Recurrent Neural Network Generates a Basis for Sensory Image Cancellation

Patrick D. Roberts

Neurological Sciences Institute
Oregon Health Sciences University
Portland, OR 97209
robertpa@ohsu.edu

Abstract

This study investigates the temporal dynamics of recurrent layers and their relevance to storage of temporal information. A recurrent layer is shown to generate a dynamical basis that allows cancellation of predictable sensory images via and adaptive mechanism based on spike-timing dependent plasticity.

Key words: recurrence, STDP, mormyrid, electrosensory

1 Introduction

Recurrent connectivity is ubiquitous in the nervous system, and many sensory processing structures adapt to the temporal pattern of repeated sensory signals to improve the detection of novel input. However, the effects of recurrence in neural networks are difficult to analyze because of their complex dynamics [1] and exhibit multiple dynamical states such as *coherent* and *asynchronous* states [2,3]. Methods for analyzing recurrent neural networks (RNN) have often relied on mean field results [4], but a significant generalization has recently been suggested [5,6]. Recent work that has focussed on the asynchronous state have led to the development of Liquid State Machines [5] or echo state networks [6]. In these networks, the strengths of recurrent connections are small so that oscillations do not develop and reverberations caused by the internal dynamics eventually decay. Although these types of recurrent neural networks are usually applied to problems of short-term memory, we show here that the reverberations of activity generates a "dynamical basis" that can be used as a reference signal to sculpt a long-term sensory image by synaptic plasticity.

Sensory image cancellation of mormyrid electric fish. The primary electrosensory processing system of mormyrid electric fish cancels predictable

sensory images to increase its sensitivity to novel stimuli. The mormyrid electric fish senses its environment by emitting an electric organ discharge (EOD) and detecting the perturbations that nearby objects cause in the self-generated electric field. The fish also detects low-frequency, externally generated electric fields. However, the electroreceptors that are sensitive to external signals also respond to the fish's own EOD so that the EOD interferes with the detection of external sources. Since the fish generates the motor command for each EOD, the reafferent sensory signal is predictable. Mormyrids have developed a mechanism to cancel predictable electrosensory signals so that the fish is exquisitely sensitive to novel electrical activity in its environment.

Electrosensory informations conveyed by afferents to the cortex of electrosensory lateral line lobe (ELL). The ELL also receives signals indicating *when* an EOD has taken place, the electric organ *corollary discharge* (CD) signals. These corollary discharge signals are time-locked with the EOD motor command which elicits the EOD. A structure similar to the cerebellar granule cell layer in mammals, the eminentia granularis posterior (EGp), receives corollary discharge signals at various delays following the EOD to at least 80 ms [7]. Granule cells in EGp give rise to the parallel fibers of ELL (Fig. 1A) that synapse onto Purkinje-like cells (PC) of the ELL. PCs also receive electrosensory input from the periphery. Pairing the corollary discharge with an electrosensory stimulus for a few minutes results in the elimination of the PC's response to the sensory stimulus [8]. Anti-Hebbian plasticity at the synapses from parallel fibers onto PCs has been demonstrated in the *in vitro* slice preparation [9], and such plasticity has been shown to explain this cancelation of predictable sensory images as observed *in vivo* [10].

Spike-timing dependent synaptic plasticity (STDP). The plasticity found at the parallel fiber synapses onto PCs [9] is STDP [11] so that the size and direction of synaptic changes depend on the exact timing between pre- and postsynaptic spikes. PCs have 2 types of spikes: (1) small, narrow axon spikes that do not invade the soma; and (2) broad, soma-dendritic spikes that propagate into the dendrites. The threshold of the broad (dendritic) spike is higher than the threshold of the small spike [12]. The dendritic spikes participates in the STDP learning rule (Fig. 1)B at the parallel fiber synapse [8,9]; the synapse is depressed only after pairings in which the dendritic spike was evoked between 0 and 60 msec after the onset of the parallel fiber postsynaptic potential, while pairings at all other delays yield potentiation. Modeling studies have helped elucidate how the learning rule measured *in vitro* can account for the plasticity of ELL cells measured *in vivo* at the systems level [10,13]. Not only does the model generate adaptive responses that are similar to those observed *in vivo*, but the STDP learning rule measured *in vitro* appears to be optimal for cancelling sensory images.

However, a critical component of the model is that the the parallel fibers must

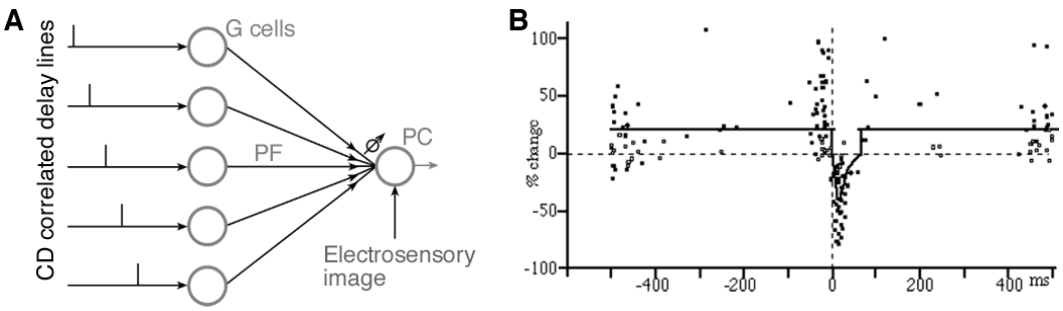

Fig. 1. *Local circuitry of the electrosensory lateral line lobe (ELL).* (A) The electric organ corollary discharge (CD) signals enter the ELL through the EGp that contains granule cells (G cells) giving rise to parallel fibers responding as tapped delay-lines following the CD. Parallel fibers (PF) synapses excite Purkinje-like cells (PC). Primary afferent fibers from electrosensory receptors enter the ELL and transmit temporal sensory pattern to PC. STDP at the PF synapse onto PC causes the output to adapt to the temporal pattern of the electrosensory image. (B) Data from [9] showing STDP at the PF synapse onto PC. The change in excitatory postsynaptic potential amplitude depends on the delay between onset of the postsynaptic potential and broad spike peak. If a dendritic spike in a PC follows the arrival of a parallel fiber spike within 60 msec, then the synapse is depressed proportionally to the excitatory postsynaptic potential; otherwise the synapse is potentiated.

act as delay-lines with respect to the EOD. Otherwise, the necessary temporal activity needed for sensory image cancellation cannot be sculpted out of the parallel fiber spikes. In addition, the spike activity of granule cell in EGp have never been recorded *in vivo* due to the difficulty of recording from such small cells. Thus, it is not known if this critical component of the model is present in ELL. Since other cerebellum-like structures, and the cerebellum, have no known tapped delay-lines reporting behaviorally important cues, the results of ELL function would be difficult to generalize. On the other hand, recurrence in the granule cell layer due to Golgi cell inhibition could generate temporal information about sensory stimuli based on principles of echo state networks [6]. In the following, we will extend our delay-line model to investigate how recurrence in a granule cell layer like EGp could generate a dynamical basis of spike patterns in parallel fibers that allows for sensory image cancellation.

2 Results

To test our hypothesis that recurrent connections in the layer of granule cells of ELL could lead electrosensory image cancellation, we conducted numerical studies of the network shown in Fig. 2A. The network consisted of a granule cell layer with synaptic connections to the PC. Each model neuron was spiking, but the neurons of the recurrent layer interacted through continuous synapses, synapses that transmitted the spike probability weighted by the recurrent synaptic weight matrix, $\mathbf{W}$. The spike probability was calculated from the membrane potential, $V_i(x_n, t)$, that was the sum of all inputs, and x_n denoted

the time within each cycle and t denoted the cycle. The spike-probability function was defined as [10,14] $P_i(x_n, t) = (1 + \exp(-\mu(V_i(x_n, t) - \theta)))^{-1}$, where θ is the threshold and μ parametrizes the noise. This function has a sigmoid form and a value of 1/2 at $\bar{V}(x_n, t) = \theta$.

In our simulations, each spiking granule cell was connected to the PC by a plastic synapse. The PC generated spikes according to its spike-probability function. At the beginning of each cycle, a subset of granule cells received a depolarizing input to represent a postsynaptic potential that initiated the recurrent dynamics. The 100 granule cells were interconnected by weights defined by an orthogonal random weight matrix [15] with eigenvalues < 1 [6]. These weights were not varied during adaptation and each synapse contributed to the membrane potential of the postsynaptic granule cell proportionally to the presynaptic spike probability.

Since the fixed-point in the learning dynamics determines whether any correlated, non-adapting sensory input will be cancelled, we need only test whether the response to the command is constant throughout the duration of the granule cell input. The fixed point of the STDP learning rule can be determined by expanding the granule cell spike-probability function, $\vec{P}(x_n, t)$, in the limit of small synaptic strength [14], $\vec{P}(x_n, t) = \sum_{K=0}^{\infty} (\mu/4)^K \mathbf{W}^K \cdot \epsilon^{(K)} * \vec{p_0}(x_n)$, where $\vec{p_0}(x_n)$ is the granule cell spike-probability without recurrent connectivity, $\epsilon^{(K)}$ is the K^{th} self-convolution of the postsynaptic potential caused by an input spike. Let $F(x_n)$ be the PC membrane potential cause by the parallel fiber inputs, $\vec{w}(t) \cdot \vec{P}(x_n, t) = F(x_n)$. Then, our expression for $\vec{P}(x_n, t)$ allows us to solve for the final weight configuration by taking the pseudo-inverse of $\vec{P}(x_n, t)$: $\vec{w}(t \to \infty) = F(x_n) \cdot \vec{P}(x_n, t)^{-1}$.

The results of simulations confirm the analysis that for the period following the initial rise in membrane potential ($x_n > 20$ ms), the membrane potential is flattened for the next 80 ms by the STDP learning rule (Fig. 2B). The granule cells exhibit a complicated set of membrane potentials that translate into parallel fiber spikes by the spike-probability function. The set of weights given by the PC spike-probability is the same as the result of STDP. We have also shown that the resulting membrane potential of the network shown in Fig. 1 for delayed input-spikes to 100 ms is nearly equivalent to the RNN result. The results of STDP agree confirming that the recurrent network effectively mimics tapped delay-lines.

3 Conclusion

Our results allow us to generalize the principles of how the ELL uses timing information and STDP to sculpt a negative image of predictable sensory inputs. Our previous requirement for delay-lines can now be relaxed, and any

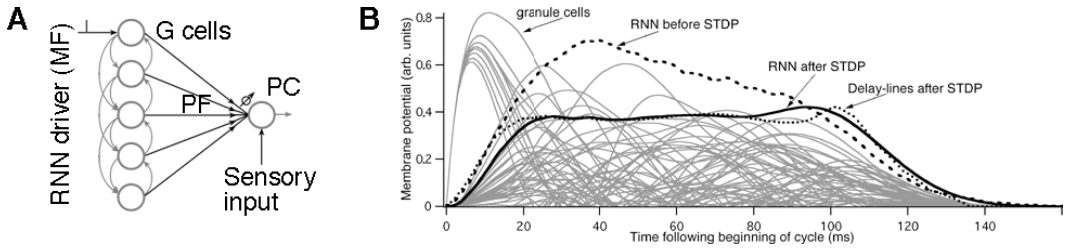

Fig. 2. *Adaptation of RNN output without delay-line input.* (A) RNN model of ELL where there are recurrent connections between the 100 granule cells (G). Ten mossy fiber inputs (MF) are synchronized so that there are no delay lines, all RNN driver inputs are synchronized. (B) The membrane potential of 100 granule cells (fine, grey traces) and PC before (dashed) and after adaptation (solid, black trace) by STDP at parallel fiber (PF) synapses. The RNN model has the same fixed point as the tapped-delay line model (dotted trace) in Fig. 1A.

cerebellum-like structure with recurrence in the granule cell layer can cancel correlated temporal patterns. The important point is that, in the recurrent case, the parallel fibers do not need to individually respond as tapped delay-lines. There is a linear combination of parallel fiber signals, that can be trained by a fixed point of the STDP learning rule to cancel correlated input. The model predicts that if recurrence is blocked in the granule cell layer of cerebellum-like structures, then the timing aspects of the function will be severely degraded. Another important remark is that the details of the recurrent synaptic weights are not important for the result. The cancellation of the sensory image is robust under a wide range of synaptic weights within the recurrent layer. The weights do not even have to be fixed, as long as of their changes occur on a longer time-scale that the STDP learning rule at parallel fiber synapses onto the PC.

Acknowledgments

I would like to thank Roberto Santiago for introducing me to echo state networks. This material is based upon work supported by the National Institutes of Health under Grant No. R01-MH60364 and by the National Science Foundation under Grant No. PHY99-07949.

References

[1] N. Brunel, Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons, J. Comput. Neurosci. 8 (2000) 183–208.

[2] L. F. Abbott, C. van Vreeswijk, Asynchronous states in networks of pulse-coupled oscillators, Phys. Rev. E 48 (1993) 14831490.

[3] C. van Vreeswijk, Analysis of the asynchronous state in networks of strongly coupled oscillators, Phys. Rev. Lett. 84 (2000) 5110–5113.

[4] A. Treves, Mean-field analysis of neuronal spike dynamics, Network 4 (1993) 259–284.

[5] W. Maass, T. Natschlager, H. Markram, Real-time computing without stable states: a new framework for neural computation based on perturbations, Neural Comput 14 (2002) 2531–2560.

[6] H. Jaeger, The "echo state" approach to analysing and training recurrent neural networks, gMD Report 148, German National Research Center for Information Technology (2001).

[7] C. C. Bell, K. Grant, J. Serrier, Sensory processing and corollary discharge effects in the mormyromast regions of the mormyrid electrosensory lobe: I. Field potentials and cellular activity in associated structures, J. Neurophysiol. 68 (1992) 843–858.

[8] C. C. Bell, A. Caputi, K. Grant, J. Serrier, Storage of a sensory pattern by anti-Hebbian synaptic plasticity in an electric fish, Proc. Natl. Acad. Sci. USA 90 (1993) 4650–4654.

[9] C. C. Bell, V. Han, Y. Sugawara, K. Grant, Synaptic plasticity in a cerebellum-like structure depends on temporal order, Nature 387 (1997) 278–281.

[10] P. D. Roberts, C. C. Bell, Computational consequences of temporally asymmetric learning rules: II. Sensory image cancellation, J. Compu. Neurosci. 9 (2000) 67–83.

[11] L. F. Abbott, S. B. Nelson, Synaptic plasticity: taming the beast, Nature Neurosci. (suppl.) 3 (2000) 1178–1183.

[12] K. Grant, Y. Sugawara, L. Gomes, V. Han, C. C. Bell, The Mormyrid electrosensory lobe *in vitro*: Physiology and pharmacology of cells and circuits, J. Neurosci. 18 (1998) 6009–6025.

[13] A. Williams, P. D. Roberts, T. K. Leen, Stability of negative-image equilibria in spike-timing-dependent plasticity, Phys Rev E Stat Nonlin Soft Matter Phys 68 (2-1) (2003) 021923.

[14] P. D. Roberts, Recurrent biological neural networks: The weak and noisy limit, Phys. Rev. E 69 (2004) 031910.

[15] O. L. White, D. D. Lee, H. Sompolinsky, Short-term memory in orthogonal neural networks, Phys. Rev. Lett. 92 (2004) 148102.