizing the intrinsic dynamics of a CANN, we use a projection method to reduce the dimensionality of the network dynamics significantly, that is, we consider only the changes in the height and position of a bump, the two dominating motion modes of the network. The simplified dynamics allows us to elucidate the effects of feedback interactions analytically and hence gives us insight into understanding the principles of feedback modulations. The simulation results agree very well with our theoretic predictions. We have observed a number of interesting behaviors that may have far-reaching implications in neural information processing.

First, we find that negative feedback increases the mobility of network states, inducing spontaneously moving bumps in a neural system. The moving bump solution may be related to the traveling wave phenomenon widely observed in the cortex (Wu, Huang, & Zhang, 2008). Previous studies have shown the Mexican hat recurrent interaction (Pinto & Ermentrout, 2001), short-term depression (York & van Rossum, 2009), and spike adaptation (Hansel & Sompolinsky, 1997) can all generate moving bumps. Our result provides another potential mechanism for the origin of traveling waves.

Second, we show that positive feedback improves the accuracy of population decoding. The working mechanism is as follows. The neural activity in the second layer holds a memory trace of external inputs due to time delay and temporal integration. This information is then passed back to the first layer through feedback connections. Consequently, the neural response in the first layer is determined by both the instant value and the history of external inputs, which largely averages out temporal noise. This view agrees with the general idea of predictive coding that recognizing objects needs the interplay between top-down and bottom-up information (Lee & Mumford, 2003; Rao & Ballard, 1999).

Third, we show that negative feedback enhances the tracking performance of the network to time-varying stimuli. For an abrupt change in stimuli, negative feedback shortens the reaction time. For continuously moving stimuli, negative feedback increases the maximal speed the network can track.

Furthermore, we find that with strong negative feedback, the network response can seamlessly track or even lead the actual position of external stimuli, achieving anticipative behavior. Anticipative neural response is important for motion extrapolation and movement control. It has been observed in the head-direction system and hippocampus (Blair & Sharp, 1995; O'Keefe & Recce, 1993). Our study reveals a potential mechanism for realizing these functions. Notably, the parameter regime for a network having anticipatory behaviors is also the one for the network holding

spontaneously moving bumps, suggesting that the widely observed traveling wave phenomenon may be an unavoidable consequence for neural systems with prediction powers.

In general, we find that positive and negative feedback tends to have opposite effects on modulating the network dynamics. The former enhances the stability of the network state, leading to improved population decoding accuracy, whereas the latter increases the mobility of the network state, accelerating the network reaction to time-varying stimuli. So can a neural system have the advantages of both? A possible solution is that the feedback interaction is biphasic in time, being positive initially and negative subsequently, as often observed in feedforward temporal filters. With biphasic feedback modulation, the network can average out high-frequency noise and track slow-frequency changes in external inputs (the latter defines the stimuli in practice, see Wiskott & Sejnowski, 2002). In future work, we seek to extend the current model to include biphasic feedback modulations.

Appendix A: The Projection Method

Substituting the gaussian ansatz, equations 3.1 and 3.2, into the network dynamics, equations 2.1 and 2.2, we obtain the following results.

For equation 2.1, its left-hand side (LHS) and right-hand side (RHS) are given by

$$\mathrm{LHS} = \tau \frac{dA_1}{dt} \exp\left[-\frac{(x-z_1)^2}{4a_1^2}\right] + \frac{\tau A_1}{2a_1}\frac{dz_1}{dt}\left(\frac{x-z_1}{a_1}\right)\exp\left[-\frac{(x-z_1)^2}{4a_1^2}\right]. \tag{A.1}$$

$$\begin{aligned}
\mathrm{RHS} &= -A_1 \exp\left[-\frac{(x-z_1)^2}{4a_1^2}\right] + \rho \int_{x'} \frac{J_0 B_1}{a\sqrt{2\pi}} \exp\left[-\frac{(x-x')^2}{2a^2}\right] \\
&\quad \times \exp\left[-\frac{(x'-z_1)^2}{2a_1^2}\right] dx' + \rho \int_{x'} \frac{J_{fb} B_2}{b_2\sqrt{2\pi}} \exp\left[-\frac{(x-x')^2}{2b_2^2}\right] \\
&\quad \times \exp\left[-\frac{(x'-z_2)^2}{2a_2^2}\right] dx' + \alpha \exp\left[-\frac{(x-z_0)^2}{4a_1^2}\right] + \sigma\xi(x,t) \\
&= -A_1 \exp\left[-\frac{(x-z_1)^2}{4a_1^2}\right] + \frac{\rho J_0 B_1 a_1}{\left(a^2+a_1^2\right)^{1/2}} \exp\left[-\frac{(x-z_1)^2}{2(a^2+a_1^2)}\right], \\
&\quad + \frac{\rho J_{fb} B_2 a_2}{\left(b_2^2+a_2^2\right)^{1/2}} \exp\left[-\frac{(x-z_2)^2}{2(b_2^2+a_2^2)}\right] + \alpha \exp\left[-\frac{(x-z_0+\sigma\xi(t))^2}{4a_1^2}\right].
\end{aligned} \tag{A.2}$$

For equation 2.2,

$$\mathrm{LHS} = \tau \frac{dA_2}{dt} \exp\left[-\frac{(x-z_2)^2}{4a_2^2}\right] + \frac{\tau A_2}{2a_2}\frac{dz_2}{dt}\left(\frac{x-z_2}{a_2}\right)\exp\left[-\frac{(x-z_2)^2}{4a_2^2}\right]. \tag{A.3}$$

$$\begin{aligned}
\mathrm{RHS} &= -A_2 \exp\left[-\frac{(x-z_2)^2}{4a_2^2}\right] + \rho \int_{x'} \frac{J_0 B_2}{a\sqrt{2\pi}} \exp\left[-\frac{(x-x')^2}{2a^2}\right] \\
&\quad \times \exp\left[-\frac{(x'-z_2)^2}{2a_2^2}\right] dx' + \rho \int_{x'} \frac{J_{ff} B_1}{b_1\sqrt{2\pi}} \exp\left[-\frac{(x-x')^2}{2b_1^2}\right] \\
&\quad \times \exp\left[-\frac{(x'-z_1)^2}{2a_1^2}\right] dx' \\
&= -A_2 \exp\left[-\frac{(x-z_1)^2}{4a_2^2}\right] + \frac{\rho J_0 B_2 a_2}{\left(a^2+a_2^2\right)^{1/2}} \exp\left[-\frac{(x-z_2)^2}{2(a^2+a_2^2)}\right] \\
&\quad + \frac{\rho J_{ff} B_1 a_1}{\left(b_1^2+a_1^2\right)^{1/2}} \exp\left[-\frac{(x-z_1)^2}{2(b_1^2+a_1^2)}\right].
\end{aligned} \tag{A.4}$$

Projecting equations A.1 and A.2 onto the motion modes, equations 3.3 and 3.4, we get

$$\begin{aligned}
\tau \frac{dA_1}{dt} &= -A_1 + \frac{\sqrt{2}\rho J_0 B_1 a_1}{(3a_1^2+a^2)^{1/2}} + \frac{\sqrt{2}\rho J_{fb} B_2 a_2}{(2a_1^2+a_2^2+b_2^2)^{1/2}} \\
&\quad \times \exp\left[-\frac{\left(z_1-z_2\right)^2}{2\left(2a_1^2+a_2^2+b_2^2\right)}\right] + \alpha \exp\left[-\frac{(z_0-z_1+\sigma\xi(t))^2}{8a_1^2}\right],
\end{aligned} \tag{A.5}$$

$$\begin{aligned}
\frac{\tau A_1}{2a_1}\frac{dz_1}{dt} &= -\frac{2\sqrt{2}\rho J_{fb} B_2 a_1 a_2}{\left(2a_1^2+a_2^2+b_2^2\right)^{3/2}}\left(z_1-z_2\right)\exp\left[-\frac{\left(z_1-z_2\right)^2}{2\left(2a_1^2+a_2^2+b_2^2\right)}\right] \\
&\quad + \frac{\alpha}{2a_1}(z_0-z_1+\sigma\xi(t))\exp\left[-\frac{(z_0-z_1+\sigma\xi(t))^2}{8a_1^2}\right].
\end{aligned} \tag{A.6}$$

When $(z_1-z_2)^2/(2a_1^2+a_2^2+b_2^2)$ is sufficiently small (which is the case for the parameters we choose) and noise is small enough $\sigma^2/(8a_1^2) \ll 1$, we obtain the dynamical equations 3.6 and 3.7.

Similarly, we project equations A.3 and A.4 onto the motion modes and obtain

$$\tau \frac{dA_2}{dt} = -A_2 + \frac{\sqrt{2}\rho J_0 B_2 a_2}{(3a_2^2 + a^2)^{1/2}} + \frac{\sqrt{2}\rho J_{ff} B_1 a_1}{(2a_2^2 + a_1^2 + b_1^2)^{1/2}} \times \exp\left[-\frac{(z_1 - z_2)^2}{2\left(2a_2^2 + a_1^2 + b_1^2\right)}\right], \tag{A.7}$$

$$\frac{\tau A_2}{2a_2}\frac{dz_2}{dt} = -\frac{2\sqrt{2}\rho J_{ff} B_1 a_1 a_2}{\left(2a_2^2 + a_1^2 + b_1^2\right)^{3/2}}\left(z_2 - z_1\right)\exp\left[-\frac{(z_1 - z_2)^2}{2\left(2a_2^2 + a_1^2 + b_1^2\right)}\right]. \tag{A.8}$$

Under the conditions that $(z_1 - z_2)^2/(2a_2^2 + a_1^2 + b_1^2)$ and $\sigma^2/(8a_2^2)$ are sufficiently small, we get the dynamical equations, 3.8 and 3.9.

Appendix B: The Error of Population Decoding

The solution of equation 3.11 is given by

$$\mathbf{Z}(t) = e^{\mathbf{M}_Z t}\left[\mathbf{Z}(0) + \int_0^t e^{-\mathbf{M}_Z s}\mathbf{I}_Z(s)ds + \int_0^t e^{-\mathbf{M}_Z s}\boldsymbol{\beta} dW_s\right], \tag{B.1}$$

where dW_s denotes the standard Wiener process.

The mean and variance of $\mathbf{Z}(t)$ are calculated:

$$E[\mathbf{Z}(t)] = e^{\mathbf{M}_Z t}E[\mathbf{Z}(0)] + \int_0^t e^{\mathbf{M}_Z (t-s)}\mathbf{I}_Z(s)ds, \tag{B.2}$$

$$\mathbf{Z}(t) - E[\mathbf{Z}(t)] = e^{\mathbf{M}_Z t}\left[\mathbf{Z}(0) - E[\mathbf{Z}(0)] + \int_0^t e^{-\mathbf{M}_Z s}\boldsymbol{\beta} dW_s\right], \tag{B.3}$$

$$\begin{aligned}\mathrm{Var}[\mathbf{Z}(t)] &= E\left[(\mathbf{Z}(t) - E[\mathbf{Z}(t)])(\mathbf{Z}(t) - E[\mathbf{Z}(t)])^{\mathrm{T}}\right] \\ &= e^{\mathbf{M}_Z t}\left\{\mathrm{Var}[\mathbf{Z}(0)] + E\left(\int_0^t e^{-\mathbf{M}_Z s}\boldsymbol{\beta} dW_s\right)\right. \\ &\quad \left.\times \left(\int_0^t e^{-\mathbf{M}_Z s}\boldsymbol{\beta} dW_s\right)^{\mathrm{T}}\right\}(e^{\mathbf{M}_Z t})^{\mathrm{T}}.\end{aligned} \tag{B.4}$$

Let $\mathbf{B}(t) = \mathrm{Var}[\mathbf{Z}(0)] + E\left(\int_0^t e^{-\mathbf{M}_Z s}\boldsymbol{\beta} dW_s\right)\left(\int_0^t e^{-\mathbf{M}_Z s}\boldsymbol{\beta} dW_s\right)^{\mathrm{T}}$, so

$$\begin{aligned}\frac{d\mathrm{Var}[\mathbf{Z}(t)]}{dt} &= \frac{de^{\mathbf{M}_Z t}}{dt}\mathbf{B}(t)(e^{\mathbf{M}_Z t})^{\mathrm{T}} + e^{\mathbf{M}_Z t}\frac{d\mathbf{B}(t)}{dt}(e^{\mathbf{M}_Z t})^{\mathrm{T}} + e^{\mathbf{M}_Z t}\mathbf{B}(t)\frac{d(e^{\mathbf{M}_Z t})^{\mathrm{T}}}{dt} \\ &= \mathbf{M}_Z e^{\mathbf{M}_Z t}\mathbf{B}(t)(e^{\mathbf{M}_Z t})^{\mathrm{T}} + \boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}} + e^{\mathbf{M}_Z t}\mathbf{B}(t)(e^{\mathbf{M}_Z t})^{\mathrm{T}}\mathbf{M}_Z^{\mathrm{T}} \\ &= \mathbf{M}_Z \mathrm{Var}[\mathbf{Z}(t)] + Var[\mathbf{Z}(t)]\mathbf{M}_Z^{\mathrm{T}} + \boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}. \end{aligned} \tag{B.5}$$

Since $\mathrm{Var}[\mathbf{Z}(t)]$ is symmetric, we have

$$\frac{d\mathrm{Var}[\mathbf{Z}(t)]}{dt} = \mathbf{M}_Z \mathrm{Var}[\mathbf{Z}(t)] + (\mathbf{M}_Z \mathrm{Var}[\mathbf{Z}(t)])^{\mathrm{T}} + \boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}. \tag{B.6}$$

The steady value of $\mathrm{Var}[Z(t)]$ is the decoding error of the network, which satisfies

$$\mathbf{M}_Z \mathrm{Var}(\mathbf{Z}) + [\mathbf{M}_Z \mathrm{Var}(\mathbf{Z})]^{\mathrm{T}} = -\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}. \tag{B.7}$$

Denote

$$\mathrm{Var}[\mathbf{Z}(t)] \equiv \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}, \quad \mathbf{M}_Z \equiv \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}. \tag{B.8}$$

where $V_{12} = V_{21}$. Equation B.7 becomes

$$\begin{aligned} &\mathbf{M}_Z \mathrm{Var}(\mathbf{Z}) + [\mathbf{M}_Z \mathrm{Var}(\mathbf{Z})]^{\mathrm{T}} \\ &= \begin{pmatrix} 2m_{11}V_{11} - 2(m_{11}+a_0)V_{21} & m_{21}(V_{11}-V_{21}) + m_{11}V_{12} - (m_{11}+a_0)V_{22} \\ m_{21}(V_{11}-V_{21}) + m_{11}V_{12} - (m_{11}+a_0)V_{22} & 2m_{21}(V_{12}-V_{22}) \end{pmatrix} \\ &= \begin{pmatrix} -\dfrac{\alpha^2\sigma^2}{\tau^2 A_1^2} & 0 \\ 0 & 0 \end{pmatrix}, \end{aligned} \tag{B.9}$$

where $a_0 = \alpha/\tau A_1$. Therefore,

$$\begin{aligned}\mathrm{Var}(z_1) &= \frac{\alpha\sigma^2}{2\tau A_1}\frac{a_0 + m_{21}}{m_{21} - m_{11}} \\ &= \frac{\alpha\sigma^2}{2\tau A_1}\frac{\sqrt{2}\alpha A_2 + \rho J_{ff} A_1 B_1}{\rho(J_{ff}A_1B_1 + J_{fb}A_2B_2) + \sqrt{2}\alpha A_2}, \end{aligned} \tag{B.10}$$

$$\begin{aligned}\mathrm{Var}(z_2) &= \frac{m_{21}}{a_0 + m_{21}}\mathrm{Var}(z_1) \\ &= \frac{\alpha\sigma^2}{2\tau A_1}\frac{\rho J_{ff}A_1B_1}{\rho(J_{ff}A_1B_1 + J_{fb}A_2B_2) + \sqrt{2}\alpha A_2}.\end{aligned} \tag{B.11}$$

Appendix C: The Network Response to an Abrupt Change in Stimuli

Without loss of generality, we consider the bump position $z_1(t) = z_2(t) = 0$, for $t < 0$. The Laplace transformations of equations 6.2 and 6.3 are

$$z_1(s) = \frac{a_0(s - m_{22})}{s^2 + 2\zeta\omega_n s + \omega_n^2}z_0(s), \tag{C.1}$$

$$z_2(s) = \frac{a_0 m_{21}}{s^2 + 2\zeta\omega_n s + \omega_n^2}z_0(s), \tag{C.2}$$

where $z_i(s)$ is the Laplace transform of $z_i(t)$. $z_0(s) = h/s$ is the Laplace transform of the abrupt change in the stimulus values, where h is the size of abrupt change.

Substituting $z_0(s)$ into equation C.2, the solution of the first-layer dynamics is divided into under-($0 < \zeta < 1$) and over-($\zeta > 1$) damping situations:

- Underdamping situation:

$$\begin{aligned}z_1(s) = \frac{h}{s} + \frac{1}{2\omega_n\sqrt{\zeta^2 - 1}}\Bigg[&\frac{a_0 - \omega_n(\zeta + \sqrt{\zeta^2 - 1})}{s + \zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1}} \\ &- \frac{a_0 - \omega_n(\zeta - \sqrt{\zeta^2 - 1})}{s + \zeta\omega_n + \omega_n\sqrt{\zeta^2 - 1}}\Bigg].\end{aligned} \tag{C.3}$$

  Using the inverse-Laplace transform, we get

$$z_1(t) = z_0 + z_0 e^{-\zeta\omega_n t}\left[\frac{a_0 - \zeta\omega_n}{\omega_d}\sin(w_d t) - \cos(\omega_d t)\right]. \tag{C.4}$$

- Overdamping situation:

$$\begin{aligned}z_1(s) = \frac{h}{s} - z_0\Bigg[&\frac{1}{s + \zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1}} \\ &- \frac{(a_0 + \omega_n\sqrt{\zeta^2 - 1} - \zeta\omega_n)}{(s + \zeta\omega_n + \omega_n\sqrt{\zeta^2 - 1})(s + \zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1})}\Bigg]\end{aligned}$$

$$\approx \frac{h}{s} - z_0 \left[\frac{1}{s + \zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1}} - \frac{(a_0 - \zeta\omega_n + \omega_n\sqrt{\zeta^2 - 1})(\zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1})}{\omega_n^2(s + \zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1})} \right] \tag{C.5}$$

In order to get the expression of reaction time t_r, an approximation is made that neglects the smaller polar point in the complex plane of frequency domain in the above equation. By using inverse-Laplace transform, we obtain

$$z_1(t) \approx z_0 - z_0 e^{-(\zeta - \sqrt{\zeta^2 - 1})\omega_n t} \times \left[1 - \left(\frac{a_0}{\omega_n} - \zeta + \sqrt{\zeta^2 - 1} \right) (\zeta - \sqrt{\zeta^2 - 1}) \right]. \tag{C.6}$$

Acknowledgments

We acknowledge valuable discussions with K. Y. Michael Wong and C. C. Alan Fung. This work is supported by the 973 program of China (No. 2011CBA00406) and the National Foundation of Natural Science of China (No. 91132702).

References

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, *27*(2), 77–87.

Ben-Yishai, R., Bar-Or, R., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences*, *92*(9), 3844–3848.

Blair, H., & Sharp, P. (1995). Anticipatory head direction signals in anterior thalamus: Evidence for a thalamocortical circuit that integrates angular head motion to compute head direction. *Journal of Neuroscience*, *15*(9), 6260–6270.

Deneve, S., Latham, P., & Pouget, A. (1999). Reading population codes: A neural implementation of ideal observers. *Nature Neuroscience*, *2*(8), 740–745.

Fung, C., Wong, K., & Wu, S. (2010). A moving bump in a continuous manifold: A comprehensive study of the tracking dynamics of continuous attractor neural networks. *Neural Computation*, *22*(3), 752–792.

Georgopoulos, A., Taira, M., & Lukashin, A. (1993). Cognitive neurophysiology of the motor cortex. *Science*, *260*(5104), 47–52.

Gulyas, B., Lagae, L., Eysel, U., & Orban, G. (1990). Corticofugal feedback influences the responses of geniculate neurons to moving stimuli. *Experimental Brain Research*, *79*(2), 441–446.

Hansel, D., & Sompolinsky, H. (1997). Modeling feature selectivity in local cortical circuits. In C. Koch & I. Segev (Eds.), *Methods in neuronal modeling: From synapse to networks* (2nd ed., pp. 499–567). Cambridge, MA: MIT Press.

Hao, J., Wang, X., Dan, Y., Poo, M., & Zhang, X. (2009). An arithmetic rule for spatial summation of excitatory and inhibitory inputs in pyramidal neurons. *Proceedings of the National Academy of Sciences*, *106*(51), 21906–21911.

Heeger, D. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, *9*, 181–197.

Latawiec, D., Martin, K., & Meskenaite, V. (2000). Termination of the geniculocortical projection in the striate cortex of macaque monkey: A quantitative immunoelectron microscopic study. *Journal of Comparative Neurology*, *419*(3), 306–319.

Lee, T., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, *20*(7), 1434–1448.

Lund, J., Angelucci, A., & Bressloff, P. (2003). Anatomical substrates for functional columns in macaque monkey primary visual cortex. *Cerebral Cortex*, *13*(1), 15–24.

McClurkin, J., Optican, L., & Richmond, B. (1994). Cortical feedback increases visual information transmitted by monkey parvocellular lateral geniculate nucleus neurons. *Visual Neuroscience*, *11*(03), 601–617.

Murphy, P., & Sillito, A. (1987). Corticofugal feedback influences the generation of length tuning in the visual pathway. *Nature*, *329*, 727–729.

Ogata, K. (2002). *Modern control engineering*. Upper Saddle River, NJ: Prentice Hall.

O'Keefe, J., & Recce, M. (1993). Phase relationship between hippocampal place units and the eeg theta rhythm. *Hippocampus*, *3*(3), 317–330.

Pack, C., & Born, R. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature*, *409*(6823), 1040–1042.

Pinto, D., & Ermentrout, G. (2001). Spatially structured activity in synaptically coupled neuronal networks: I. Traveling fronts and pulses. *SIAM Journal on Applied Mathematics*, *62*(1), 206–225.

Rao, R., & Ballard, D. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79–87.

Samsonovich, A., & McNaughton, B. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. *Journal of Neuroscience*, *17*(15), 5900–5920.

Sillito, A., Cudeiro, J., & Jones, H. (2006). Always returning: Feedback and sensory processing in visual cortex and thalamus. *Trends in Neurosciences*, *29*(6), 307–316.

Van Horn, S., Erişir, A., & Sherman, S. (2000). Relative distribution of synapses in the α-laminae of the lateral geniculate nucleus of the cat. *Journal of Comparative Neurology*, *416*(4), 509–520.

Wang, W., Jones, H., Andolina, I., Salt, T., & Sillito, A. (2006). Functional alignment of feedback effects from visual cortex to thalamus. *Nature Neuroscience*, *9*(10), 1330–1336.

Wiskott, L., & Sejnowski, T. (2002). Slow feature analysis: Unsupervised learning of invariances. *Neural Computation*, *14*(4), 715–770.

Wu, J., Huang, X., & Zhang, C. (2008). Propagating waves of activity in the neocortex: What they are, what they do. *Neuroscientist*, *14*(5), 487–502.

Wu, S., & Amari, S. (2005). Computing with continuous attractors: Stability and online aspects. *Neural Computation*, *17*(10), 2215–2239.

Wu, S., Amari, S., & Nakahara, H. (2002). Population coding and decoding in a neural field: A computational study. *Neural Computation*, *14*(5), 999–1026.

Wu, S., Hamaguchi, K., & Amari, S. (2008). Dynamics and computation of continuous attractors. *Neural Computation*, *20*(4), 994–1025.

York, L., & van Rossum, M. (2009). Recurrent networks with short term synaptic depression. *Journal of Computational Neuroscience*, *27*(3), 607–620.

Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *Journal of Neuroscience*, *16*(6), 2112–2126.

Received September 30, 2011; accepted December 21, 2011.