Perturbations of cortical "ringing" in a model with local, feedforward, and feedback recurrence

Kimberly E. Reinhold Department of Neuroscience University of California, San Diego San Diego, CA 92037 *kreinhol@ucsd.edu*

Abstract

Recurrence is one of the fundamental functions attributed to neocortex. Classically, studies of recurrence in cortex have focused on higher-order association areas, such as prefrontal cortex, involved in working memory tasks and showing neural correlates of representation maintenance. However, the role of recurrence in primary sensory cortices is debated. Recurrence is thought to orchestrate, on a moment-by-moment basis, the balance of excitation and inhibition within a local network. What is controversial is the degree to which recurrence in primary sensory areas also enables the maintenance of a representation over time. We may begin to tackle this question using optogenetic methods. Genetic manipulations in the mouse allow the expression of channelrhodopsin (ChR) in parvalbumin-positive (PV+) interneurons, which comprise 95% of the cortical interneurons. By synchronous activation of many inhibitory cells, we can transiently interrupt recurrent activity in primary visual cortex (V1) after a stimulus and determine the effect of this manipulation on maintenance of the cortical representation. Furthermore, many of these effects can be recapitulated in a rate model of two simple, coupled cortical areas, to allow insights into the subtleties of the studied system and to suggest theories for the observed phenomena. The model explains the preferentially sustained network response in the gamma frequency range (30-80 Hz) and the ability of the representation to return to V1after transient silencing. This model suggests that, *in vivo*, a different cortical area may store the representation of the visual stimulus, while V1 is silenced, and then feed that representation back into V1.

1 Introduction

Local cortical circuitry is highly recurrent. One often-proposed function of this recurrence is representation maintenance by recurrent excitation. Such a mechanism is believed to underlie, for instance, stimulus-specific neural activity observed in the prefrontal cortex after the removal of a stimulus during working memory [1]. However, it is true that highly recurrent connectivity is a feature of both prefrontal cortex and primary sensory cortices, such as primary visual cortex (V1) of the mouse [2]. Primary sensory cortical areas are not classically believed to play a role in working memory; however, the neural substrate for recurrence, and thus, perhaps, for representation maintenance, also exists in V1. The function of excitatory and inhibitory local circuits in V1 is not understood, but several roles may be proposed by theory [3]. First, it is possible these recurrent circuits exist simply to maintain an appropriate balance of excitation and inhibition in the network, or to regulate ratios of activity across cortical layers. Second, local recurrence may serve to temporally pattern neural activity (e.g., to organize spiking into widespread, oscillatory neural activity at a given frequency) but may have no hand in maintaining a specific representation after the stimulus has been removed, or in such attractor-like dynamics. Finally, it is possible that even V1, the lowest-level visual processing stage in neocortex, shows a

sort of working memory, or hysteresis, that is specific to past stimuli. This would imply that previously presented visual stimuli might influence the neural encoding of present stimuli, beyond simply determining instantaneous change in the visual world. It also implies certain stimuli may be temporally overrepresented in cortical encoding relative to less salient stimuli. There is some evidence that this is the case in V1 [4]. To test the role of local recurrence in such a function, we might examine the effects of transiently interrupting recurrence (transient silencing) on subsequent neural encoding. If recurrence (uninterrupted local activity) affects encoding, a change in encoding should be seen after the interruption of recurrence.

This "loss of recurrent function" approach is enabled by current optogenetic methods. Transiently silencing excitatory neuronal activity in an area of cortex will interrupt the locally recurrent excitation "ringing" in that area. The tool for this manipulation is channelrhodopsin (ChR), a blue light-gated ion channel that, when open, leads to neuronal depolarization and action potential generation, and which allows fine temporal and spatial control of neural spiking in cortex. Temporal control results from the rapid activation and deactivation kinetics of the channel, and spatial control results from the ability to express this genetically encoded channel in specific subpopulations of cells. For example, we can selectively express ChR in parvalbumin-positive (PV+) interneurons, which make up 95% of all cortical interneurons, by injecting a virus containing the ChR gene flanked by loxP sites into a transgenic mouse expressing Cre recombinase under the control of the parvalbumin promoter. In the described experiment, the virus infects all nearby neurons, but ChR is only expressed in cells that also express Cre (the PV+ cells), because Cre is necessary to reorient the open reading frame of ChR to permit its transcription. In this way, the cortical inhibitory cell population can be specifically controlled and caused to spike synchronously. Reason suggests, and experiments confirm, that such strong, synchronous activation of inhibition robustly silences the spiking of pyramidal neurons.

We use such a silencing approach to explore the role of cortical recurrence in sustaining the neural response to a brief, high contrast visual stimulus. In the anesthetized mouse, presentation of a visual full-field transition from dark to light produces a strong "ringing" response in V1, where "ringing" refers to the sustained network response specifically at gamma frequencies (30 to 80 Hz) for a second or more after the stimulus transition. This is a particularly long-lived response to an instantaneous stimulus, and we hypothesized that the later part of the response might be partially sustained by local excitatory recurrence. Interestingly, we find that interrupting local recurrence in vivo, or transient silencing of V1, a few hundred milliseconds after the visual transition does not eliminate the later part of the response or prevent the return of stimulus-specific activity in V1 following silencing. In fact, contradicting our hypothesis, it appears that interrupting local recurrent activity has little effect on V1 maintenance of the representation. Using a model, we explore in this paper the possibility that inter-areal cortical interactions may underlie this "representation return" phenomenon. Another cortical structure (e.g., V2) may maintain the representation of the stimulus while V1 is silenced before reactivating V1 once V1 is released from inhibition. We use a simple, four-dimensional linear rate model with two internally identical but differentially connected cortical areas to explore the dynamics of coupling and necessary parameter regimes to fit the experimental data. We go on to add non-linearities to the model to prevent negative firing rates and add N-methyl D-aspartate receptor (NMDAR) currents to reproduce the "representation return" result. The simple model reproduces many but not all of the in vivo effects and provides insights into possible mechanisms of cortico-cortical interaction and representation maintenance.

2 Methods

2.1 *In vivo* experiments

2.1.1 Expressing channelrhodopsin in parvalbumin-positive cortical interneurons

To express channelrhodopsin (ChR) specifically in parvalbumin-positive (PV+) cortical interneurons, we use a transgenic mouse line expressing Cre recombinase under the control of the PV promoter in a C57 Black 6 background. On post-natal day 0 or 1, each mouse pup receives



Figure 1: Methods. a) A mouse anesthetized with isoflurane and expressing ChR in PV+ cells of V1 is situated in front of a computer monitor presenting visual stimuli. A blinder has been placed on the animal's head to prevent the light from the blue LED from entering the eye and constituting a visual stimulus in its own right. The local field potential is recorded in V1, and the blue light can be turned on to activate the PV+ cortical cells. b) The basic organization of early cortical stages for visual processing. V1 and V2 each contain excitatory and inhibitory cell populations, and recurrent, feedforward, and feedback connections within and across areas. c) The author's model as analogous to the visual system in (b). Rate variables are assigned to each cell population, and all connections are given a weight. Inputs are denoted by I_E and I_I , the inputs to V1 excitatory and inhibitory populations, respectively. These inputs are not scaled by a weight.





viral injection (AAV, serotype 2/1) of a vector containing, crucially, a fluorescent reporter protein (TdTomato) and a floxed (flanked by loxP sites) ChR gene under the control of the strong and ubiquitous neuronal promoter EF1alpha. At P0 or P1, the skull is thin enough that a glass pipette containing virus may be advanced through it, into primary visual cortex. Virus is then pressure injected into the brain.

2.1.2 Electrophysiology

The pups recover, and, at one month of age, when the critical period for vision has ended, the mouse is taken for acute surgery and electrophysiology. The mice are anesthetized with isoflurane inhalant anesthesia and put into a stereotax for checking expression and surgery. The fluorescent protein reporter of viral expression may be visualized through the skull. Expression of ChR in V1 is thus verified, a craniotomy is made at this location, and a glass pipette is advanced about 100 microns into the brain for recording of the local field potential (Fig. 1a).

Visual stimuli are presented on a computer monitor situated in front of the contralateral eye (Fig. 1a). A screen transition from black to white is presented at a frame rate of 60 Hz, while data is recorded in V1. In some trials, a blue light-emitting diode (LED) situated over the craniotomy illuminates the brain and activates the ChR-expressing PV+ cells at some time after the screen transition. Data is collected and analyzed in Matlab.

2.2 Model

2.2.1 Linear model To explain the observed experimental data, a simple rate model of two coupled "cortical" areas was designed (Fig. 1b). Each area contained an excitatory and an inhibitory population. Within an area, self-recurrence existed from excitatory cells onto the same excitatory cells and from inhibitory cells onto themselves. Furthermore, local excitatory and inhibitory populations were reciprocally connected. Between the two areas, the only existing projections were from excitatory cells onto both excitatory and inhibitory populations in the other area, in keeping with cortical anatomy.

A linear rate model was derived from this structure. The model here has some similarities to the model described in Kang et al.. For the derivation of the rate model from a spiking model, see Kang et al.. Each distinct cell population is assigned a rate function variable (i.e., m(t), n(t), p(t), and q(t)), which represents the low-pass filtered (smoothed and continuous) firing rate of cells in that population. Each rate variable is influenced by the firing rate of presynaptic populations, which project with some weight (i.e., S, U or V) to the postsynaptic population, and the rate variables thus evolve in time according to the following differential equations:

$$\frac{dm}{dt} = \frac{1}{\tau_e} (-m + S_{EE} m - S_{EI} n + U_{EE} p + I_E)$$

$$\frac{dn}{dt} = \frac{1}{\tau_i} (-n + S_{IE} m - S_{II} n + U_{IE} p + I_I)$$

$$\frac{dp}{dt} = \frac{1}{\tau_e} (-p + S_{EE} p - S_{EI} q + V_{EE} m)$$

$$\frac{dq}{dt} = \frac{1}{\tau_i} (-q + S_{IE} p - S_{II} q + V_{IE} m)$$
(1.0)

where τ_e and τ_i are the synaptic time constants of excitation and inhibition, respectively. We choose 3 ms for τ_e based on the kinetics of the alpha-amino-3-hydroxyl-5-methyl-4-isoxazolepropionate (AMPA) receptor, a major receptor for glutamate, the main excitatory neurotransmitter in the mouse central nervous system, and we choose τ_i equals 6 ms, based on the longer kinetics of the gamma-amino-butyric acid A (GABA-A) receptor, which is the major receptor for inhibition in the mouse brain. m(t) and n(t) are the rate variables for the excitatory and inhibitory populations in V1, respectively; similarly, p(t) and q(t) are the rate variables for the excitatory and inhibitory populations in V2. V1 and V2 are assumed to be internally identical (this assumption may be supported, to some degree, by a similar anatomy of V1 and V2 in mice) but differentially connected. Within either V1 or V2, S_{EE} and S_{IE} are the weights of local synapses onto excitatory and inhibitory populations, respectively, from the excitatory population, while S_{EI} and S_{II} are the local synaptic weights onto the excitatory and inhibitory populations from the inhibitory cells. The coupling between the two areas is determined by the weights of feeedforward and feedback connections. U_{EE} and U_{IE} are the feedforward synaptic weights from V1 onto excitatory and inhibitory cells in V2. V_{EE} and V_{IE} are the feedback synaptic weights from V2 onto excitatory and inhibitory cells in V1. Finally, I_E and I_I represent the feedforward inputs from the lateral geniculate nucleus (LGN) to V1 and are the sole sources of input to the network.

2.2.1 Model with rectifying non-linearity

The linear model does not constrain firing rates to be positive, but of course there are no negative firing rates in the brain. For this reason, we decided to add a rectifying non-linearity to the model, to test the findings in a more realistic model. Rectification was applied to the presynaptic firing rate of a population when calculating the effects on a postsynaptic population's rate, that is, in each differential equation, the presynaptic firing rate is rectified (multiplied by 1 if it is positive and by 0 if it is not) before multiplying it by the weight of the connection. Note that the sum of all effects on the postsynaptic firing is not explicitly rectified. Simulation allows the examination of the behavior of the model under these conditions.

2.2.3 Model with rectifying non-linearity and NMDA receptor currents

In this section, we add a second excitatory current to the model, one with a greater synaptic time constant (slower channel kinetics), based on the N-methyl-D-aspartate (NMDA) receptor. This new current has a synaptic time constant of 60 ms and turns on only when the postsynaptic



Figure 2: *In vivo* data. **a**) A high contrast visual stimulus (screen transition from black to white) produces a long-lasting LFP response in V1. The gamma frequency of the "ringing" LFP decreases as the response decays. This response pattern can be seen both for single trials (top) and in an average (bottom). **b**) A brief LED pulse (red line below) transiently silences the "ringing" response following a stimulus. The black trace is the LFP response to a transition in screen color; the red trace shows the same response interrupted by silencing with the LED. Because in this layer, an LFP response is downward, the upward deflection in the red trace indicates silencing. **c**) Spectrograms for the responses in (b). Effects at right and left edges of the graphs can be ignored.

population is sufficiently active (depolarized), both conditions for the activation of the real NMDA receptor. An NMDA receptor-like current, *a*, is added to the excitatory cell populations in V1 and V2. The magnitude of this current is itself determined by a differential equation depending on the postsynaptic firing rate:

$$\frac{da}{dt} = \frac{1}{\tau_{e,NMDAR}} \left(-a + S_{NMDAR,E} * r(m) + U_{NMDAR,E} * r(m) \right)$$
(2.0)

where $\tau_{e,NMDAR}$ is the 60 ms synaptic time constant of the NMDA current and r(V) is activation as a function of voltage. Here, we use the terms "voltage" and "population activity" interchangeably, because the rate model is a low-pass filtered version of the spiking network model. Again, we examine the behavior of the model with respect to the experimental data.

3 Results

3.1 In vivo experiments

3.1.1 Expressing channelrhodopsin in parvalbumin-positive cortical interneurons

We are reliably able to express ChR at high levels in the PV+ interneurons in V1.

3.1.2 Electrophysiology

The local field potential (LFP) in V1 shows strong stimulus-evoked responses. For example, a simple, high contrast visual stimulus, the transition of the computer screen from black to white, elicits an early, strong, and sharp response in the LFP followed by a "ringing" of the LFP (Fig. 2a). This "ringing" is in the gamma frequency band (30-80 Hz), greatly outlasts the stimulus itself (can be seen for 1-2 seconds after the screen transition), and gradually decays in amplitude. This is interesting, because it has been classically proposed (and many studies support) that the visual system encodes the derivative of a visual stimulus rather than the absolute stimulus. The persistence of the representation is therefore a puzzle. Either continual feedforward drive from the lateral geniculate nucleus (the thalamic relay that receives information from the retina and sends it to cortex) or recurrence may explain the persistence. By interrupting recurrence, we ask: what is the contribution of cortical recurrence to the maintenance of the representation?

In an attempt to interrupt recurrence, we transiently silence V1. This is accomplished by activating the ChR-expressing PV+ cells, a manipulation we show to robustly silence cortical activity. A flash of blue light over the craniotomy causes synchronous action potentials in many PV+ cells, which inhibit cortical pyramidal neurons. Because pyramidal cells contribute most to the LFP (due to their asymmetric morphologies and aligned orientations with respect to the pial surface), strong activation of the inhibitory neurons and thus strong silencing of the excitatory pyramidal cells results in a silencing of the LFP signal.

When this silencing is preceded by a strong visual stimulus, does interrupting excitatory cortical recurrence in V1 eliminate the representation of the stimulus? In vivo experiments suggest that, surprisingly, the representation returns after silencing (Fig. 2b and 2c). This raises a question. Where is the representation stored while V1 pyramidal cell activity is missing? It may be that feedforward drive from the lateral geniculate nucleus (LGN) continues even into this post-silencing period and this is the source of the activity return. Alternatively, another structure may share recurrence with V1, and this structure may store the representation while V1 is silenced.

2.2 Model

2.2.1 Linear model

A model was built to test whether two recurrent, coupled cortical areas could recapitulate the experimental data. Specifically, would it be possible for one cortical area to store the representation by recurrent excitation while the other was silenced? Would it be possible for the continuously active cortical area to then feed that representation back into the silenced area once inhibition was removed?

The exact ratios of synaptic weights determine the system's behavior, and it is not straightforward to assign a specific function to each connection. However, certain connections can be seen to play a larger role in certain aspects of the response. I will describe features of the model's behavior and mention parameters that strongly modulate these features. Computer simulation with parameter



Figure 3: Linear model. **a**) The response of the linear model to an input pulse with matched excitation and inhibition. **b**) Under-, critically and over-damped regimes for model operation determined, largely by the strength of inhibitory self-recurrence. **c**) Increasing the strength of the inhibitory feedback loop to V1 increases the frequency of the network oscillation. **d**) Even a very large input to the inhibitory cells of V1 cannot silence the linear model, because positive and negative firing rates are symmetric in this model.



Figure 4: Model with rectifying non-linearity. **a**) No oscillations and no activity return after silencing. **b**) Unmatched excitatory and inhibitory inputs into the system rescue oscillations, but this may not be physiologically relevant.



Figure 5: Model with NMDAR currents. **a**) A pulse to the system with matched excitation and inhibition produces a ringing in the gamma frequency range. A subsequent silencing pulse to the inhibitory cell population in V1 (akin to the LED flash in experiments) transiently silences V1 but does not completely silence V2. Following this silencing, activity returns to V1, driven by the continued resonance in V2. **b**) Closeup of the silencing pulse in (a).

sweeps provides a number of insights in the system's behavior, as will be described below.

A first major result: the linear model easily reproduces the "ringing" response of the LFP (Fig. 3a). A brief step pulse to the linear model matched in excitatory (I_E) and inhibitory (I_I) drive kicks off an oscillation in the network that decays away to baseline after some period of time. The persistence of this ringing, or the damping time constant of the system, depends largely on the ratio of S_{EE} to S_{II} . As S_{II} increases, the behavior of the system transitions from under-damped to over-damped (Fig. 3b). The natural frequency of the system, or the frequency of the undriven response as it decays, is determined in large part by the strength of the inhibitory feedback connection relative to the other synaptic connections (Fig. 3c).

Connection strengths must be rather balanced in the model to prevent unbounded activity. For instance, if S_{II} or S_{EE} becomes too strong, the model enters a regime of unbounded oscillatory growth. This is also true for excitatory and inhibitory feedforward and feedback loops. Excitation and inhibition must be rather balanced to prevent runaway oscillatory activity.

One feature of the linear model that makes it non-ideal for comparison with the experimental data is an inability to silence the system by perturbing n, the firing rate of inhibitory cells in V1 (Fig. 3d). This perturbation matches, conceptually, the *in vivo* manipulation of PV+ cell activation. Because inhibition and excitation are symmetric in the model (negative firing rates are valid), perturbation of either population acts as an input to the network and elicits oscillatory activity.

2.2.2 Model with rectifying non-linearity

Adding a rectification non-linearity to prevent negative firing rates eliminates the symmetry of excitation and inhibition. Thus, in the rectified, non-linear model, silencing is possible by perturbation of n (Fig. 4a). Strong, simultaneous activation of many interneurons in cortex can also be thought of, conceptually, as a strong pulse input to inhibitory cells exclusively. This is the formulation used in the model to perturb n with respect to m, p, and q. The inhibitory pulse increases n and decreases m, as would be expected based on the analogous experimental data.

Interestingly, however, a tendency for oscillation is also somewhat disrupted in the non-linear model. While the linear model can ring for some time without continual drive to the network, the non-linear model damps faster without input, because entry into what would be a negative firing rate regime shuts down all spiking and returns the network to its baseline, inactive state (Fig. 4a). With zero drive, the non-linear, rectified network tends to oscillate around zero; thus, the response to a pulse input damps rapidly. Increasing excitatory recurrence can lengthen the duration of the

response to a pulse input, but if excitatory recurrence is increased beyond a certain point, the system becomes unstable.

Based on these results, it follows that non-matched excitatory and inhibitory inputs, or rather excitatory drive that sufficiently exceeds the inhibitory drive, is able to drive oscillations in the network. These oscillations as a result of unmatched excitation and inhibition decay away as the difference between excitatory and inhibitory inputs decreases (Fig. 4b).

When parameters are tuned such that the response outlasts the stimulus (i.e., response lasts about 1 second, in keeping with the experimental data), a pulse to the V1 inhibitory cell population is able to silence V1 activity. (The pulse also decreases but does not silence V2 activity.) Notably, the response does not return in V1 after silencing, and parameter sweeps suggest that no reasonable configuration of synaptic weight parameters in this set-up, with these synaptic time constants (τ_e = 3 ms and τ_i = 6 ms), will give "representation return" after silencing. The logic behind these results is: because the longer synaptic time constant for inhibition ensures inhibition always lags excitation, and because the fairly balanced excitatory and inhibitory synaptic weights (as dictated by the electrophysiology) ensure that excitation and inhibition in the network are closely coupled, the shutting down of excitation will always be followed at a short delay by the shutting down of inhibition. Furthermore, feedback excitation will be lagged by feedback inhibition. In this system, the lagging inhibition will prevent the regeneration of excitation after silencing, unless the system is itself unstable, showing oscillations that grow without bound.

2.2.3 Model with rectifying non-linearity and NMDA receptor currents

NMDA currents are much slower than AMPA currents; thus, they may provide the necessary stored excitation in the network to explain the "representation return" phenomenon in the experimental data. In fact, adding these NMDA receptor currents has a number of interesting effects on the network (Fig. 5a and 5b). First, matched excitation and inhibition can now drive gamma-frequency oscillations in V1 and V2. Second, a pulse to the V1 inhibitory neurons accomplishes robust silencing, but V2 is able to feed a stimulus representation back into V1 after silencing. In the essence of these details, the model exactly matches experimental data.

2.2.4 Differences between the best model and experimental data

Some differences exist between the simple model and the experimental data. Of course, the model captures only a tiny fraction of the complexity of the real brain. The model does not show activity in multiple frequency bands simultaneously, as does the real cortex, nor does the model's "ringing" response to an instantaneous stimulus decrease in frequency as it decays, as in the real cortex.

4 Discussion

The final model, with a rectifying non-linearity preventing negative firing rates and the addition of NMDA receptor currents, is able to demonstrate many of the features of the in vivo physiology. The model is able to reproduce gamma frequency band "ringing," the persistence of a response that far outlasts the duration of the stimulus, silencing by the activation of many inhibitory neurons, and representation return following silencing (once V1 has been relieved from inhibition). Interestingly, the longer synaptic time constant of the NMDA receptor currents transitions the network from an over-damped or critically damped state to a state that can show prolonged ringing based on excitatory and inhibitory recurrence, and the NMDA receptor currents allow V2 to feed a representation back into V1 after V1 silencing.

Interestingly, recurrence in this model depends on the presence of inhibition. In a purely excitatory model, the system is unstable or does not oscillate. The important role of inhibition for recurrence is also seen in the fact that the inhibitory feedback loop to V1 is what largely decides the frequency of the persistent response to a transient stimulus. Increasing local excitation or feedback excitation tends to decrease this "resonant" frequency, because the strength of inhibitory feedback is lessened relative to these excitatory connections. If either the excitatory or inhibitory loops

become too strong, the system becomes unstable, and in fact this stability requirement constrains the excitatory and inhibitory synaptic weights to a relatively balanced regime. This matches the physiological reality – in the cortex, and particularly in feedforward projections to the cortex, excitation and inhibition are generally relatively balanced.

In conclusion, this very simple model can explain some phenomena of the physiological data. Whether or not this model has anything to do with the real mechanisms in the brain is a question that I shall not attempt to answer. However, it is interesting that even a simple model of two recurrent, coupled cortical areas can recapitulate the experimental data, suggesting the mechanisms at work in the brain may have something akin to the fundamental behaviors of coupled oscillators with balanced excitation and inhibition.

Acknowledgements

I would like to acknowledge my advisor, Massimo Scanziani, for his support with the science, and the students, teachers, and assistants of Neurodynamics for their help with the concepts.

References

[1] Pasupathy and Miller EK. (2005) Different time courses of learning-related activity in the prefrontal cortex and striatum. Nature 433:873-876.

[2] Wang Q and Burkhalter A. (2007) Area map of mouse visual cortex. J Comp Neurol 502(3):339-57.

[3] Curtis CE and Lee D. (2010) Beyond working memory: the role of persistent activity in decision making. Trends Cogn Sci 14(5):216-22.

[4] Benucci A, Ringach DL and Carandini M. (2009) Coding of stimulus sequences by population responses in visual cortex. Nat Neurosci 12(10):1317-24.