Unsupervised Formation of Vocalization-Sensitive Neurons: A Cortical Model Based on Short-Term and Homeostatic Plasticity

Tyler P. Lee
tylerlee@berkeley.edu
Helen Wills Neuroscience Institute, University of California, Berkeley,
CA 94720 U.S.A., and Departments of Neurobiology and Psychology,
and Brain Research Institute, University of California, Los Angeles,
CA 90095, U.S.A.

Dean V. Buonomano
dbuono@ucla.edu
Departments of Neurobiology and Psychology and Brain Research Institute,
University of California, Los Angeles, CA 90095, U.S.A.

The discrimination of complex auditory stimuli relies on the spatiotemporal structure of spike patterns arriving in the cortex. While recordings from auditory areas reveal that many neurons are highly selective to specific spatiotemporal stimuli, the mechanisms underlying this selectivity are unknown. Using computer simulations, we show that selectivity can emerge in neurons in an entirely unsupervised manner. The model is based on recurrently connected spiking neurons and synapses that exhibit short-term synaptic plasticity. During a developmental stage, spoken digits were presented to the network; the only type of long-term plasticity present was a form of homeostatic synaptic plasticity. From an initially unresponsive state, training generated a high percentage of neurons that responded selectively to individual digits. Furthermore, units within the network exhibited a cardinal feature of vocalization-sensitive neurons in vivo: differential responses between forward and reverse stimulus presentations. Direction selectivity deteriorated significantly, however, if short-term synaptic plasticity was removed. These results establish that a simple form of homeostatic plasticity is capable of guiding recurrent networks into regimes in which complex stimuli can be discriminated. In addition, one computational function of short-term synaptic plasticity may be to provide an inherent temporal asymmetry, thus contributing to the characteristic forward-reverse selectivity.

1 Introduction

Sensory stimuli are defined by the spatiotemporal pattern of action potentials they elicit at the level of our sensory organs. Based on these spatiotemporal patterns, the central nervous system makes sense of the external world. Some stimuli, such as a vertical or horizontal line or tones of different pitches, are characterized primarily by their spatial signature—that is, by which sensory afferents are active. Other stimuli, such as a 100 or 200 ms tone, must be discriminated based on their temporal signature. Many stimuli, perhaps most, are defined by their spatial and temporal features. Speech, for example, is rich in spatiotemporal structure, and degrading either the spatial or temporal features impairs speech recognition (Drullman, 1995; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995).

Recordings in the auditory system of birds (Margoliash, 1983; Doupe, 1997; Sen, Theunissen, & Doupe, 2001; Gentner & Margoliash, 2003; Prather, Nowicki, Anderson, Peters, & Mooney, 2009) and of mammals (Kilgard & Merzenich, 2002; Engineer et al., 2008; Yin, Mishkin, Sutter, & Fritz, 2008; Sadagopan & Wang, 2009) have revealed neurons that respond selectively to complex spatiotemporal stimuli. It is generally assumed that these neurons contribute to vocalization and speech recognition. The mechanisms responsible for the selectivity to spatiotemporal stimuli, however, remain unknown. Indeed, the neural mechanisms underlying the discrimination of even simple intervals remain a mystery (Mauk & Buonomano, 2004; Ivry & Schlerf, 2008). This stands in contrast to the mechanisms underlying simple forms of spatial selectivity, such as pitch or orientation tuning (Bienenstock, Cooper, & Munro, 1982; Ferster & Miller, 2000; Song & Abbott, 2001).

A number of models have proposed that the sensitivity to the temporal features of stimuli may rely on specialized neural mechanisms, such as delay lines, or ad hoc circuitry in which the timing of inhibition is tuned to generate interval- and order-selective neurons (Moore, Desmond, & Berthier, 1989; Lewicki & Konishi, 1995; Saitoh & Suga, 1995; Fiala et al., 1996; Drew & Abbott, 2003; Aubie, Becker, & Faure, 2009; Razak & Fuzessery, 2009). It has also been proposed that recurrent neural networks are intrinsically capable of temporal processing as a result of the inherently time-varying nature of neuronal properties and network dynamics. Specifically, in the same manner that the ripples produced by two raindrops falling on a pond contain a signature of which fell first and the interval between them, the dynamics of neural networks might naturally encode the recent temporal history of stimuli. Selectivity arises from the inherent complexity and nonlinearity of cortical architecture and the interaction between incoming information and the internal state of the network. This framework has been referred to as a state-dependent network (Buonomano & Merzenich, 1995; Buonomano, 2000) or, in the context of machine learning, a liquid-state machine (Maass, Natschläger, & Markram, 2002, 2003) or reservoir

computing. Useful to understanding this framework is the realization that at any given point in time, the internal state of a network is characterized by two components (Buonomano & Maass, 2009): the active state, defined by which neurons are active (if any), and the hidden state, which captures the time-dependent changes in synaptic and cellular properties such as short-term synaptic plasticity (STP). STP refers to the ubiquitous use-dependent changes in synaptic strength that occur over a timescale of tens to hundreds of milliseconds (Zucker, 1989; Abbott & Regehr, 2004).

A weakness of this framework, to date, however, has been that while the weights and connectivity are random, the average values still have to be fairly finely tuned. Additionally, few of these models have incorporated synaptic plasticity within the recurrent connections. Specifically, due to the highly nonlinear dynamics of recurrent cortical circuitry, it has proven challenging to incorporate synaptic plasticity rules that guide these networks to appropriate regimes (Doya, 1992; Pearlmutter, 1995). Here we show that when the state-dependent network framework is coupled with a homeostatic plasticity rule, units in the network develop selective responses to spatiotemporal stimuli in an unsupervised fashion. Specifically, after exposing the network to spoken digits, many units became selective to these digits. This provides one of the first models demonstrating how vocalization-selective neurons can emerge in complex recurrent networks in an unsupervised fashion. Additionally, we explicitly examined the importance of short-term synaptic plasticity in the formation of digit-selective units and demonstrate that it significantly enhances the discrimination of forward versus reverse digits. In other words, short-term plasticity may contribute to the brain's sensitivity to temporal direction that is responsible for the unmistakable perceptual differences between forward and reverse vocalizations.

2 Methods

2.1 Data Set. We used a benchmark data set of 100 spoken digits: 10 utterances of each of the digits 0 through 9, from speaker 1 of the TI46 database. Starting with the stimulus waveforms, we used the Auditory Toolbox (Malcolm Slaney) and the LyonPassiveEar command to generate a cochleagram that was used to drive 18 integrate-and-fire units representing cochlear nerve fibers (with a maximum frequency of 5 kHz). The spiking responses of these units served as the input to our cortical network.

2.2 Cortical Network. All cortical network simulations were performed using the NEURON simulation package (Hines & Carnevale, 1997). The cortical network was designed to represent a simplified cortical column; 18 input neurons (IN) relayed information from the auditory periphery and synapsed onto L-IV neurons. L-IV neurons both received and sent projections to a population of L-II/III neurons. In total, the column was composed

of 400 excitatory (Ex) and 100 inhibitory (Inh) integrate-and-fire neurons. This ratio of Ex to Inh cell counts approximates what has been estimated throughout cortex, including primary auditory cortex (Prieto, Peterson, & Winer, 1994). Neurons were modeled as cylinders with both a length and diameter of 10 μm. Ex and Inh cells had a resting membrane potential of -60 mV each, a membrane time constant of 30 and 10 ms, and an action potential threshold drawn from a normal distribution $N(-40\,\text{mV}, 2\,\text{mV})$ and $N(-45\,\text{mV}, 2.25\,\text{mV})$, respectively. Each neuron possessed four sources of current: a leakage current, an afterhyperpolarization (AHP) current, synaptic currents, and a noise current. The leak conductance for all cells was set at 0.1 mS/cm^2 with a reversal potential of -60 mV, providing an input resistance of 3.18 GΩ. When the action potential threshold was reached, the cell's voltage was set to 40 mV for 1 ms and then reset to -60 and -65 mV for Ex and Inh units, respectively. Upon reset, the AHP was incremented by .07 and .2 mS/cm^2 and then decayed with a time constant of 10 and 2 ms for Ex and Inh units, respectively. The AHP reversal potential was set at -90 mV. A noise current pulled from a uniform distribution between $-.5$ and .5 pA was injected into each cell at each 0.1 ms time step.

The network was initialized with random connectivity, with connection probabilities of 12.5% for Ex→Ex, 10% for Inh→Ex, and 2.5% for Ex→Inh synapses. Each input neuron projected to 12 Ex and 3 Inh neurons. All synapses were simulated using a kinetic model (Destexhe, Mainen, & Sejnowski, 1994; Buonomano, 2000). IN→Ex and IN→Inh synaptic efficacies were fixed at 2 nS and 1.67 nS, respectively, with synaptic delays of 0.8 ms. The input thus strongly drove activity in L-IV neurons. All cortical excitatory synapses contained plastic AMPA ($E_{AMPA} = 0$ mV) and NMDA ($E_{NMDA} = 50$ mV) components. NMDA receptor conductances were given a weight of 0.3 times that synapse's AMPA conductance. The weights of the inhibitory GABAergic synapses ($E_{GABA} = -70$ mV) were fixed with a mean of 20 nS (except for Figure S2, where this parameter was systematically varied). (All figure numbers preceded by "S" are in the online supplement.) AMPA Ex→Ex and Ex→Inh weights were initialized with a mean of .02 nS and .1 nS, respectively. To prevent any individual synapses from causing a postsynaptic neuron to fire, Ex→Ex and Ex→Inh weights were bounded from above at 2 nS and 5 nS, respectively. Synaptic delays were set at 1.4 ms for Ex→Ex, 0.8 ms for Ex→Inh, and 0.6 ms for Inh→Ex. All synapses exhibited STP modeled according to Markram, Wang, and Tsodyks (1998). In this model, synaptic efficacy is determined at each spike by the product of two variables, R and u. R, which has an initial value of 1, represents the amount of remaining synaptic efficacy. After the first action potential in a trial, R is

$$R_1 = 1 - U, \tag{2.1}$$

where U is a constant representing the synaptic efficacy utilized by the first action potential. For subsequent spikes, the remaining and utilized efficacy are characterized by two parameters, τ_{rec} and τ_{facil}, representing the time constant with which each decays to its initial value. Thus, the remaining synaptic efficacy following the $(n + 1)$th spike is governed by

$$R_{n+1} = R_n(1 - u_{n+1}) \exp\left(-\frac{\Delta t}{\tau_{rec}}\right) + 1 - \exp\left(-\frac{\Delta t}{\tau_{rec}}\right), \qquad (2.2)$$

where Δt is the interspike interval and u_{n+1} is the efficacy utilized by the $(n + 1)$th spike. u, which incorporates a facilitation term, is governed by

$$u_{n+1} = u_n \exp\left(-\frac{\Delta t}{\tau_{facil}}\right) + U\left(1 - u_n \exp\left(-\frac{\Delta t}{\tau_{facil}}\right)\right). \qquad (2.3)$$

In the simulation, excitatory synapses exhibited short-term facilitation, with $\tau_{facil} = 250$ (500) ms, $\tau_{rec} = 25$ (125) ms, and $U = .25$ (.2) for Ex→Ex (Ex→Inh) synapses. Inhibitory synapses exhibited short-term depression, determined by $\tau_{facil} = 20$ ms, $\tau_{rec} = 700$ ms, and $U = .25$. Finally, input synapses exhibited facilitation, characterized by $\tau_{facil} = 500$ ms, $\tau_{rec} = 125$ ms, and $U = .2$.

2.3 Training and Synaptic Plasticity. Training occurred across 1000 stimulus blocks, each block consisting of 10 trials (two utterances of five digits). Only 10 stimuli were used during each training block for reasons of computational load. The choice of the digits used during training significantly affected the ability of different digits to drive network activity and the forward selectivity (see section 4), but did not produce significant differences in the direction selectivity between the trained and untrained digits. Synaptic weights were updated at the end of each trial. After the completion of training, plasticity was turned off, and 10 more blocks were run using the entire stimulus set, including both forward and reverse digit presentations (for a total of 2000 trials).

In the initial state, the input projections to the L-IV units were strong and robustly drove activity in these units. However, the recurrent synapses between all Ex→Ex and Ex→Inh units were set to very low (but nonzero) values. Consequently, none of the L-II/III Ex units spiked during early stimulus presentations. The network was trained using presynaptic-dependent scaling (PSD). PSD, a modified form of synaptic scaling, is consistent with current experimental data, but in contrast to synaptic scaling, it generates stable dynamics within recurrent neural networks (Buonomano, 2005).

Synaptic weights were updated according to

$$W_{ij}^{\tau+1} = W_{ij}^{\tau} + \alpha_W A_i^{\tau} (A_{goal} - A_j^{\tau}) W_{ij}^{\tau}. \tag{2.4}$$

Here W_{ij}^{τ} is the synaptic weight from neuron i to j at trial τ; α_W represents the "learning" rate, set to 5×10^{-4}; and A_{goal} is the target activity for the postsynaptic neuron. Target activity levels were randomly set at .25, .5, or .75 for each neuron. By requiring activity values to average between 0 and 1 spikes per trial, the cortical network sparsely coded the input stimulus as well as refrained from runaway excitation, a common problem in recurrent neural networks (Buonomano, 2005). A_i^{τ} is a running average activity of the ith neuron in trial τ, calculated as

$$A_i^{\tau+1} = A_i^{\tau} + \alpha_A (S_i^{\tau} - A_i^{\tau}). \tag{2.5}$$

$\alpha_A = .04 \; (.2)$ controls the across-trial integration of activity for Ex (Inh) cells. S_i^{τ} is the number of spikes by neuron i in trial τ.

2.4 Quantification of Selectivity. In order to estimate the preference of a neuron for different digits, a selectivity index was defined as

$$SI = \frac{RS_{max} - \overline{RS}}{RS_{max} + \overline{RS}}, \tag{2.6}$$

where RS_{max} is the maximal mean response to any digits (averaged over all utterances), RS is the mean response to the remaining $N - 1$ digits. Values of SI are bounded between 0 and 1. To test for significant selectivity, we shuffled the responses of each neuron to all stimuli and calculated an $SI_{shuffled}$. This was done 1000 times, and the SI was considered significant if it was greater than 99% of all $SI_{shuffled}$ values.

In the same fashion, we calculated the direction selectivity of each cell for each digit by computing

$$SI_{dir} = \frac{RS_{For} - RS_{Rev}}{RS_{For} + RS_{Rev}}, \tag{2.7}$$

where RS_{For} and RS_{Rev} are the mean responses to the forward and reverse presentations of that digit, respectively. Significance was tested by comparing the absolute value of SI_{dir} and the 99% threshold derived from shuffling all responses to that digit.

2.5 Readout Units. In some experiments, we wanted to determine the ability of the population of all Ex neurons to discriminate the stimulus set. Toward this end, we added a layer of readout units, each of which received

an input from all of the excitatory units in the cortical network. The readout neurons were trained using a tempotron learning rule to discriminate the digits (Gutig & Sompolinsky, 2006). This is a supervised learning rule in which, if the postsynaptic neuron did not fire to the target ("positive") stimulus, the weights of the synapses whose activity contributed to the maximum voltage are increased in a manner proportional to that contribution. If the readout unit fires at any point during a "negative" stimulus, the synaptic weights of each synapse are decreased in proportion to their contribution to the incorrect spike. All weights were initialized to zero and allowed to range between .2 and $-.2\ \mu$S. Weights were increased with a learning rate of .001 when the output neuron's response was a false negative and decreased with a learning rate of .0001 when the response was a false positive. A population of five output neurons was assigned to each stimulus class (i.e., a digit). Upon spiking, each neuron evoked a lateral inhibitory current in all other readout neurons ($g_{LI} = .15\ \mu$S), encouraging the neurons in each population to extract distinct spatiotemporal patterns from the cortical network. Classification was determined based on the group with the highest total number of spikes. We trained the network using a leave-one-out cross-validation scheme, withholding one utterance of each digit to test for generalization. When training readouts to discriminate forward and reverse digits, we used 20 populations (10 digits $\times$ 2 directions) of 5 units (100 total readout units). We used the tempotron learning rule strictly to estimate the amount of spatiotemporal information present in the cortical network, but we emphasize that we do not view it as a mechanistically realistic supervised learning rule.

2.6 Removal of Short-Term Plasticity. In order to examine the effect of short-term plasticity on the spatiotemporal selectivity of the cortical network, we removed short-term plasticity by setting all values for τ_{facil} and τ_{rec} to zero. This alone, however, resulted in a net decrease in network activity because of the absence of short-term facilitation. To ensure that the network received the same average level of input, we calculated the average activity of all L-IV neurons over 200 stimulus presentations. We then varied the IN$\rightarrow$Ex weights in the network without STP until the activity matched that of the network with STP intact. The approach revealed that an increase in the "baseline" strength of IN$\rightarrow$Ex weights of 20% (to 2.4 nS) compensated for the absence of STP; thus this value was first used for all NoSTP simulations. Additionally, to compensate for the need for stronger synapses, we doubled the maximum Ex$\rightarrow$Ex and Ex$\rightarrow$Inh weights.

3 Results

The network architecture is schematized in Figure 1. It contained 400 excitatory and 100 inhibitory integrate-and-fire units. Input was provided
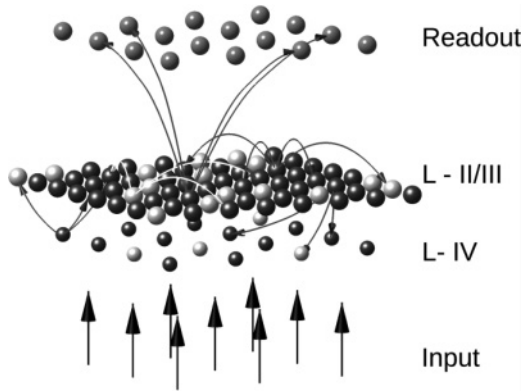
Figure 1: Network architecture. A layer of input channels feeds into a recurrent cortical network composed of excitatory (dark gray) and inhibitory (light gray) integrate-and-fire neurons. A subset of neurons (L-IV) is driven strongly by the input via synapses that do not undergo plasticity, whereas the remaining (L-II/III) are driven solely by the propagation of activity within the cortical network. Cortical excitatory neurons synapse onto both excitatory and inhibitory cells, while inhibitory neurons synapse only onto excitatory cells. All excitatory connections between the cortical cells were plastic. In some simulations, a layer of readout units was used to perform digit classification.

by 18 input fibers, each representing a frequency band. The stimulus set consisted of a benchmark of 10 (0–9) spoken digits (10 utterances of each). Spoken digits were converted into spatiotemporal patterns of spikes using a simulated cochleagram (see section 2). Since our focus is on early stages of sensory processing, all digits were from a single speaker. We do not expect our model to generate significant speaker invariance; however, it is noteworthy that even the different utterances from a single speaker vary significantly in their spatial and temporal features. Figure 2 illustrates the range of the temporal variability of utterances by showing the shortest and longest instances of each digit, as well as the duration spread (see Figure 2B).

In the initial state, all cortical excitatory (Ex→Ex; Ex→Inh) synapses were weak; thus, as illustrated in Figure 3, in the initial state only the L-IV cortical units, which received direct input, fired. This initial state loosely reflects early developmental stages in which most cortical neurons are weakly responsive or unresponsive to external stimulation and exhibit weak synapses (Hubel & Wiesel, 1963; Echevarria & Albus, 2000).

For a neuron to create robust representations of sensory stimuli, an initially unresponsive neuron must develop responses to some of the stimuli to which an animal is exposed. A number of synaptic plasticity rules are likely to contribute to this process, but since most of the units in the
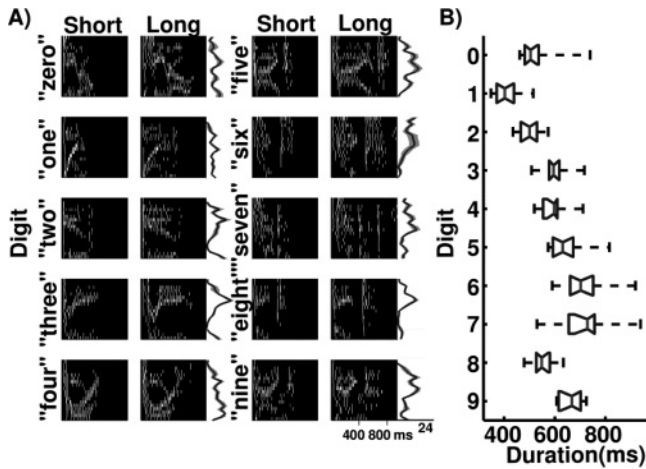
Figure 2: Variation in spectral and temporal properties of utterances. (A) Input spike rasters depicting the shortest and longest utterance for each digit (left). Rasters display action potentials in each input nerve fiber (*y*-axis) across time and have been padded with zeros to 939 ms to maintain a constant timescale for visualization. (Right) The thick line shows the average number of spikes from that channel, with the shading indicating one standard deviation from the mean. (B) Box-and-whiskers plot of utterance durations for each digit.

network are initially unresponsive, homeostatic forms of plasticity are obvious candidates (note that the absence of postsynaptic spikes in the initial state makes STDP ineffective in the early stages of processing; see section 4). One type of homeostatic plasticity rule, synaptic scaling, has been shown to be effective in bringing neurons embedded in feedforward networks to desired average levels of activity (van Rossum, Bi, & Turrigiano, 2000; Turrigiano & Nelson, 2004). However, theoretical work has suggested that when used in recurrent networks, synaptic scaling can produce unstable and "pathological" states (Buonomano, 2005; Houweling, Bazhenov, Timofeev, Steriade, & Sejnowski, 2005; Frohlich, Bazhenov, & Sejnowski, 2008; Liu & Buonomano, 2009). Here we used a modified version of synaptic scaling (see section 2), termed presynaptic-dependent (PSD) scaling, in which changes in synaptic weights take into account the average levels of activity in the presynaptic neurons (Buonomano, 2005).

For computational load reasons, only a subset of 10 stimuli (out of the 100) was presented in each training block (see section 2). At a learning rate, α_w, of 5×10^{-4} the mean level of activity (spikes per trial) generally converged toward the mean set point within 1000 training blocks (see Figure 3B). The speed of convergence, however could be dramatically accelerated by using a higher values of α_w (see Figure 3B, gray line). Before
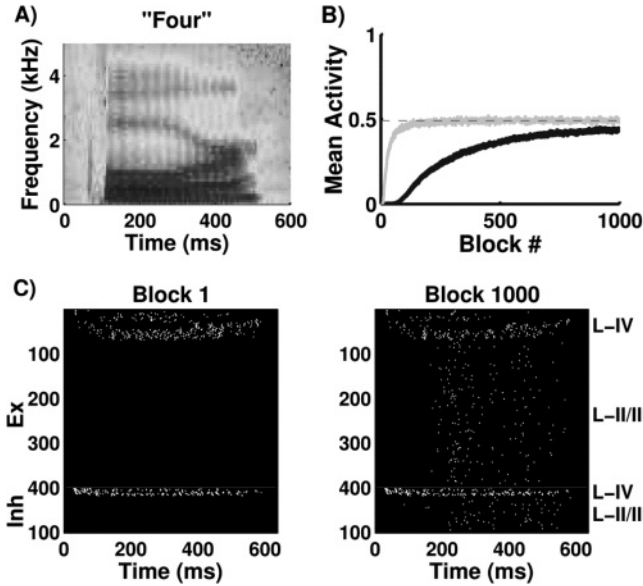
Figure 3: Training of the cortical network. (A) Spectrogram of one of the digit "four" utterances. (B) Convergence over training of the average activity to a predetermined target value (set-point), averaged across all L-II/III Ex cells. Synaptic weights were trained according to the presynaptic-dependent scaling rule with $\alpha_W = 5 \times 10^{-4}$. The light gray line represents training with $\alpha_W = .005$, showing faster, and still monotonic, convergence (even faster learning rates can still converge, but in a nonmonotonic fashion). (C) Raster of cortical activity prior (left) to and after (right) training. Notice that only L-IV neurons are initially active due to strong and fixed input synaptic weights.

training with PSD, only the subset of units that received direct input fired in response to a stimulus; after training, a significant number of Ex units responded to the presentation of digit 4, and together 96% of all the neurons responded to at least one of the stimuli used during training—indicating that the excitatory cortical connections were now driving many of the units and effectively spreading activity across the entire network (see Figure 3C).

3.1 Selectivity of the Cortical Units. After the completion of the 1000 training blocks, we presented each of the 10 utterances of the 10 digits a total of 10 times in order to characterize the selectivity of each unit. Since one of our major goals is to examine the emergence of neurons that are spatiotemporally selective, it is necessary to determine if any observed selectivity is simply attributable to the spatial (spectral) content rather than the spatiotemporal features of the spoken digits. A unit that responds selectively to
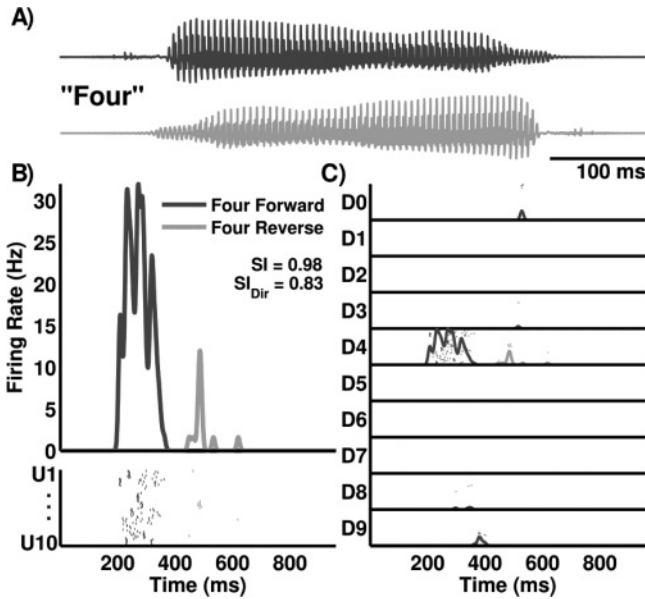
Figure 4: Example of an L-II/III unit with strong digit and direction selectivity. (A) Sonograms of the forward (top) and reverse (bottom) presentations of digit "four." (B) Poststimulus time histogram for a single neuron across all utterances of digit "four" (top). It depicts the average firing rate over time for this cell in response to the forward presentation (dark gray) and reverse presentation (light gray), smoothed with a 5 ms wide gaussian filter. Raster of responses by the same neuron across each trial and utterance (10 trials of each of the 10 utterances) for that digit (bottom). Dark pixels represent spikes in response to the forward presentation and light to the reverse presentation. (C) Raster plots showing the response of this unit to all stimulus presentations.

"four" may simply be a frequency-tuned neuron responding to a frequency present in "four" and may be totally insensitive to the temporal structure of the stimulus. The traditional experimental method to determine if a neuron is truly selective to spatial and temporal information is to contrast the responses to forward and reverse vocalizations (Wang, Merzenich, Beitel, & Schreiner, 1995; Doupe, 1997; Recanzone, 2008; Huetz, Philibert, & Edeline, 2009; Razak & Fuzessery, 2008). Thus we also presented all stimuli reversed in time.

We first characterized the receptive field characteristics of all the L-II/III Ex units by calculating two selectivity indexes (see section 2): one to the forward presentation of all stimuli (SI) and one based on the number of spikes elicited by each digit played forward and backward (SI_{Dir}).

Figure 4 illustrates the receptive fields of an L-II/III unit that was highly selective to the forward presentation of "four" compared to all other
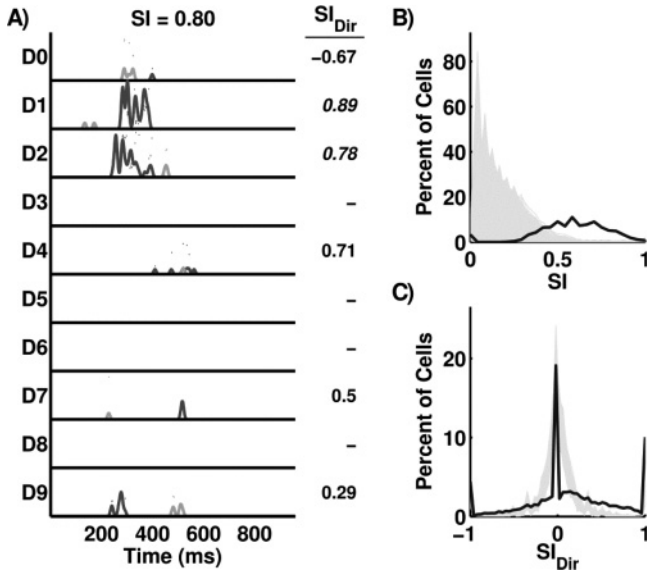
Figure 5: Distribution of selectivity within the cortical network. (A) As in Figure 4, the poststimulus time histogram for a different single neuron to all 10 digits. Dark gray traces and pixels indicate responses to forward stimuli and light gray to reverse. This cell responds strongly to "one" and "two" with very high direction selectivity (SI_{Dir} on right). Significant SI_{Dir} values are indicated in italics. (B) Distribution of SI for all cortical L-II/III cells. The distribution peaks at an SI of $\sim$.6, demonstrating that a large fraction of neurons in the network was digit selective. (C) Distribution of SI_{Dir} for cortical L-II/III Ex cells. Notice that the distribution has heavy tails with a large number of cells at each extreme. Values range from -1 to 1, indicating high reverse or forward direction selectivity, respectively. (B, C) Light gray lines represent distributions of $SI_{shuffled}$ values for all 4000 shuffles (1000 shuffles for each of four seeds).

forward digits (a high SI value). Note that even though different utterances of the same digit vary in duration and spatiotemporal structure, this unit responded reliably to all utterances of its preferred digit. This unit also fired selectively to the forward utterances of "four" compared to the reverse utterances (a high SI_{Dir}). This unit did, however, fire to the reverse presentation of one of the utterances of the digit "four." It is noteworthy that the selectivity to digits and direction emerged in a totally unsupervised manner. Figure 5A provides an example of a slightly less selective unit that fires to most of the utterances of two digits. Together 95% of the L-II/III Ex units, all of which were unresponsive before training, exhibited a significant SI value. The distributions of SI and SI_{Dir} of all L-II/III Ex units are shown in Figures 5B and 5C. Note the asymmetry in the distribution of the SI_{Dir}

values, indicating that units tend to prefer the forward compared to the reverse direction. This average preference for forward digits remained if the network was trained on reverse stimuli only; importantly, however, the forward selectivity was significantly less (see Figure S1, t-test, $p = .001$). The preference for forward stimuli is readily explained by the temporal asymmetry in speech itself, where strong broadband onsets are common, inducing greater firing, on average, to forward stimuli. However, the fact that reverse training decreased average forward selectivity indicates that the training stimuli per se are in part responsible for the final selectivity profile. In other words, the final network state can be said to be experience dependent.

To investigate the parameter sensitivity of the model, we parametrically varied the mean strength of Inh→Ex synapses. This parameter is critical because these synapses are not plastic in our implementation and because the balance of excitation and inhibition is critical to stimulus selectivity and network dynamics (Wilent & Contreras, 2005; Wu, Li, Tao, & Zhang, 2006; Froemke, Merzenich, & Schreiner, 2007; Carvalho & Buonomano, 2009). Figure S2 shows the average level of digit selectivity and direction selectivity among L II/III excitatory neurons in the network. Our results show robust levels of selectivity within the network across a six-fold range of mean Inh→Ex strengths (4 nS to 24 nS), with decreasing levels of digit selectivity and increasing levels of direction selectivity at higher levels of inhibition ($p = .0086$, $p = .014$, respectively, one-way ANOVA).

3.2 Role of Short-Term Plasticity in Selectivity. Spatiotemporal selectivity, by definition, requires that neurons be sensitive to the order or interval between different spectral features. In the state-dependent network model, this ability derives from the fact that the state of the network at any point in time t encodes the recent features of the incoming stimuli. As mentioned in section 1, the internal state is defined by the active and hidden state (Buonomano & Maass, 2009). In the current simulations, the hidden state is primarily constrained to the short-term changes in synaptic strength imposed by short-term synaptic plasticity, while the active state refers to any ongoing firing in the population. To attempt to dissect the contribution of the short-term synaptic plasticity to the spatiotemporal selectivity, we repeated the above simulations after removing short-term plasticity from all the synapses in the network (see section 2).

Short-term plasticity provides an inherent temporal asymmetry to the network; for example, the EPSP amplitude of the third pulse of the pattern 0-50-200 ms, will be different from that of the pattern 0-150-200 ms. Thus, we hypothesized that short-term plasticity may contribute primarily to direction selectivity. After we trained a network without any short-term plasticity, a large number of units continued to be selective for the forward digits, although the overall distribution of SI shifted significantly downward (see Figure 6A, KS-test, $p < 0.01$). More important, there was a
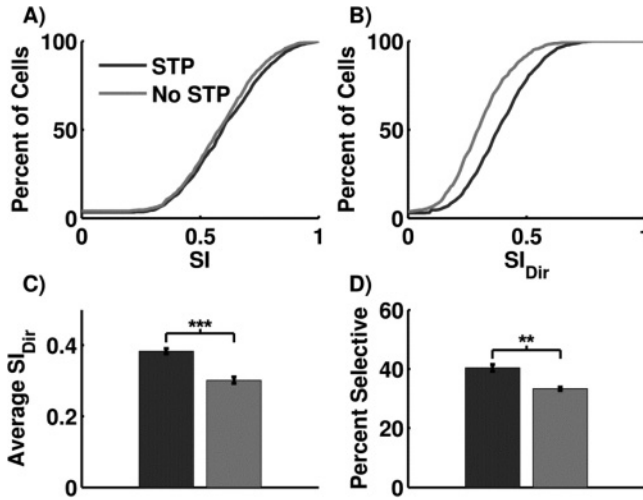
Figure 6: Comparison of selectivity between STP and no STP groups. (A) Cumulative distribution of selectivity indexes (SI) of all L-II/III cells for STP and no STP groups. Short-term plasticity enhances digit selectivity within the cortical network ($p = .006$, two-sample KS-test; subtle rightward position of the default network with STP). (B) To compare the overall levels of direction selectivity in the presence and absence of STP, we averaged the absolute value of each L II/III cell's SI_{Dir} across all digits. The cumulative distribution of these averaged values depicts greatly increased direction selectivity in cortical networks with STP ($p = 10^{-38}$, two-sample KS-test). (C) Comparison of the average SI_{Dir} values between the STP and no STP groups. The difference between groups is significant ($p = 3 \times 10^{-4}$, t-test). (D) Comparison of the number of direction-selective cells between groups. There are significantly more direction-selective neurons in the STP group ($p = .0016$, t-test). Error bars represent SEM across four simulations with different random seeds.

significant shift in the distribution of SI_{Dir} values toward decreased selectivity (see Figure 6B; KS test, $p = 10^{-38}$) with a corresponding significant decrease in the mean SI_{Dir} (t-test, $p < 0.001$) and the number of units with significant SI_{Dir} values (t-test, $p = 3 \times 10^{-4}$, $p = .0016$; see Figures 6C and 6D). This finding was robust across levels of inhibition within the network (see Figure S2).

3.3 Digit Discrimination. The above results demonstrate that a simple homeostatic plasticity rule can lead to the emergence of units that respond selectively to spoken digits across different utterances. But the fact that most cells exhibited significant SI and SI_{Dir} values does not mean that every digit was well represented in the network. Indeed, some digits were the preferred stimulus of only a few neurons in the network. But as previously proposed, the notion is that together, the varying degrees of selectivity serve
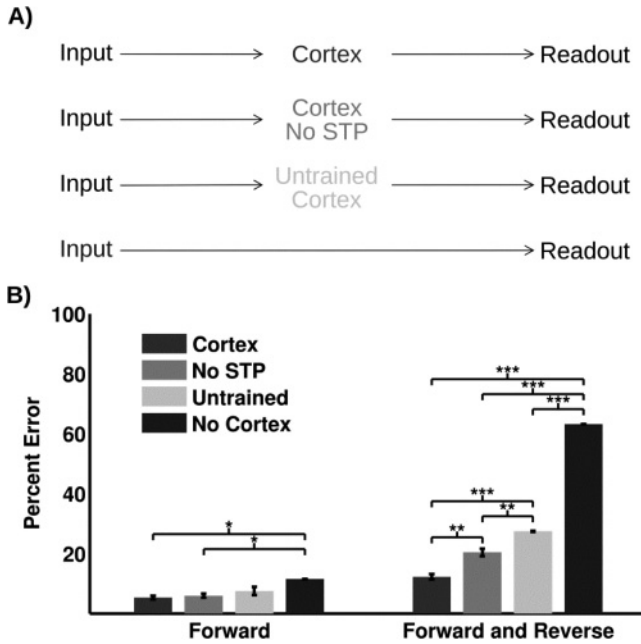
Figure 7: Performance on digit discrimination task. (A) Readout neurons were trained to discriminate each of the 10 digits using input from the trained cortical network (STP), the cortical network trained in the absence of short-term plasticity (no STP), the untrained cortical network (Untrained), and directly from the input channels (no cortex). (B) Percentage error in classification when discriminating the forward (left) and forward and reverse (right) digits across all groups. Error bars represent SEM across four simulations with different random seeds. The readout network performed similarly on the forward digit discriminations across all cortical groups (5.3–7.5% error). Performance on the no cortex group, however, was significantly worse than the cortex group (11.5% versus 5.3% error, $p = .016$, t-test). Performance on the forward and reverse discrimination task was best on the cortex group with an error rate of 12% (20%, $p = .0016$, 27%, $p = 2.4 \times 10^{-6}$, 63%, $p = 1.2 \times 10^{-4}$, t-test, from left to right).

as a basis for digit discrimination by neurons downstream from neurons in primary sensory cortex. So to examine the potential ability of downstream neurons to discriminate digits based on the pattern of activity in the cortical units, we used a supervised learning rule to train a layer of output units to discriminate among the 10 digits. Specifically we used the tempotron learning rule (Gutig & Sompolinsky, 2006), which provides the ability to process continuous time signals, to train sets of readout neurons (see section 2).

As controls we compared performance of the readout units trained on patterns from the cortical network versus (see Figure 7A): readout units trained on the cortical network without STP; the untrained cortical network

(i.e., the output of the L-IV units); and no cortex (i.e., the spike patterns generated from the cochleagram). Note that the spatiotemporal patterns from the cochleagram must contain all the relevant information for digit discrimination, so any improvement in performance over the no cortex group can be attributed to transformations imposed by the cortical circuits that improve the readout of the information.

In the forward condition, the readout units were trained on 9 of the 10 utterances of each digit and tested on the omitted utterance (see Figure 7B, left panel). In all four conditions, the readout units were able to discriminate among the forward digits well above chance level (error rates of less than 12%, compared to chance error rates of 90%). However, the no cortex condition performed the worst, and the error rate was significantly higher than the cortex condition (11.5% versus 5.3%, t-test, $p = .016$), indicating that the transformation imposed by the cortical circuitry significantly enhances the ability of the readout units to capture the digit-specific signature of the input patterns.

The reasonable performance of the readout units even in the absence of the cortex is expected given the fact that many digits can be discriminated based solely on their spectral signature, independent of their spatiotemporal features. To examine true spatiotemporal discrimination and better capture the electrophysiological and perceptual findings that clearly show high degrees of direction selectivity, we next trained the readout units to discriminate among the forward and reverse instances of all the digits (see Figure 7B, right panel). Here the performance of the readout units in the no cortex condition was very bad (>60% error rates). Again the cortical network performed the best (12.2% error rate) and was significantly better than training in the absence of STP (20.4%, t-test, $p = .002$) or on the naïve network (27.4%, t-test, $p < 10^{-5}$). These results further support the conclusion that STP may play a critical role in direction selectivity.

4 Discussion

A large number of in vivo electrophysiological studies have demonstrated that auditory cortex neurons of mammals (Wang et al., 1995; Kilgard & Merzenich, 2002; Yin et al., 2008; Sadagopan & Wang, 2009; Zhou, de Villers-Sidani, Panizzutti, & Merzenich, 2010) or neurons in the song areas of songbirds (Margoliash, 1983; Lewicki & Arthur, 1996; Doupe, 1997) can be highly selective to the spatiotemporal structure of vocalizations. It is generally believed that this selectivity in the early stages of sensory processing is critical to vocal communication.

Despite the importance of selectivity for the processing of complex stimuli, relatively little is known about the mechanisms and ontogenesis of this selectivity. Previous models have demonstrated that selectivity to complex stimuli can be observed in randomly connected networks (Buonomano & Merzenich, 1995; Buonomano, 2000; Maass et al., 2002, 2003; Jaeger,

Lukoševičius, & Popovici, 2007; Buonomano & Maass, 2009). However, these state-dependent/liquid state machine approaches have required fine-tuning the average weights to ensure the network operates in an appropriate regime. Additionally, in the previous studies, as the result of the absence of recurrent plasticity, there was no "tuning" to the stimulus set at hand. Here we have shown for the first time that the appropriate regimes can be reached in an entirely unsupervised manner; furthermore, there is some degree of tuning of the network to the stimulus set at hand. In the initial state of our cortical network model the L-II/III units were nonresponsive to sensory stimulation, but the presynaptic-dependent scaling rule led to the emergence of activity within the recurrent network and to neurons that exhibited a wide range of selectivity values. This variability captures the inherent diversity observed in vivo, in which some cells exhibit high degrees of selectivity, while others seem to exhibit little or no selectivity. Furthermore, we demonstrated that the distribution of selective neurons provided a robust spatiotemporal code that allowed downstream neurons to discriminate among digits with an accuracy of approximately 95%.

4.1 Forward Versus Reverse. To determine if neurons are truly selective to the spatiotemporal structure of stimuli, as opposed to some characteristic spectral signature, electrophysiological studies typically contrast the number of spikes elicited by forward and reverse presentation of stimuli. Results often reveal a high degree of direction selectivity (Wang et al., 1995; Doupe, 1997; Theunissen & Doupe, 1998). Some studies, however, depending on cortical area and the stimuli used, have not revealed significant selectivity between forward and reverse stimuli, yet even in these studies, there was more information present about the forward stimuli than the reverse (Schnupp, Hall, Kokelaar, & Ahmed, 2006; Recanzone, 2008). The reported direction selectivity of some neurons is consistent with the obvious perceptual difference between forward and reverse stimuli. It is also important to note that perceptual direction selectivity seems to be inherent in the underlying neural mechanisms—not learned by experience. That is, cells in the birdsong auditory system respond selectively to forward presentations of song even though they were never exposed to the reverse presentations during the life of the bird. Similarly, although someone may never have heard his or her name said backward, nobody is in danger of confusing the forward and reverse presentations of his or her name. Indeed, human and animal studies show that animals are generally very sensitive to time-reversed stimuli (Doupe & Kuhl, 1999; Saberi & Perrott, 1999; Ghazanfar & Hauser, 2001), although there are exceptions (Mathevon & Aubin, 2001; Ghazanfar, Smith-Rohrberg, Pollen, & Hauser, 2002).

Here we show that the state-dependent nature of cortical network dynamics seems to naturally account for the emergence of direction selectivity—at least in a significant percentage of neurons. We further establish that short-term synaptic plasticity is partly responsible for the observed

direction sensitivity of the units. Indeed, as previously suggested, short-term synaptic plasticity provides an inherent temporal asymmetry, thus allowing for the processing of both temporal and spatiotemporal discrimination (Buonomano, 2000). Importantly, while STP contributes in a fundamental manner to imposing a temporal asymmetry that permits direction selectivity, homeostatic plasticity drives the network to regimes capable of efficiently coding for different digits. This will occur in either the presence or absence of STP. Thus, both mechanisms act essentially independently.

Together our results lead us to the prediction that if short-term plasticity could be selectively turned on or off (or altered in a controllable fashion), there would be a significant decrease in the number of direction-selective neurons. Moreover, the discrimination of complex spatiotemporal patterns, such as speech, would be much more severely affected than the discrimination of complex spatial stimuli, such as images.

4.2 Homeostatic Plasticity. The simulation presented here used the PSD plasticity rule, a hypothesized form of homeostatic plasticity (Buonomano, 2005; Liu & Buonomano, 2009). This rule is distinct from the traditional synaptic scaling plasticity rule (van Rossum et al., 2000) in that the homeostatic changes in synaptic strength also depend on the average levels of activity in the presynaptic cells. The rule overcomes the instability observed in computational models of recurrent networks that only incorporate synaptic scaling (Buonomano, 2005; Houweling et al., 2005; Frohlich et al., 2008). While PSD has not been directly tested experimentally, it is important to highlight that it is consistent with most of the experimental data used to support synaptic scaling. Additionally, in comparison to traditional synaptic scaling, PSD better accounts for observations in recurrent circuits, demonstrating that decreased activity can lead to both potentiation and depression of synapses (Thiagarajan, Lindskog, & Tsien, 2005; Mitra, Mitra, & Tsien, 2012). We stress, however, that we are not suggesting that this is the only rule that may lead to the presented results.

4.3 Weaknesses. One limitation of our model is that all of our analysis was based on selectivity to digits from a single speaker. It was, however, not our goal to account for the more complex computational problem of invariant pattern recognition. Indeed, there is little evidence that neurons in early sensory cortices exhibit significant temporal or spectral invariance. As has been hypothesized in the visual system with regard to scalar and position invariance (Fukushima, 1988), speaker invariance may arise from the increasing convergence of cells with little invariance onto cells with progressively more insensitivity to local spatial and temporal manipulations of their preferred stimuli, although alternate models have been proposed (Gutig & Sompolinsky, 2009).

A second limitation of our simulations was that we included only one form of long-term synaptic plasticity: presynaptic-dependent scaling, a

hypothesized form of homeostatic plasticity (Buonomano, 2005). Another candidate form of plasticity that has been suggested to play a role in processing complex time-varying stimuli is STDP (Abbott & Nelson, 2000; Dan & Poo, 2004). While we expect STDP to contribute to general cortical plasticity and organization, previous simulations have suggested that STDP may not be ideally suited to be the principal plasticity rule for spatiotemporal stimuli (Buonomano, 2005; Liu & Buonomano, 2009). One reason why relates to what can be thought of as the latency reduction problem: STDP by its very nature increases the synaptic strength between neurons that are sequentially active within a time window of tens of milliseconds (Song, Miller, & Abbott, 2000). As synapses become stronger, the latency of the postsynaptic neuron becomes shorter and may shift toward the onset of a stimulus, such as a spoken word. This latency reduction process can decrease the neuron's selectivity to the temporal features of the stimuli. For example, in the extreme, a neuron that fires at the onset of a spoken word cannot be selective to any temporal features. Nevertheless it is our view, and a line of future research, that if STDP is incorporated into recurrent networks in a balanced manner with other synaptic plasticity rules operating, it may contribute to the formation of spatiotemporal selectivity. Indeed, in a simpler model that focused on the detection of sequences of events, it has been shown that STDP, together with synaptic scaling, enhances the formation of sequence-selective neurons (Lazar, Pipa, & Triesch, 2009).

4.4 State-Dependent Networks. A cardinal feature of state-dependent network models (Buonomano & Merzenich, 1995; Buonomano, 2000) and so-called reservoir computing in general (Maass et al., 2003; Jaeger & Haas, 2004; Buonomano & Maass, 2009) is that the discrimination of complex time-varying stimuli relies heavily on the inherent interaction between stimuli and the complexity and highly nonlinear nature of neural networks (Buonomano & Maass, 2009). In other words, in the previous instantiations of this class of models, selectivity is in a sense stochastic: given a large enough network, complex circuitry, and a rich set of time-varying synaptic properties, some subset of neurons should respond with some degree of selectivity to almost any stimulus set (Buonomano & Merzenich, 1995; Maass et al., 2002; Haeusler & Maass, 2007; Karmarkar & Buonomano, 2007; Buonomano & Maass, 2009). This notion stands in contrast with the more typical view that a neuron that responds selectively to a specific stimulus, direction, or interval is the product of specific mechanisms and plasticity rules "designed" to solve the particular problem at hand. Given the obvious complexity, the statistical nature of network connectivity, and the apparent importance of network size, we would maintain that our framework contributes to the selectivity to spatiotemporal patterns in vivo. Indeed it is of relevance that even when animals are presented ecologically unnatural or novel stimuli (e.g., marmoset vocalizations to a cat) and there is no clear population bias toward forward or reverse stimuli, some subset of neurons nevertheless

prefers the forward or reverse stimuli—as if by chance (Wang & Kadia, 2001; Huetz et al., 2009). And the fact that some units fire selectively to the reverse stimuli strongly suggests that in addition to the obvious learned component of vocalization selectivity, there is also a stochastic component.

The work described here demonstrates that a simple form of homeostatic plasticity is sufficient to guide the network into regimes where there is a reasonable balance of excitation and inhibition (i.e., most neurons fire, but there is no runaway excitation), and, more important, the neurons exhibit a rich diversity of preferred responses. But the results extend beyond finding a random pattern of weights that satisfies the above conditions and indicates that in some sense, the network is tuned by stimulus set. First, we established a significant decrease in direction selectivity when the network is trained on the reverse stimuli (see Figure S1). Second, if the performance of the network was entirely a product of finding an appropriate distribution of synaptic weights, there would be no decrement in performance if the plastic weights were shuffled. As expected, shuffling resulted in a dramatic decrease in overall activity (which limits its value in making a direct comparison with the unshuffled runs). Nevertheless, shuffling the weights did produce a highly significant ($p < 0.001$) decrease in direction selectivity (see Figure S3). Furthermore, and most important, during training, only half of the digits were used (see section 2); thus, it was possible to examine if more cells were selective to the trained digits. The number of cells that exhibited forward selectivity for digits 0 to 4 was more than were selective for digits 5 to 9 when the network was trained on 0 to 4, and, conversely, more cells were selective for 5 to 9 when trained on 5 to 9 (see Figure S4). The training set, however, did not result in a difference in direction selectivity between the trained and untrained digits. Together, our results establish that the performance of the network relies on two factors: finding the appropriate regime (which can also be achieved by randomly assigning weights from an appropriate distribution) and stimulus-specific forms of experience-dependent plasticity that tune specific weights in a manner that amplifies performance.

While we argue that the reliance on the size and complexity of cortical networks is both physiological and critical, it is nevertheless clear that experience-dependent plasticity plays an important role in cortical function. Indeed a cardinal feature of cortical circuitry is its ability to adapt and reorganize in accordance to the stimuli to which they are exposed (Buonomano & Merzenich, 1998; Feldman & Brecht, 2005; Karmarkar & Dan, 2006). The synapses responsible for the recurrent architecture of cortical circuits undergo a number of different forms of plasticity (Abbott & Nelson, 2000; Dan & Poo, 2004; Turrigiano, 2008; Pozo & Goda, 2010). While we have reasonable models of how these plasticity rules contribute to the processing of spatial stimuli, an open challenge remains how these rules contribute to the processing of complex time-varying stimuli. Here we have taken an important step in this direction by showing that a relatively simple form

of homeostatic plasticity is sufficient to guide naïve networks into regimes where stimulus selectivity emerges.

Acknowledgments

References

Abbott, L., & Nelson, S. (2000). Synaptic plasticity: Taming the beast. *Nat. Neurosci.*, *3*, 1178–1183.

Abbott, L., Regehr, W. (2004). Synaptic computation. *Nature*, *431*, 796–803.

Aubie, B., Becker, S., & Faure, P. (2009). Computational models of millisecond level duration tuning in neural circuits. *J. Neurosci.*, *29*, 9255–9270.

Bienenstock, E., Cooper, L., & Munro, P. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *J. Neurosci.*, *2*, 32–48.

Buonomano, D. (2000). Decoding temporal information: A model based on short-term synaptic plasticity. *J. Neurosci.*, *20*, 1129–1141.

Buonomano, D. (2005). A learning rule for the emergence of stable dynamics and timing in recurrent networks. *J. Neurophysiol.*, *94*, 2275–2283.

Buonomano, D., & Maass, W. (2009). State-dependent computations: Spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.*, *10*, 113–125.

Buonomano, D., & Merzenich, M. (1995). Temporal information transformed into a spatial code by a neural network with realistic properties. *Science*, *267*, 1028–1030.

Buonomano, D., & Merzenich, M. (1998). Cortical plasticity: From synapses to maps. *Annu. Rev. Neurosci.*, *21*, 149–186.

Carvalho, T., & Buonomano, D. (2009). Differential effects of excitatory and inhibitory plasticity on synaptically driven neuronal input-output functions. *Neuron*, *61*, 774–785.

Dan, Y., & Poo, M. M. (2004). Spike timing-dependent plasticity of neural circuits. *Neuron*, *44*, 23–30.

Destexhe, A., Mainen, Z., & Sejnowski, T. (1994). An efficient method for computing synaptic conductances based on a kinetic model of receptor binding. *Neural Comput.*, *6*, 14–18.

Doupe, A. (1997). Song- and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development. *J. Neurosci.*, *17*, 1147–1167.

Doupe, A., & Kuhl, P. (1999). Birdsong and human speech: Common themes and mechanisms. *Annu. Rev. Neurosci.*, *22*, 567–631.

Doya, K. (1992). Bifurcations in the learning of recurrent neural networks. In *Proceedings of the 1992 IEEE International Symposium on Circuits and Systems* (pp. 2777–2780). Piscataway, NJ: IEEE.

Drew, P., & Abbott, L. (2003). Model of song selectivity and sequence generation in area HVc of the songbird. *J. Neurophysiol.*, *89*, 2697–2706.

Drullman, R. (1995). Temporal envelope and fine structure cues for speech intelligibility. *J. Acoust. Soc. Am.*, *97*, 585–592.

Echevarria, D., & Albus, K. (2000). Activity-dependent development of spontaneous bioelectric activity in organotypic cultures of rat occipital cortex. *Brain Res. Dev. Brain Res.*, *123*, 151–164.

Engineer, C., Perez, C., Chen, Y., Carraway, R., Reed, A., Shetake, J., et al. (2008). Cortical activity patterns predict speech discrimination ability. *Nat. Neurosci.*, *11*, 603–608.

Feldman, D., & Brecht, M. (2005). Map plasticity in somatosensory cortex. *Science*, *310*, 810–815.

Ferster, D., & Miller, K. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annu. Rev. Neurosci.*, *23*, 441–471.

Fiala, J., Grossberg, S., & Bullock, D. (1996). Metabotropic glutamate receptor activation in cerebellar Purkinje cells as substrate for adaptive timing of the classically conditioned eye-blink response. *J. Neurosci*, *16*, 3760–3774.

Froemke, R., Merzenich, M., & Schreiner, C. (2007). A synaptic memory trace for cortical receptive field plasticity. *Nature*, *450*, 425–429.

Frohlich, F., Bazhenov, M., & Sejnowski, T. (2008). Pathological effect of homeostatic synaptic scaling on network dynamics in diseases of the cortex. *J. Neurosci.*, *28*, 1709–1720.

Fukushima, K. (1988). Neocognitron: A hierarchical neural network capable of visual patten recognition. *Neural Networks*, *1*, 119–130.

Gentner, T., & Margoliash, D. (2003). Neuronal populations and single cells representing learned auditory objects. *Nature*, *424*, 669–674.

Ghazanfar, A., & Hauser, M. (2001). The auditory behaviour of primates: A neuroethological perspective. *Current Opinion in Neurobiology*, *11*, 712–720.

Ghazanfar, A., Smith-Rohrberg, D., Pollen, A., & Hauser, M. (2002). Temporal cues in the antiphonal long-calling behaviour of cottontop tamarins. *Animal Behaviour*, *64*, 427–438.

Gutig, R., & Sompolinsky, H. (2006). The tempotron: A neuron that learns spike timing-based decisions. *Nat. Neurosci.*, *9*, 420–428.

Gutig, R., & Sompolinsky, H. (2009). Time-warp—Invariant neuronal processing. *PLoS Biol.*, *7*, e1000141.

Haeusler, S., & Maass, W. (2007). A statistical analysis of information-processing properties of lamina-specific cortical microcircuit models. *Cerebral Cortex*, *17*, 149–162.

Hines, M., & Carnevale, N. (1997). The NEURON simulation environment. *Neural Comput.*, *9*, 1179–1209.

Houweling, A., Bazhenov, M., Timofeev, I., Steriade, M., & Sejnowski, T. (2005). Homeostatic synaptic plasticity can explain posttraumatic epileptogenesis in chronically isolated neocortex. *Cerebral Cortex*, *15*, 834–845.

Hubel, D., & Wiesel, T. (1963). Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *J. Neurophysiol.*, *26*, 994–1002.

Huetz, C., Philibert, B., & Edeline, J.-M. (2009). A spike-timing code for discriminating conspecific vocalizations in the thalamocortical system of anesthetized and awake guinea pigs. *J. Neurosci.*, *29*, 334–350.

Ivry, R., & Schlerf, J. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences*, *12*, 273–280.

Jaeger, H., Lukoševičius, M., & Popovici, D. (2007). Optimization and applications of echo state networks with leaky integrator neurons. *Neural Networks*, *20*, 335–352.

Jaeger, H., & Haas, H. (2004). Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, *304*, 78–80.

Karmarkar, U., & Buonomano, D. (2007). Timing in the absence of clocks: Encoding time in neural network states. *Neuron*, *53*, 427–438.

Karmarkar, U., & Dan, Y. (2006). Experience-dependent plasticity in adult visual cortex. *Neuron*, *52*, 577–585.

Kilgard, M., & Merzenich, M. (2002). Order-sensitive plasticity in adult primary auditory cortex. *Proc. Natl. Acad. Sci. USA*, *99*, 3205–3209.

Lazar, A., Pipa, G., & Triesch, J. (2009). SORN: A self-organizing recurrent neural network. *Front. Comput. Neurosci.*, *3*, 23.

Lewicki, M., & Arthur, B. (1996). Hierarchical organization of auditory temporal context sensitivity. *J. Neurosci.*, *16*, 6987–6998.

Lewicki, M., & Konishi, M. (1995). Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. *Proc. Natl. Acad. Sci. USA*, *92*, 5582–5586.

Liu, J., & Buonomano, D. (2009). Embedding multiple trajectories in simulated recurrent neural networks in a self-organizing manner. *J. Neurosci.*, *29*, 13172–13181.

Maass, W., Natschläger, T., & Markram, H. (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Comput.*, *14*, 2531–2560.

Maass, W., Natschläger, T., & Markram, H. (2003). A model of real-time computation in generic neural microcircuits. In S. Becker, S. Thrün, & K. Obermayer (Eds.), *Advances in neural information processing systems*, *15* (pp. 229–236). Cambridge, MA: MIT Press.

Margoliash, D. (1983). Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J. Neurosci.*, *3*, 133–143.

Markram, H., Wang, Y., & Tsodyks, M. (1998). Differential signaling via the same axon of neocortical pyramidal neurons. *Proc. Natl. Acad. Sci. USA*, *95*(9), 5323–5328.

Mathevon, N., & Aubin, T. (2001). Sound-based species-specific recognition in the blackcap *Sylvia atricapilla* shows high tolerance to signal modifications. *Behaviour*, *138*, 511–524.

Mauk, M., & Buonomano, D. (2004). The neural basis of temporal processing. *Ann. Rev. Neurosci.*, *27*, 307–340.

Mitra, A., Mitra, S., & Tsien, R. (2012). Heterogeneous reallocation of presynaptic efficacy in recurrent excitatory circuits adapting to inactivity. *Nat. Neurosci.*, *15*, 250–257.

Moore, J., Desmond, J., & Berthier, N. (1989). Adaptively timed conditioned responses and the cerebellum: A neural network approach. *Biol. Cybern.*, *62*, 17–28.

Pearlmutter, B. (1995). Gradient calculation for dynamic recurrent neural networks: A survey. *IEEE Trans. on Neural Network*, *6*, 1212–1228.

Pozo, K., & Goda, Y. (2010). Unraveling mechanisms of homeostatic synaptic plasticity. *Neuron*, *66*, 337–351.

Prather, J., Nowicki, S., Anderson, R., Peters, S., & Mooney, R. (2009). Neural correlates of categorical perception in learned vocal communication. *Nat. Neurosci.*, *12*, 221–228.

Prieto, J., Peterson, B., & Winer, J. (1994). Morphology and spatial distribution of GABAergic neurons in cat primary auditory cortex (AI). *J. Comp. Neurol.*, *344*, 349–382.

Razak, K., & Fuzessery, Z. (2008). Facilitatory mechanisms underlying selectivity for the direction and rate of frequency modulated sweeps in the auditory cortex. *J. Neurosci.*, *28*, 9806–9816.

Razak, K., & Fuzessery, Z. (2009). GABA shapes selectivity for the rate and direction of frequency-modulated sweeps in the auditory cortex. *J. Neurophysiol.*, *102*, 1366–1378.

Recanzone, G. (2008). Representation of con-specific vocalizations in the core and belt areas of the auditory cortex in the alert macaque monkey. *J. Neurosci.*, *28*, 13184–13193.

Saberi, K., & Perrott, D. (1999). Cognitive restoration of reversed speech. *Nature*, *398*, 760–760.

Sadagopan, S., & Wang, X. (2009). Nonlinear spectrotemporal interactions underlying selectivity for complex sounds in auditory cortex. *J. Neurosci.*, *29*, 11192–11202.

Saitoh, I., & Suga, N. (1995). Long delay lines for ranging are created by inhibition in the inferior colliculus of the mustached bat. *J. Neurophysiol.*, *74*, 1–11.

Schnupp, J., Hall, T., Kokelaar, R., & Ahmed, B. (2006). Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *J. Neurosci.*, *26*, 4785–4795.

Sen, K., Theunissen, F., & Doupe, A. (2001). Feature analysis of natural sounds in the songbird auditory forebrain. *J. Neurophysiol.*, *86*, 1445–1458.

Shannon, R., Zeng, F., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.

Song, S., & Abbott, L. (2001). Cortical development and remapping through spike timing-dependent plasticity. *Neuron*, *32*, 339–350.

Song, S., Miller, K., & Abbott, L. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nat. Neurosci.*, *3*, 919–926.

Theunissen, F., & Doupe, A. (1998). Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVc of male zebra finches. *J. Neurosci.*, *18*, 3786–3802.

Thiagarajan, T., Lindskog, M., & Tsien, R. (2005). Adaptation to synaptic inactivity in hippocampal neurons. *Neuron*, *47*, 725–737.

Turrigiano, G. (2008). The self-tuning neuron: Synaptic scaling of excitatory synapses. *Cell*, *135*, 422–435.

Turrigiano, G., & Nelson, S. (2004). Homeostatic plasticity in the developing nervous system. *Nat. Neurosci. Rev.*, *5*, 97–107.

van Rossum, M., Bi, G., & Turrigiano, G. (2000). Stable Hebbian learning from spike timing-dependent plasticity. *J. Neurosci.*, *20*, 8812–8821.

Wang, X., & Kadia, S. (2001). Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. *J. Neurophysiol.*, *86*, 2616–2620.

Wang, X., Merzenich, M., Beitel, R., & Schreiner, C. (1995). Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: Temporal and spectral characteristics. *J. Neurophysiol.*, *74*, 2685–2706.

Wilent, W., & Contreras, D. (2005). Dynamics of excitation and inhibition underlying stimulus selectivity in rat somatosensory cortex. *Nat. Neurosci.*, *8*, 1364–1370.

Wu, G., Li, P., Tao, H., & Zhang, L. (2006). Nonmonotonic synaptic excitation and imbalanced inhibition underlying cortical intensity tuning. *Neuron*, *52*, 705–715.

Yin, P., Mishkin, M., Sutter, M., & Fritz, J. (2008). Early stages of melody processing: Stimulus-sequence and task-dependent neuronal activity in monkey auditory cortical fields A1 and R. *J. Neurophysiol.*, *100*, 3009–3029.

Zhou, X., de Villers-Sidani, É., Panizzutti, R., & Merzenich, M. (2010). Successive-signal biasing for a learned sound sequence. *Proc. Natl. Acad. Sci. USA*, *107*, 14839–14844.

Zucker, R. (1989). Short-term synaptic plasticity. *Annu. Rev. Neurosci.*, *12*, 13–31.