ELSEVIER Neurocomputing 44–46 (2002) 789–794

NEUROCOMPUTING

www.elsevier.com/locate/neucom

Modeling temporal combination selective neurons of the songbird

Patrick J. Drew*, L.F. Abbott

Volen Center for Complex Systems and Department of Biology, Brandeis University, Waltham, MA 02454, USA

Abstract

Some neurons in the nucleus HVc of the songbird respond vigorously to sequences of syllables as they appear in the bird's own song (such as AB), but they respond weakly or not at all when the same syllables are played individually (A or B) or in a different order (BA). We have constructed a network model that replicates this temporal sequence selectivity. The model is based on recurrently connected networks that produce strong resonant responses when the pattern of excitation evoked by a stimulus matches the pattern of excitation generated internally within the network. In the model, syllable B does not generate such a resonant response by itself. However, if syllable A is presented to the network followed by syllable B, the activity generated by A modifies the effective connectivity of the network making it resonantly responsive to B. This produces a highly selective response to the sequence of syllables AB, but not to any other combination. © 2002 Published by Elsevier Science B.V.

Keywords: Temporal sequence selectivity; Birdsong

1. Introduction

Many neurons in the nucleus HVc of certain songbirds are well tuned to specific syllables of the bird's own song [3,4], and of these, some are responsive only to specific temporal combinations of syllables [4,2]. These are the so-called temporal combination selective neurons, which only respond to a syllable combination such as AB, but not to BA, AA, or BB. This selectivity requires the integration of information over time periods of up to several hundred milliseconds [4,2]. Neurons in field *L*, which provides the primary input, both directly and indirectly, to HVc, do

---

* Corresponding author.
E-mail addresses: drew@brandeis.edu (P.J. Drew), abbott@brandeis.edu (L.F. Abbott).

0925-2312/02/$ - see front matter © 2002 Published by Elsevier Science B.V.
PII: S0925-2312(02)00473-3

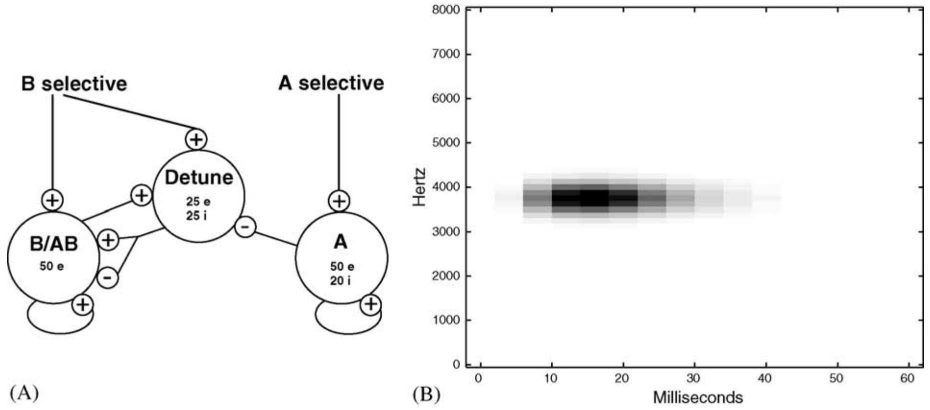

Fig. 1. (A) Network model of temporal combination selective neurons in the HVc of the songbird. Numbers indicate how many excitatory and inhibitory neurons are in each subpopulation. (B) A typical STRF used to generate input to the network. The dark area shows the frequency and time range over to which the STRF is sensitive.

not display marked temporal combination selectivity [2], and their spectral temporal receptive fields (STRFs) indicated that they do not integrate over periods as long as those needed to produce temporal sequence selectivity [6,5]. This suggests that much of the temporal combination selectivity arises within HVc.

2. The network model

Any model of temporally selective neuronal responses, such as selectivity for the combination AB, requires two basic features: a mechanism to retain a memory of stimulus A, at least until the onset of B; and a way of allowing this memory trace of A to affect the response the B evokes. In our model, recurrent excitation plays a key role in both of these mechanisms. Networks of neurons with a high degree of recurrent excitation have two features that are well suited for these tasks. First, they can generate "reverberatory" activity that can sustain itself for a period that significantly outlasts the stimulus. These reverberations endow the network with a form of short-term memory and provide the memory trace in our model. Second, recurrent networks can amplify responses to stimuli, but only if they are configured appropriately. In our model, the presence of stimulus A enhances the response to stimulus B by configuring a recurrent network selective for B in such a way that its response is highly amplified.

The network model we discuss is constructed from integrate-and-fire neurons receiving synaptic input modeled as a combination of AMPA, NMDA, and GABA conductances (see [7]). Following the general ideas outlined in the previous paragraph, the network consists of three subpopulations of interconnected spiking neurons, A selective, AB selective, and "detuning" (see Fig. 1A). These are characterized by their responses to two temporally adjacent syllables (A and B) taken from real zebra finch songs. The

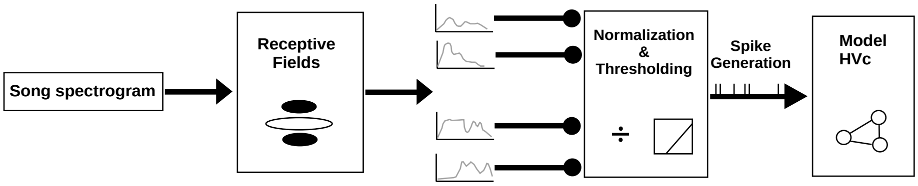

Fig. 2. Model layout: birdsong is transformed into firing rates by receptive fields. After normalization and thresholding, the rates are converted into spikes that are fed into the recurrent network.

A-selective population, composed of both excitatory and inhibitory neurons, is driven by inputs that are themselves selective for A through a mechanism explained later in this section. Selectivity is further amplified by strong recurrent connections between the excitatory A-selective neurons, which have the additional effect of extending the activity in response to syllable A so that it outlasts the duration of that syllable.

The AB selective subpopulation receives input that is selective for syllable B and, like the A-selective network, it has strong recurrent excitation. When appropriately configured, the AB selective subnetwork "resonates" when B is presented, strongly amplifying its B-selective input. However, this amplification only occurs in the presence of the sustained activity of the A-selective network evoked by syllable A. In the absence of this activity, a set of detuning neurons spoils the connectivity of the AB-selective network preventing it from amplifying syllable B. The detuning subpopulation, composed of both excitatory and inhibitory neurons, receives B-selective input as well as excitatory input from the AB-selective neurons. The detuning neurons both excite and inhibit members of the AB-selective subpopulation. These interactions serve to disrupt the resonant response of the AB selective group to stimulus B by silencing some cells and stimulating others, perturbing the precise tunings that is required for selective amplification of the B signal. If firing of the detuning cells is not suppressed, the amplification of the B input in the AB subnetwork is drastically reduced. The activity pattern generated in the network by the presentation of A tunes the network by silencing the detuning cells, allowing a strong resonant response to B. If B is presented alone or prior to A, so that the network has not been pretuned by A, the response to B is significantly weaker than if A had preceded B. In this way, temporal combination selectivity is generated.

It only remains to explain how A- and B-selective inputs to the networks of the model are generated. Our model takes real digitized bird songs and syllables, and using realistic STRFs (Fig. 1B), generates a time-varying output similar to what might be produced within field *L*. This output is then fed into a "syllable detector" that generates spikes when the chosen syllable is detected (see Fig 2). These spikes provide the syllable-selective input to the network of spiking neurons shown in Fig. 1.

The STRFs we use are based on those obtained experimentally from field *L* [6]. These receptive fields do not integrate information any further back than 50 ms, so they do not span enough time to be able to individually detect a sequence of syllables. A sample STRF is shown in Fig. 1B. An entire bank of such filters was used, with

STRFs given by

$$\mathbf{r}(f,t) = \frac{\alpha(\alpha t)^5}{2\pi f_{\text{width}}} \exp\left(-5\left(\frac{f-f_i}{f_{\text{width}}}\right)^2\right)\exp(-\alpha t),$$

where $\alpha = 3/\text{s}$, $f_{\text{width}} = 225\,\text{Hz}$, f_i is the frequency where the STRF labeled by i is centered, and f and t are frequency and time, respectively. The entire bank consists of 37 filters with different f_i evenly spaced between 600 and 7000 Hz.

The spectral temporal receptive fields are convolved with a spectrogram of syllable sequences or entire songs taken from zebra finch recordings supplied by K. Sen and A.J. Doupe. The resulting output from the entire bank of filters then forms a vector of firing rates. This is fed into a syllable detector that performs two operations. First, it normalizes the firing rate vector at each point in time to allow amplitude-independent syllable identification. This is similar to models of contrast normalization in primary visual cortex [1]. Second, a weighted sum of all the filter outputs is computed and thresholded. The weights are set equal to the normalized output of the filter bank during the syllable we wish to recognize. Above the threshold, which is set by hand, spikes are generated using a Poisson process. The spikes then act as input to the recurrent network.

3. Results

We presented the network with two sequential syllables from a zebra finch song. One syllable is designated A, the second, B. The syllables were presented in normal and reversed order. The results are shown in Fig. 3. The top panels in this figure show the syllables A and B presented in the order AB (left) and BA (right). The panels second from the top show the response of a typical A-selective neuron. Note that its activity outlasts the presentation of syllable A providing a memory trace. This is due to the recurrent coupling among the A-selective neurons. The third panels from the top in Fig. 3 show a typical AB selective neuron and indicate its strong preference for B only if preceded by A (left plot), but not the reverse (right plot). If B is played before A or in isolation, the detuning cells are excited, as indicated by the bottom panel on the right of Fig. 3. When A is played before B, the detuning neurons are inhibited enough (bottom panel at the left of Fig. 3) so they cannot disrupt the resonate response of the AB selective neurons, allowing strong amplification of the B syllable.

There is a gradation of temporal combination selectivity present in the AB selective neurons, as the perturbation affects individual neurons slightly differently. Some (approximately a quarter) only fire when a specific temporal combination is presented (similar to the AB-selective neuron shown in Fig. 3), while approximately another quarter are almost completely insensitive to the context in which their prefered syllable is presented. In between there is a graded continuum of responses.

4. Discussion

Our model makes use of the fact that complex sounds can be discriminated using the output of simple spectral temporal receptive fields if these are appropriately normalized

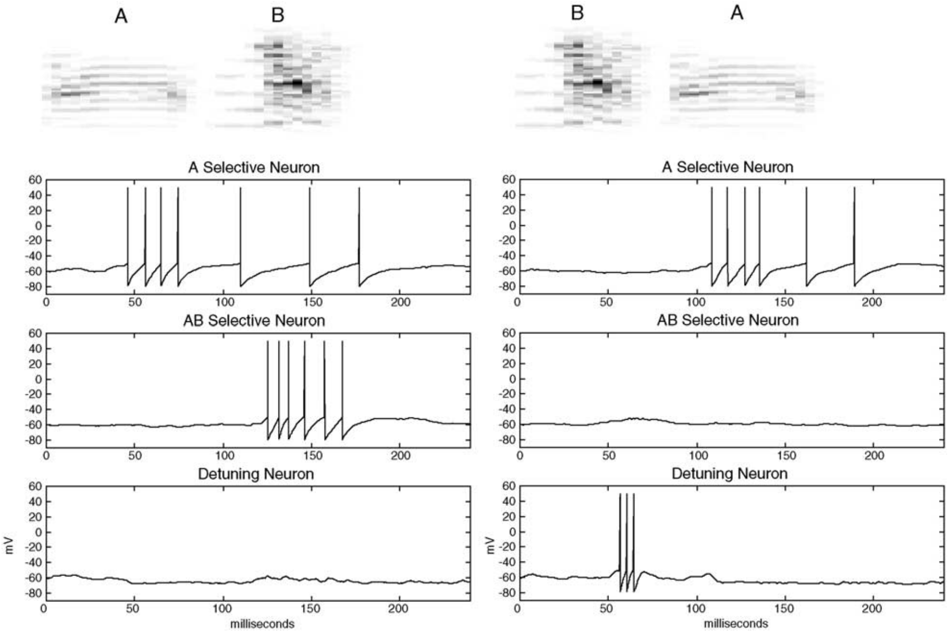

Fig. 3. Sample neural activity from neurons in each of the subpopulations in response to in order (left) and out of order (right) syllable presentations.

and thresholded. The normalization of the output is critical for this recognition; otherwise any type of loud sound could cause a false positive syllable detection. The relative intensities at the various frequencies are more important for discrimination than the absolute amplitudes.

In our model of temporal combination selectivity, prior input alters the effective connectivity of the network, enabling it to respond to later input patterns more strongly than if they were not preceded by the gating stimulus. This leads to several predictions. First, the network activity induced by the presentation of a single syllable should outlast the duration of the syllable. The duration of this persistance should be of the same order as the largest silent interval that can separate two syllables but still excite the temporal combination selective cells for the sequence. Another prediction of our model is that there will be neurons that respond to B presented in isolation, but not B preceded by A. These cells will be inhibited when A is presented. Though no such cells have been found, this is most likely a consequence of the experimental paradigm. Using the bird's own song as a probe, which is the typical stimulus used when looking for auditory cells in HVc, should tend to inhibit detuning cells, making them obscure. An optimal stimulus to find such cells would be to play the bird's own song with syllables in reverse order. Cells that are excited by this stimulus, but that are also surpressed by the bird's own song, could be the detuning cells of our model.

References

[1] D.J. Heeger, Normalization of cell responses in cat striate cortex, Visual Neurosci. 9 (1992) 181–197.

[2] M.S. Lewicki, B.J. Arthur, Hierarchical organization of auditory temporal context sensitivity, J. Neurosci. 16 (1996) 6987–6998.

[3] D. Margoliash, Acoustic parameter underlying the response of song-specific neurons in the white-crowned sparrow, J. Neurosci. 3 (1983) 1039–1057.

[4] D. Margoliash, E.S. Fortune, Temporal and harmonic combination-sensitive neurons in the song-system of white-crowned sparrows, J. Neurosci. 12 (1994) 4309–4326.

[5] K. Sen, F.E. Theunissen, A.J. Doupe, Feature analysis of natural sounds in the songbird auditory forebrain, J. Neurophysiol. 86 (2001) 1445.

[6] F.E. Theunissen, K. Sen, A.J. Doupe, Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds, J. Neurosci. 20 (2000) 2315–2331.

[7] X.J. Wang, Synaptic basis of cortical persistant activity: the importance of NMDA receptors to working memory, J. Neurosci. 19 (1999) 9587–9603.