

# Songbirds, Sequences, and Timing

---

*Abstract: This paper outlines a model of bird song timing based on existing research on the actual neuroanatomy of the zebra finch.*

The zebra finch, *Taeniopygia guttata*, is one of the most popular domestic birds, prized for its beauty, ease of care, and lovely, boisterous song. Song production is dimorphic – males produce a more complex and stereotyped repertoire than females and the male zebra finch brain has a song nucleus eight times larger than the female.

Some songbirds, canaries for example, may develop a new repertoire each year; the male zebra finch, however, has a single set of songs learned during its first 30-90 days of life, though refinement of the songs continues until about 200 days post-hatch. Significant neurogenesis is associated with song acquisition during either juvenile to adult transition or during changes in repertoire in species for which the repertoire has a seasonal nature. Each male has a distinctive song (or motif) that incorporates elements of its father's song along with other sounds from its natural environment. Songs are typically preceded by a series of introductory sounds while the motif is composed of a temporally precise sequence of 3-7 syllables, though tempo changes may occur for a variety of reasons (Glaze & Troyer, 2006). A song syllable is 50-250 ms long, separated by short pauses in vocal production that are accompanied by respiratory expirations or sometimes abrupt changes in acoustic characteristics. Notes within a syllable are on the order of 30-70 ms. The length of the pauses and of each syllable are not highly correlated with each other and the pauses tend to be much more variable in length (Yu & Margoliash; see fig. 2), which is evidence that in favour of the proposition that the syllable as the fundamental unit of a motif, however, there is some evidence favouring the note (Glaze & Troyer, 2007).

The earliest song-selective neurons are found in the premotor telencephalic nucleus high vocal center (HVC, as a proper name) (Lewicki & Arthur, 1996). The posterior descending pathway (PDP) from HVC is responsible for song production. It consists of neurons from HVC denoted  $HVC_{(RA)}$  that project into the premotor area RA (nucleus robustus archistriatalis or robust nucleus of the archistriatum). The RA in

turn innervates the tracheosyringeal half of the hypoglossal nucleus (nXIIts) that projects to the vocal muscles as well as the nucleus ambiguus/nucleus retroambiguus (nAM/nRAm) controlling respiration. Vates et al. (1997) showed there exists a reafferent pathway from RA back to HVc through the dorsomedial thalamic nucleus (DMP) which connects via the medial magnocellular nucleus of anterior neostriatum (mMAN) back to HVc. They postulated that this reafferent loop has properties similar to that proposed for the thalamo-cortical loops in humans and other mammals, thought to have a key function in sequencing action. However, they also suggested that this pathway is primarily for song maintenance and correction, or possibly for bilateral coordination of the song – the zebra finch syrinx has two halves that are controlled by ipsilateral vocal circuits and can operate separately - and therefore is not required for sequencing of song syllables.

HVc also connects to RA through the anterior forebrain pathway (AFP) which, in contrast to the PDP, is only required for song acquisition – ablation of this pathway in adults does not affect song recognition or production except over long time frames although it does halt the development of new seasonal songs. In immature birds, ablation of the pathway prevents the refinement of early vocal babbling into distinct songs (Troyer & Bottjer, 2001; Nottebohm, 2005). Female songbirds with bilateral lesions of HVc respond to heterospecific mating songs as well as conspecific songs indicating that there is some critical function in recognition of song characteristics associated with HVc. In the songbird brain, therefore, it appears that recognition and production of songs are separable from the circuitry directly involved in learning, and it is reasonable to focus on the circuitry specific to the HVc without regard to reafferent loops in investigating singing as a motoric sequential task. While studying how new songs are learned is an important area of research in its own right, and though there is some evidence that AFP is required for long-term maintenance of the integrity of an existing repertoire, this project is interested in timing and sequencing, hence the focus on HVc and the PDP.

## **Song Production in HVc**

Disruptions to HVc in male songbirds produce severe effects on song production. For example, stimulating HVc during singing can reset the song, suggesting that sequencing and temporal control of action over long time frames takes place within HVc. Sequence-specific neurons (SSNs) exist in HVc that are highly responsive to auditory playbacks of the bird's own song (BOS) but not to modifications of the song or to conspecific songs. Furthermore, neurons in afferent nuclei such as region L do not display a similar special tuning to syllable sequence (Lewicki & Arthur, 1996; see fig. 1a for schematic of structure

in bird brain and fig. 1b for correspondence with human structures) though they are tuned for amplitude and frequency modulations. For example, if A and B are two syllables of a BOS, SSNs in HVC will only show significant spiking activity to the sequence AB and not to AA, BB, or BA (see fig. 3). As well, intrasyllable reversals do not activate SSNs. SSNs do not exist in the RA, however, and there is further indirect evidence that temporal sequencing of syllables is controlled by HVC while the production of individual notes is governed by the RA (Yu & Margoliash, 1996; Spiro, Dalva, & Mooney, 1999). There is no evidence to date that RA has SSNs related to syllable production. It appears, then, that HVC creates the temporal structure allowing for the integration and production of signals over long periods of time. It acts as a trigger for smaller segments of a song produced by RA. For this reason and because the songbird is a simplified system amenable to experimental probes, HVC has been proposed as a useful starting point for modeling sequential action in other animals, including motoric timing in humans (Buonomano, 2005).

### *Three types of HVC neurons*

The three known types of neurons in HVC (see Kubota & Taniguchi, 1998, for evidence of four types of neurons) include those projecting into the RA, called  $HVC_{(RA)}$ , those projecting into area X, called  $HVC_{(X)}$ , and interneurons. Typical response of these neurons to BOS is shown in fig. 4 (Mooney, 2000). As we see from figure 4, and from Kozhevnikov & Fee (2006),  $HVC_{(RA)}$  neurons burst at most once during production of a motif. The burst duration is between 3.0 and 9.1 ms, is composed of 2.0 to 6.8 spikes each, and is between 330 Hz and 890 Hz. The initial spike for a particular  $HVC_{(RA)}$  neuron is tightly locked to syllable timing with root mean square time jitter of  $0.73 \pm 0.3$  ms. When the bird is not singing these  $HVC_{(RA)}$  neurons do not burst (less than 0.001 spike/s). As well, some  $HVC_{(RA)}$  neurons only appear to fire during the introductory notes.  $HVC_{(RA)}$  membrane potentials resting polarization changes at the start of the BOS, going from -76.5mV to around -65mV (Mooney, 2000), and maintains this depolarized state for some time after the song. Not shown in fig. 4 is that RA-projecting neurons fail to spike during playing of the reverse BOS however the membrane resting state still depolarizes. Action potentials from  $HVC_{(RA)}$  neurons to other HVC neurons typically evoked depolarizing post-synaptic potentials (Mooney & Prather, 2005) and the post-synaptic current from a single  $HVC_{(RA)}$  can induce spiking. The other synaptic connection to  $HVC_{(RA)}$  is from HVC interneurons connected via inhibitory post-synaptic action potentials.

Although  $HVC_{(X)}$  neurons can burst up to four times during a motif (Kozhevnikov & Fee, 2006), most are time locked to the motif. As well, these neurons burst for longer periods of time than RA-projecting ones, and continue to spike in awake nonsinging birds. Unlike RA-projecting neurons, the membrane of

X-projecting neurons does not depolarize during either the forward or reverse BOS but like the X-projecting neurons, there is no spiking during playback of reverse BOS.

The final type is the HVC interneuron. Interneurons have a spontaneous firing rate of 2-14 Hz when there is no song, and an evoked firing rate of 25-262 Hz (Kozhevnikov & Fee, 2006) during playback of BOS. They spike continuously during playback of BOS and minimally during the reverse BOS.

## Timing in Humans

Although interesting by itself, much of the interest in zebra finch song is because it can be used to model sequencing of action. There are numerous models of timing in humans implicating many parts of the brain depending on the type of task, interval, and state. For complex motor and sequential action evidence from Parkinson's patients (Harrington & Haaland, 1991), Huntington's patients (Gabrieli, 1995) and fMRI (Rao, Mayer, & Harrington, 2001; Nenandic, Gaser, Volz, Rammsayer, Hager, & Sauer, 2003) implicate the basal ganglia as a locus of timing; the striatum sends signals to premotor (supplementary) cortex which initiates distinct elements of the sequence. In this way, the basal ganglia acts as an overall timer for motor action while leaving shorter sequences under the control of motor and premotor areas.

How the striatum codes time is unknown, but it is reasonable to assume it is because of the unique nature of striatal neurons and architecture (Beiser, Hua, & Houk, 1997). The striatum is mostly composed of medium spiny neurons, each of which has 10,000-30,000 afferent cortical inputs per neuron. It has been suggested that the medium spiny neuron acts as a coincidence detector. It is also believed that this neuron is bistable, with the two states characterized as:

- down-state: hyperpolarized; -61 to -94 mV; resists spiking
- up-state: polarization at -71 to -40 mV; spiking facilitated; relaxes slowly to down-state

The transition to an up-state requires highly correlated wide scale cortical activity and (possibly) the suppression of the lateral inhibitory network that also makes up the structure of the striatum.

In rats, the depolarized state appears to last from 0.1-3 seconds (Wilson & Kawaguchi, 1996) or 284 ms (Blackwell, Czubayko, & Plenz, 2003), with an average of 217 synaptic inputs required to activate an up-state (Blackwell et al., 2003) although there is some evidence that other voltage-dependent mechanisms are at play (Wilson & Kawaguchi, 1996). Spiny neurons burst fire in typified ways during motor action, while remaining quiescent much of the time.

Bistable neurons are believed to exist in the mammalian striatum for awake and asleep animals (for example, Kitano et al., 2002; Stern, Jaeger, & Wilson, 1998). In addition, the change in bistable membrane potential is highly synchronized across striatal microzones although spiking is not coordinated within a particular microzone (Stern et al., 1998). This suggests that there is a mechanism which (a) produces a bistable condition, and (b) pushes neurons into the up state that can cover regions of striatum, not just individual neurons. However, there is a separate mechanism which induces any particular neuron to spike.

Dutar, Huan, & Perkel (1998) speculate that  $HVC_{(RA)}$  neurons undergo a change in electrophysiological properties that allows them to fire rapidly only around the time of singing. Indeed, there are many similarities between  $HVC_{(RA)}$  neurons and the medium spiny neuron other than just morphological. For instance, RA-projecting neurons have a hyperpolarized resting state of around -80mV and a depolarized 'up' state of -65mV (see earlier discussion; note that striatal states are also separated by about 15mV – Stern et al., 1998). Interestingly, the depolarized state appears both in response to the BOS and the reverse BOS, suggesting that the song sequence is not coded into the mechanism that initiates the transition from down to up state. There may well be something which is similar to the combination of dopamine and input current in the mammalian striatum. This depolarized state in  $HVC_{(RA)}$  neurons then persists beyond the end of the BOS, suggesting a hysteresis similar to that found in a bistable system. The short time-locked bursts and the nearly complete lack of spiking otherwise are also characteristic of both types of neurons. So, while the neurons are not the same and the identified region in the bird brain is cortical rather than basal, perhaps features and architecture of one (RA-projecting neurons) can be used to model potential uses of the other (mammalian medium spiny neurons).

## System Description

Here is a summary of the salient points.

### Connectivity:

1.  $HVC_{(RA)}$ -  $HVC_{(RA)}$  connections are unidirectional and depolarizing.
2. INT-  $HVC_{(RA)}$  connections can be either direction.
  - a.  $HVC_{(RA)}$  to INT connections are excitatory, and a single spike in the  $HVC_{(RA)}$  neuron is able to evoke a spike in the interneuron.
  - b. INT to  $HVC_{(RA)}$  connections are inhibitory and there are multiple such connections per

- c. There are some reciprocal connections.
- 3. INT-INT connections are unidirectional and inhibitory.
- 4.  $HVC_{(RA)} - HVC_{(X)}$  connections could be unidirectional or reciprocal, and could have either inhibitory or excitatory post-synaptic responses.

#### **Firing characteristics:**

- 5.  $HVC_{(RA)}$  neurons can integrate a signal over a long time (Kubota & Taniguchi, 1998) and may be sensitive to harmonics (Margoliash & Fortune, 1992).
- 6. At least some  $HVC_{(RA)}$  neurons are tightly locked to syllables but *some or possibly many* are not. Although not stated above, it is also possible that not all  $HVC_{(RA)}$  neurons actually project into RA and do not produce signals that get interpreted as a timing code by RA.
- 7. Interneurons fire tonically during singing of the BOS but not the reverse BOS (at least fire at a much slower rate).
- 8. Interneuron firing patterns are weakly correlated with each other, 0.14 (Kazhenikov & Fee, 2006) and with patterns of sound in the song (0.08) where it leads by 30-50ms.
- 9. The auditory input signal is presumed to come from a population of neurons that are tuned to the signal at any appropriate level. The neural dynamics of this population have not been shown to encode the timing.

## **Design Specifications**

Imagine we have a sequence of subsyllables or notes,  $A_1$  to  $A_4$ . The model is as follows:

- 1. The timescale of neural dynamics in HVC is 5ms (Kozhevnikov & Fee, 2006). I'm not sure, however, what this means for the various dynamics of the neurons and these parameters will have to be adjusted as required.
- 2. The timescale of syllable dynamics is 50-250 ms. We may be able to use the shorter notes instead with a timescale of only 30-70 ms. Because of this, and item 5. in the design specifications, I am going to assume that a note can be interpreted by the neuron as a single event.
- 3.  $HVC_{(RA)}$  neurons burst fire at a rate of 330-890 Hz only once during a specific syllable in a motif. A proportion of these are song and syllable specific and we will denote these as SSN-RA neurons. Other  $HVC_{(RA)}$  neurons will be denoted EUN for excitatory unspecific neurons.
- 4. Interneurons fire tonically at a rate of 25-262 Hz during BOS.
- 5. Temporal accuracy of the signal to RA is required to be on the order of 1ms or less.

## Implementation

Each SSN-RA neuron is tuned to a specific frequency of inputs such that it burst fires at a specific point of the syllable (often near the end). Each microzone or small population of SSN-RA neurons are therefore tuned to one particular pattern of inputs in frequency space (I will call this the time-frequency domain from now on and Tripp & Eliasmith refer to the encoding by the weights as temporal basis functions) which can be characterized for syllable A by the function,  $f_A(t)$ . Strictly speaking  $f_A$  could be multiple-valued in the frequency domain since harmonics are part of the birdsong, however this unnecessarily complicates the situation so we will make the simplifying assumption that  $f_A: \mathfrak{R} \rightarrow \mathfrak{R}$ . The output function of the SSN-RA neuron is equivalent to a step function,  $g_A$ , where the step occurs at the end of the syllable for a very short time. Following Tripp & Eliasmith (2007), an SSN-RA neuron can preselect for this pattern by weighting synaptic inputs in such a way as to minimize the error term,  $E$ , given by:

$$E = \int_0^T \left[ g_A - \sum w_i P_i(t) \right]^2 dt$$

In this equation,  $w_i$  is the weight of the  $i$ th synapse and  $P_i$  is the unweighted post-synaptic current. While I will not discuss how these weights develop in the zebra finch, it is most likely due to error monitoring signals driven by HVC<sub>(x)</sub> neurons during the days post-hatch when the juvenile zebra finch develops its songs.

Regarding temporal coding by a single neuron or small population of neurons such as a microzone, Tripp & Eliasmith note the following:

1. The ability to represent any arbitrary function over time depends on presynaptic firing patterns. In particular, *regular* presynaptic firing patterns were associated with a greater error term,  $E$ , because regular patterns preselect certain regions of the time-frequency domain. If there were a mismatch between  $f_A$  and  $g_A$  in the time-frequency domain, then a regular firing pattern would not allow for a good transformation of the signal. Thus, it makes sense to have interneurons and EUNs that do not fire in a regular pattern. If an SSN-RA neuron were to attempt to transform  $f_A$  to  $g_A$  directly, the output ( $\widetilde{g_A}$ ) is likely to be much noisier than if it, instead, was provided with input spikes that effectively tile (or form a basis for) the time-frequency domain.
2. Error increases with increasing spike time correlation. Interneuron spike time correlation was in fact very low, 0.14, as we would predict.

3. Patterns could be transformed on the order of hundreds of milliseconds. This is critical because the timescale of interest in our case is up to 250 ms.
4. The technique of extracting  $g_A$  maintains a low mean-square error (MSE) even when spike-time jitter and extraneous spikes were included in the model. Even under considerable noise (half the spikes were extraneous and the ones that were not were subject to up to 20 ms time jitter) the resulting pattern was still fairly accurately reproduced.
5. Large populations of presynaptic neurons were not required for high accuracy. It is unclear how this relates to the model for now because I have not found data estimating the number of INT-SSN-RA connections. However, given the extent of the dendritic tree for  $HVC_{(RA)}$  neurons it would be surprising if the number of post-synaptic connections was small.

Next, correlations between interneuron and EUN firing and the BOS is very low, 0.08. Tripp & Eliasmith<sup>1</sup> predict this when they note that firing patterns can be difficult to distinguish from random firing. This effect is exacerbated by the relative imperviousness of the process to time jitters and random spiking. Therefore, the combination of creating an irregular pattern to allow for coding and introducing noise most likely explains this apparent low correlation. Interestingly, there is some subthreshold activity of SSN-RA membranes during BOS but prior to the time-locked firing. A model that includes noise would predict exactly this.

However, once the neural network is trained to respond to the given input with a threshold response, it has additional two properties that make it insufficient to explain how sequential coding works:

- i. The same threshold response occurs whenever the latter part of the input signal is presented. In other words, assume that a signal is actually made up of the syllables ABCD and a response is only required at the end of D, then anytime D appears there will be a threshold response.
- ii. The neuron only fails to respond to A, B, and C. Any other arbitrary input will produce a response (almost always, a.a., in the measure-theoretic sense of the term). It is not possible to overcome this limitation by training the neurons to fail to respond to many possible input signals as the Hilbert space is infinite in extent. If the possible inputs are extremely limited, however, the Tripp & Eliasmith network could model the final threshold output. As we shall see in the next section, this scenario is a possible consequence of the model which produces sequencing behaviour.

---

<sup>1</sup> Interestingly, the paper cites songbird vocalization as an example of this (p. 1868).



## Sequencing

We need a way to keep the useful properties (1-5) of the SSN while overcoming limitations i & ii. In the songbird HVC, this appears to be solved by having a network that is *only active* during the playing of the song. The EUN and interneurons must spike in response to sequence AB, but not after AA, BA, or BB. There are examples of this in neural network models. For example, in Chapter 6 of his PhD thesis, B. Tripp proposes a model that uses the bistability of Gruber neurons to create such a sequenced response. Each Gruber neuron is conceptualized as having an up-state tuning curve that is larger and encompasses the down-state tuning curve. Each element in the sequence is represented by one set of neurons, with the high-state tuning curves for this state encompassing all the subsequent up-state tuning curves in the sequence. Thus, as the trajectory of the system moves through the state space, more neurons are activated while none of the previous neurons are deactivated. At the end of the sequence, all neurons have been activated and this combined signal causes post-sequence neurons to fire (Tripp proposes that the post-sequence neurons are inhibitory, but I don't believe this is essential). While this model is possible, and while some HVC neurons have many of the properties of Gruber neurons, there is no experimental evidence for this sort of spreading activation as a sequence progresses.

A second model for sequential action uses a synfire chain, i.e., a wave of neuronal activation that moves through layers of neurons. Each layer represents one element of the sequence. While synfire chains do not require any special neural dynamics, there is no evidence for layered activation in HVC.

The challenge with both models described above is that they are unsupported by observations of HVC neural dynamics. As well, neither specifically explains how neurons can react exclusively to sequential external inputs. While this is a similar problem to creating a sequence, it is not the same, and reacting to external inputs may be the more difficult neural problem to solve because it requires both a sequence and a continual check to see if the input is the same as the predicted sequence.

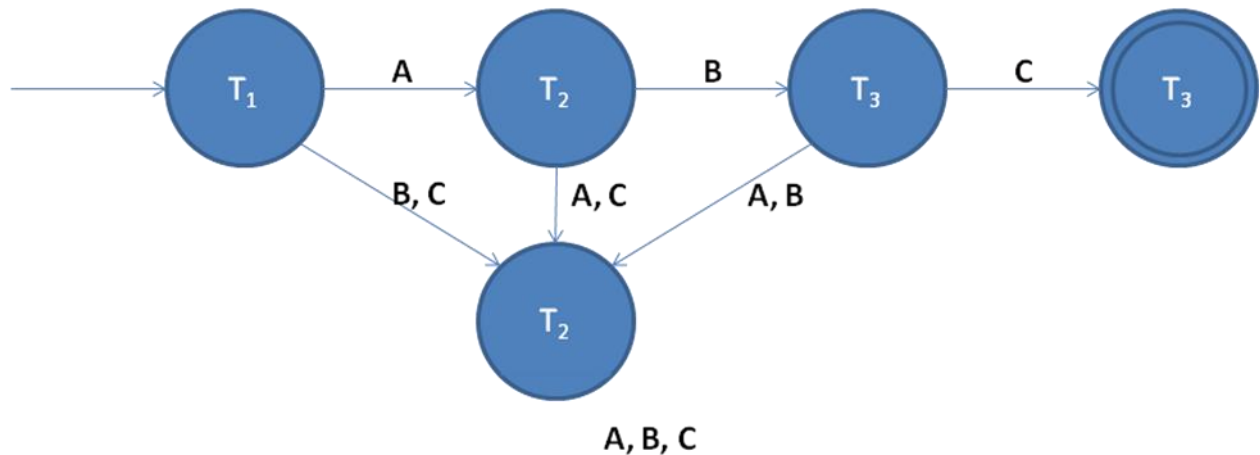
## Elements of my Model

Consider the simplest model consisting of only three sequential elements, A, B, and C. Here is what each element must do:

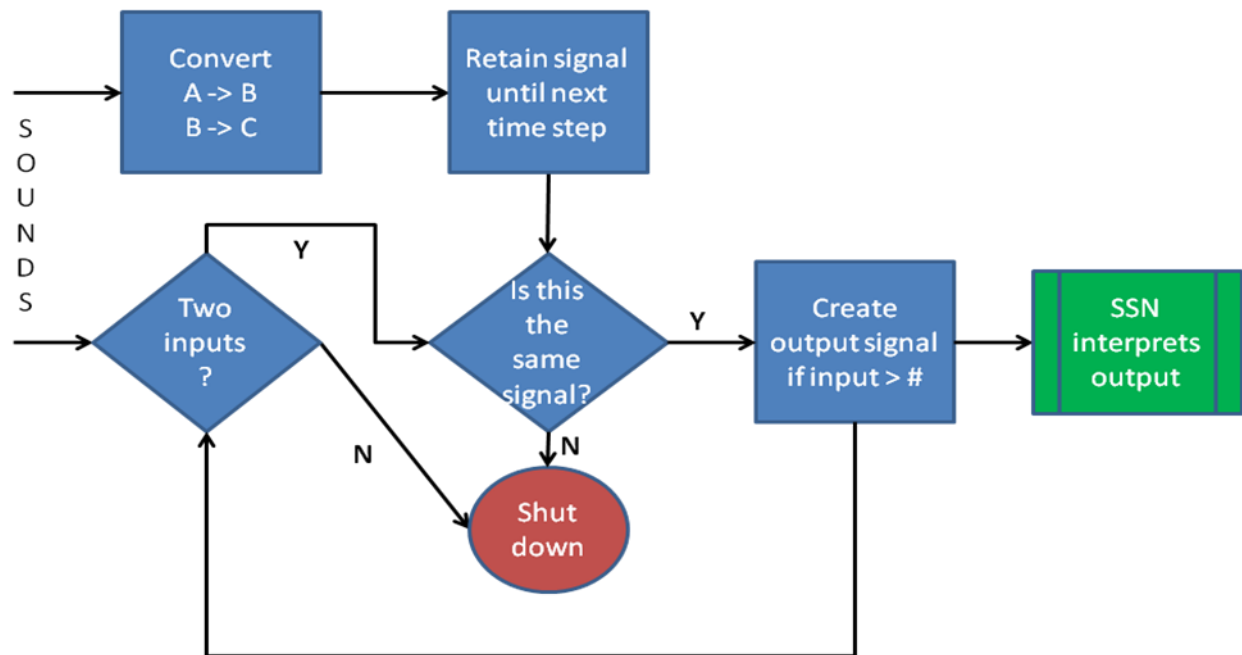
Element	Appearing at...		
	Time 1	Time 2	Time 3
A	- Start system - Prepare system to receive B at next interval	- Shut down system	- Shut down system
B	- Shut down system	- Keep system running - Prepare system to receive C at next interval	- Shut down system
C	- Shut down system	- Shut down system	- Signal output

As well, once the system is shut down, it should only be able to restart with the re-presentation of A.

The state diagram for this is given by

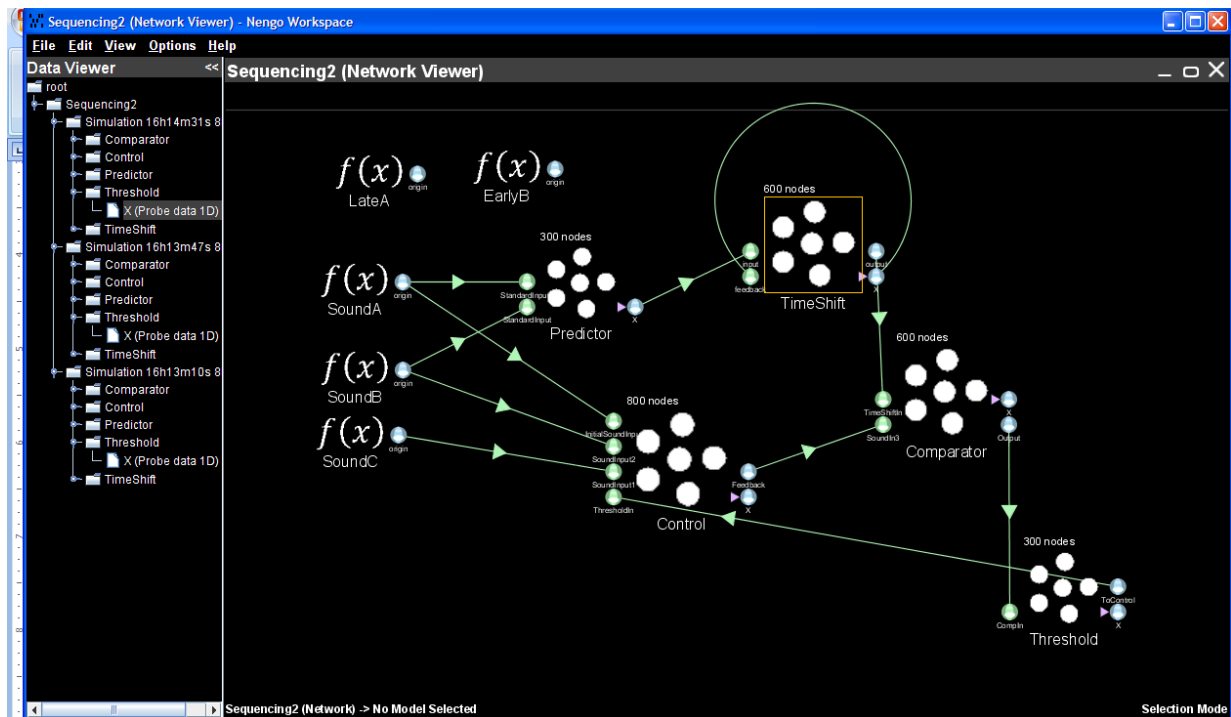


Where  $T_i$  means time  $i$ . As we have seen, multiple possible neural networks can perform the above computation. One that respects the experimental data on HVC is as follows:



From an implementation perspective, one problem with this schematic is that the system shuts down on the very first sound input (because there is no input from the Convert-Retain path until the second sound). Therefore, an actual implementation will need to give a special role to the first sound, or create introductory notes, that prevent shut down at  $t=1$ .

The following is a screenshot of the Nengo model which implements the above ideas. I will send you a copy of the model (Note: the file is about 52M – I will try to send it, but if not, I've loaded it onto my memory key and will drop by with it next week).



The Nengo model does not include the response to the final output signal coming from Threshold. There are a number of ways of doing this, but one which produces very precise timing even with some jitter in output is the Tripp & Eliasmith burst firing neurons. The SSNs then need only be trained to respond to the final input signal and not to respond to a finite set of other responses, i.e., it fires in response to the final element of the sequence ABC..., but not to any incomplete version such as A or AB. Since the system is shut down otherwise, the SSN neurons need only shut down for A, B, C, etc., until the last element in the sequence appears.

From now on, I will consider a sequence of length  $N$ , with the first element called A, the final element called C, with intermediate elements  $B_i$ . Each sound is conceptualized as occurring in a space of dimension  $N$ , where each dimension refers to one independent component of the temporal basis function (A,  $B_1$ ,  $B_2$ , ..., C). For example, (0,1,0,...,0) would represent  $B_1$ . The various components of the Nengo model are:

Element	Description	Inputs	Outputs
Predictor	N-di neural system. It acts like a Markov chain, i.e., given $B_i$ , it produces $B_{i+1}$ . Each input persists for a set time and then stops. C does not need need to connect to this circuit, though doing so has no effect.	<i>StandardInput</i> : ( $N \times 1$ ) vectors input from any of the sounds, coupled to Predictor with an ( $N \times N$ ) matrix consisting of 1's in the off-diagonal immediately below the	X: ( $N \times 1$ ) vector output to TimeShift with identity component of signal in each dimension.

		diagonal and 0's everywhere else. This transforms an input at time $t$ into the next element in the sequence, but still at time $t$ . The output therefore 'predicts' what should come next in the sequence.	
Time Shift	This is an N-di integrator circuit with feedback of 0.95, i.e., it integrates each dimension incompletely and then decays back to a resting state. Choosing the right feedback constant allows the predicted signal from the Predictor circuit to exist in time $t+1$ (which is when it should be input from the external BOS), but only minimally time $t+2$ . Because it is an integrator, the predicted signal will also appear near the end of time $t$ .	<i>Input:</i> (N x 1) vector. Note that the coupling constant is the same as the PSC, as required for NEF integrator circuits. <i>Feedback:</i> Vector of length N. The coupling constant of 0.95 makes this an incomplete integrator with decay.	X: Simple feedback for integration. Output: (N x 1) vector output to comparator with identity component of signal in each dimension
Control	(N+1)-di system. The first N-dimensions correspond to the sequence, and the N+1 <sup>st</sup> dimension is the degree of activation of the system (Threshold input or initial input). This final dimension will be used as the "Shut Down" control element seen in the schematic.	<i>InitialSoundInput:</i> (N x 1) vector which accepts the input of the first sound and transcribes it to the N+1 <sup>st</sup> dimension. <i>SoundInput:</i> (N x 1) vector which accepts the input of any other sound except the initial one. The input signal is coupled to an (N x N+1) identity matrix. <i>ThresholdIn:</i> Scalar which takes output signal from Threshold and puts it in the N+1 <sup>st</sup> dimension. A coupling constant of 0.72 keeps the input signal at about the right level (though the value of it is not that sensitive as long as it is big enough).	<i>Feedback:</i> N-di vector which is the product of the N+1 <sup>st</sup> dimension with each of the dimension 1, 2, ..., N. In other words, if there is no initial input or if the Threshold response drops to zero, this eliminates further output from Control. It is the "Shut Down" control element seen in the schematic.
Comparator	This is a 2N-di system. The first N-di are the time-shifted inputs. For example, if A is input to the Predictor at time 1, the Predictor outputs B to TimeShift which makes B persist into time 2 (in the first N-dimensions). The last N-di are used to store output from the comparator. A signal is only considered active if there is a signal in both the $j^{\text{th}}$ and (N+j) <sup>th</sup> dimension.	<i>TimeShiftIn:</i> 2N-di coupling matrix that puts the N-di input from TimeShift into the first N-di. <i>SoundIn3:</i> 2N-di coupling matrix that puts the N-di input from Control into the last N-di.	<i>Output:</i> A scalar constructed by summing the products of the $j^{\text{th}}$ and (N+j) <sup>th</sup> dimensions. This ensures a signal passes through to the Threshold if any signal is the same in TimeShift and Control.
Threshold	A collection of neurons tuned only to a portion of the range (0.5 to 1.0,	CompIn: A scalar that enhances the input signal	<i>ToControl:</i> Scalar representing the degree of

	compared to -1 to 1 for all the others). Thus, it is only activated by a sufficiently strong input signal.	(couples with 10.0). This coupling constant is mostly because of signal loss earlier on.	activation of the system. Required to keep the Control signal non-zero. X: Output to SSN
A	An N-di vector with 1 in the first component and 0 elsewhere. It is defined as a step function over a length of time equal to the length of the sound input.		Connects to predictor. Because it is the initial input, it must also connect to the N+1 <sup>st</sup> dimension of the control to indicate that the system is active (since there is no Threshold signal initially)
B <sub>i</sub> (B in the example)	An N-di vector with 1 in the (i+1) component and 0 elsewhere. It is defined as a step function over a length of time equal to the length of the sound input.		Connects to predictor.
C	An N-di vector with 1 in the last component and 0 elsewhere. It is defined as a step function over a length of time equal to the length of the sound input.		
LateA	An N-di vector with 1 in the second component and 0 elsewhere. To show what happens with the sequence AA, replace all B connections with LateA.		
EarlyB	An N-di vector with 1 in the first component and 0 elsewhere. To show what happens with the sequence BB, replace all A connections with LateB. Note: if you want to see what happens with BA, replace use both LateA and EarlyB.		

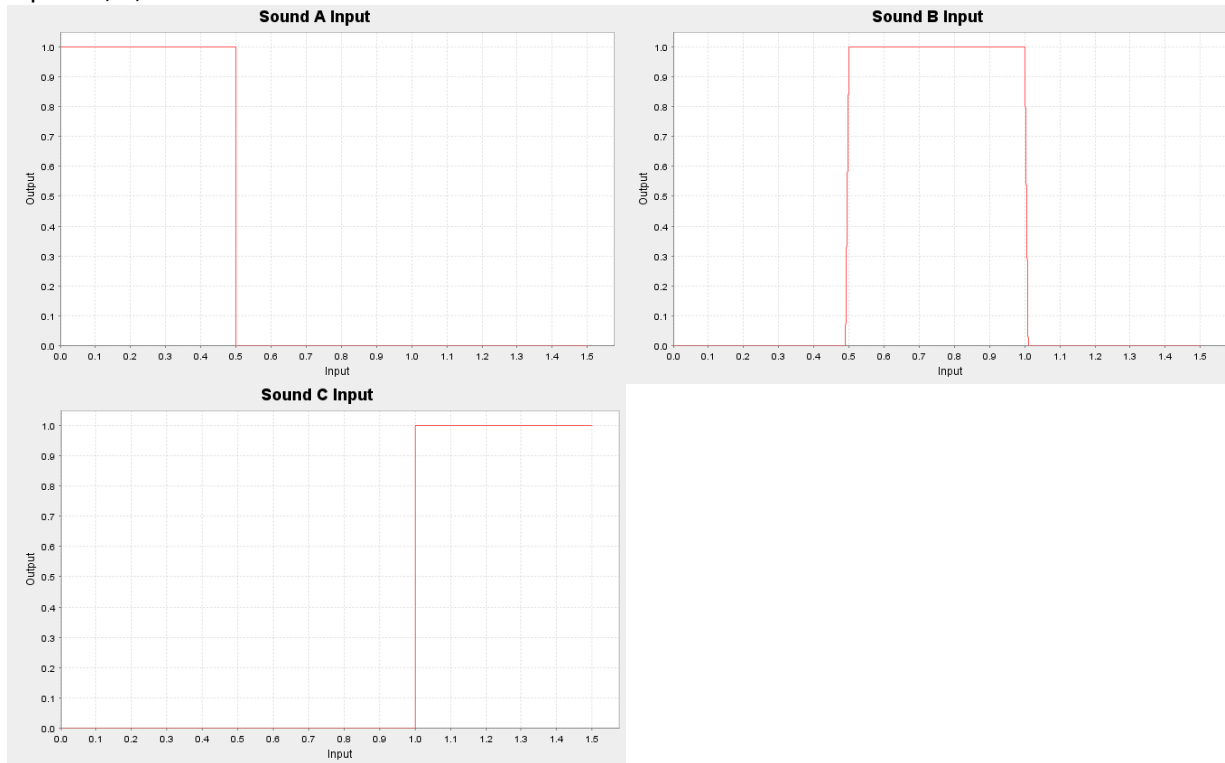
We can see that the above system is able to handle an arbitrary number of items in a sequence and, though it will not be shown here or in sample output, is reasonably robust over different timescales. It also does so with a minimum of neurons – the example I am running has a combined total of only 2400 neurons and works perfectly well. In fact, it is relatively insensitive to changes in the number of neurons. The one aspect that could require adjusting depending on the precise temporal dynamics of each element is the leaky integrator constant of 0.95.

The neurons in all parts of the simulation were tuned to (-1,1), with firing rates of 25-262 Hz. While these characteristics were only valid for interneurons (no specifications were available for EUNs), the model is not particularly sensitive to the firing rate.

### Output from the ABC example

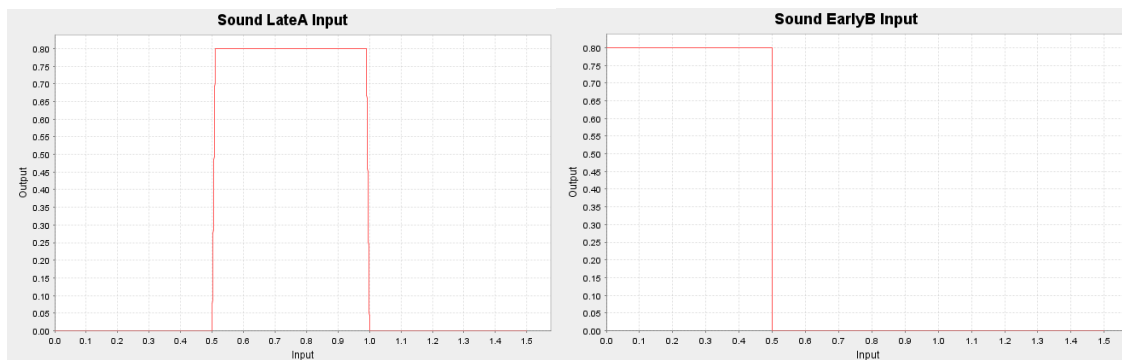
(Note: for the purposes of this example, the simulation was run for 1.5 s.)

Inputs A, B, C



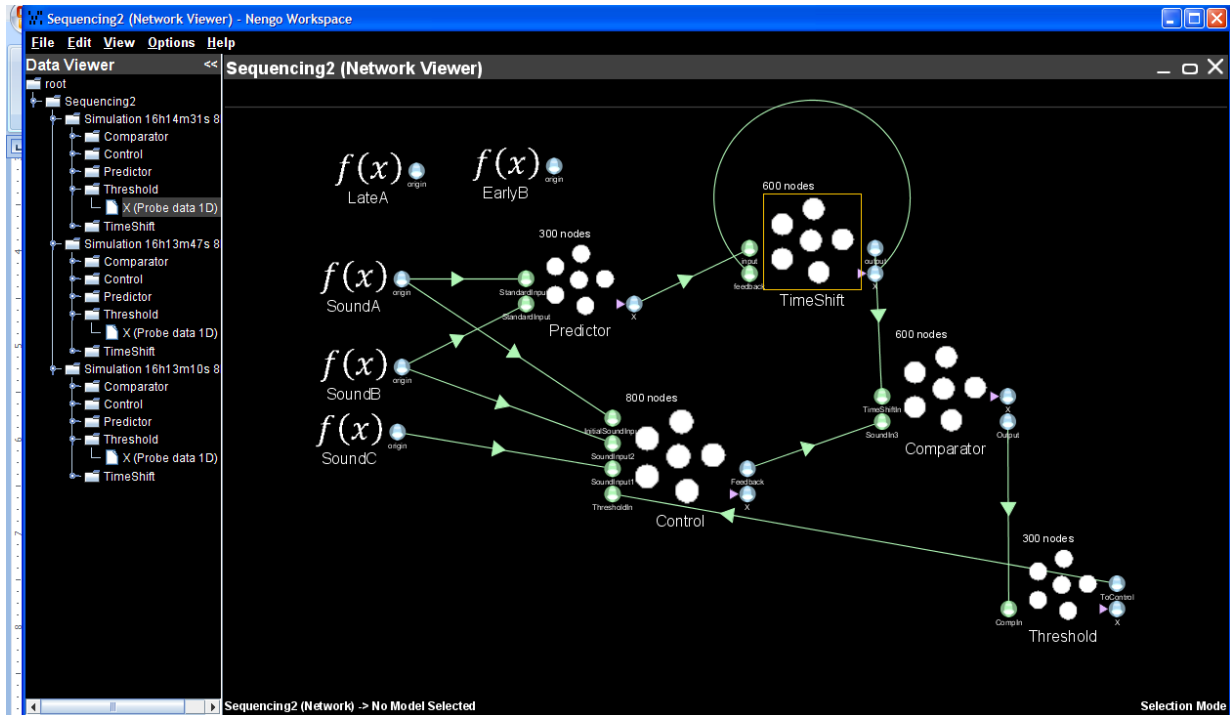
Note that I am only showing the non-zero dimension. Each sound input is a 3-di vector. For A, I have only shown the first component, for B only the second, and for C only the third. All other components are identically 0.

The non-standard inputs are LateA and EarlyB, with non-zero components in the first and second components respectively:



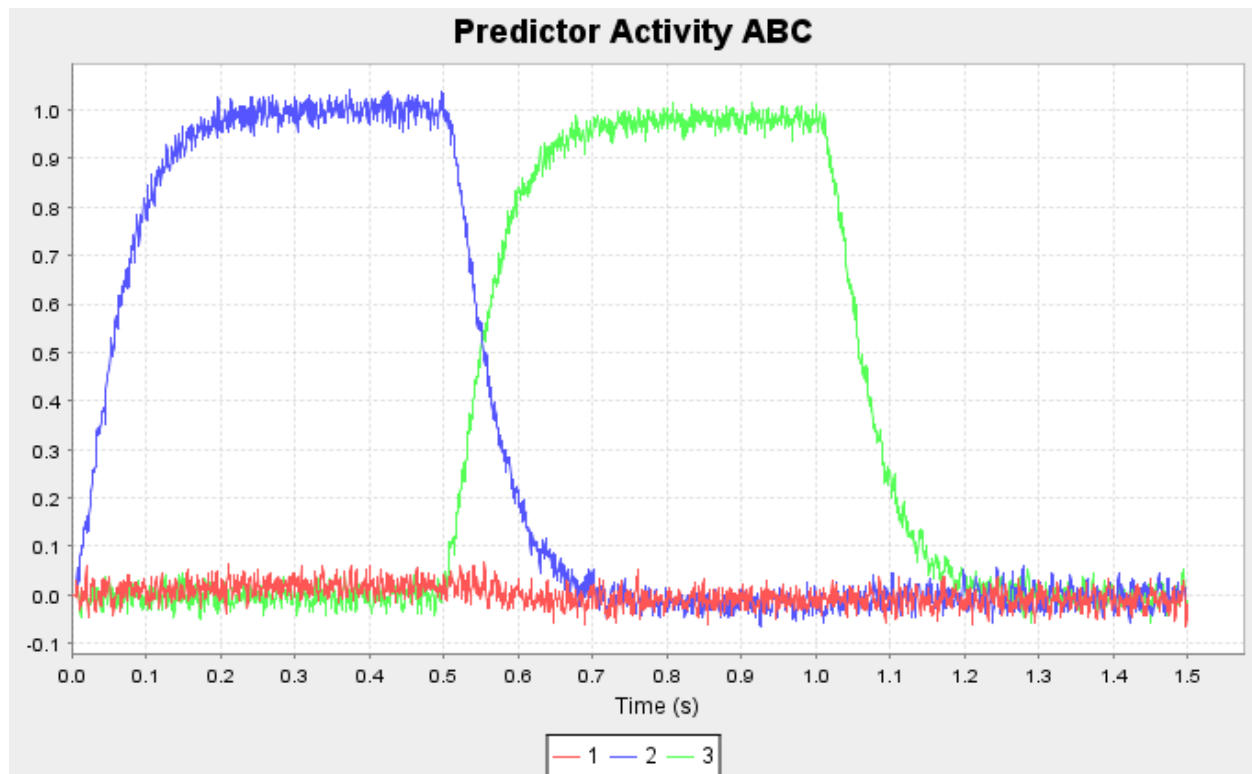
These are identical to A and B other than their positions in the sequence.

## Running the ABC Sequence:

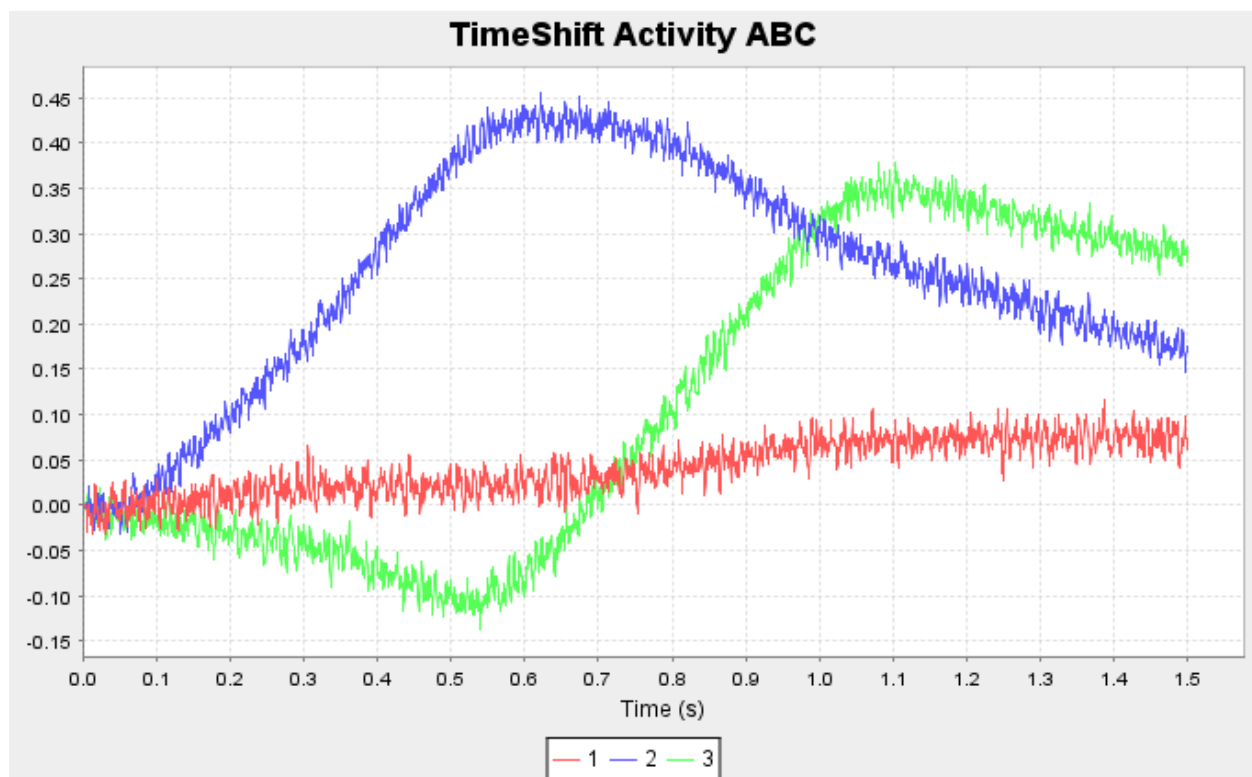


We obtain the following results:

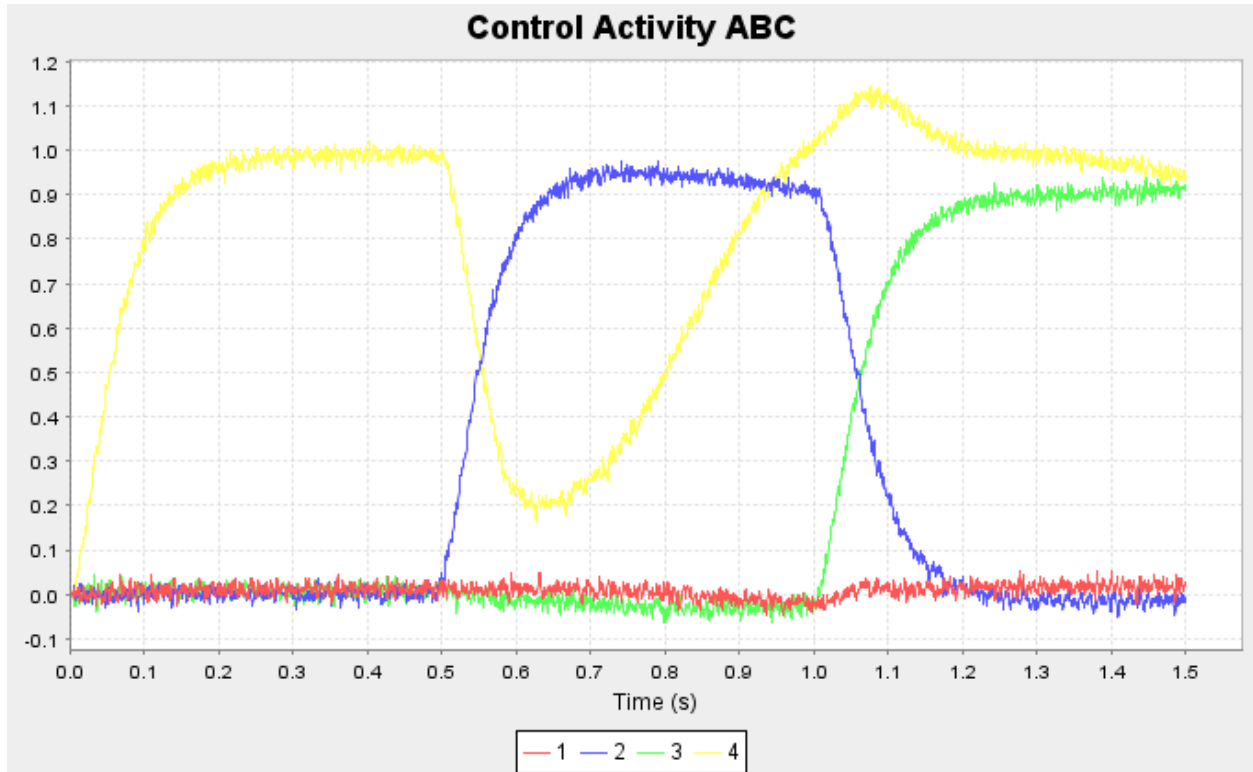




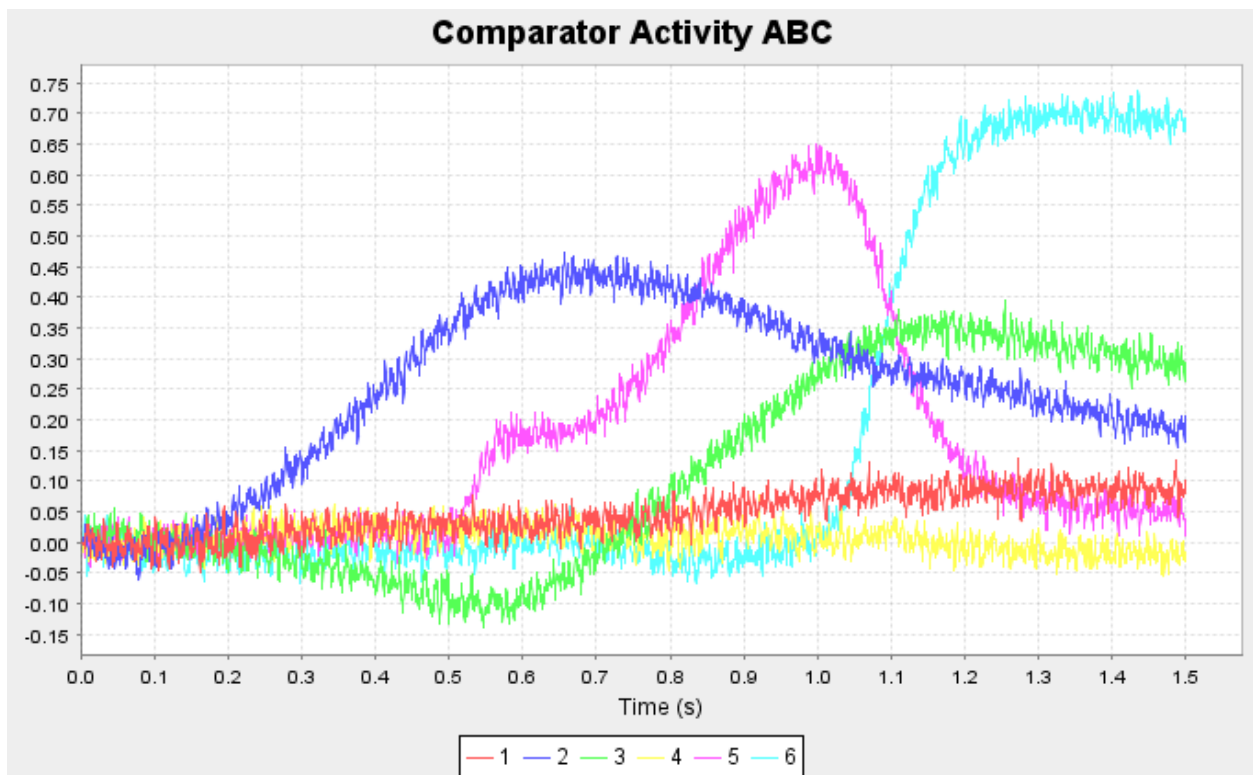
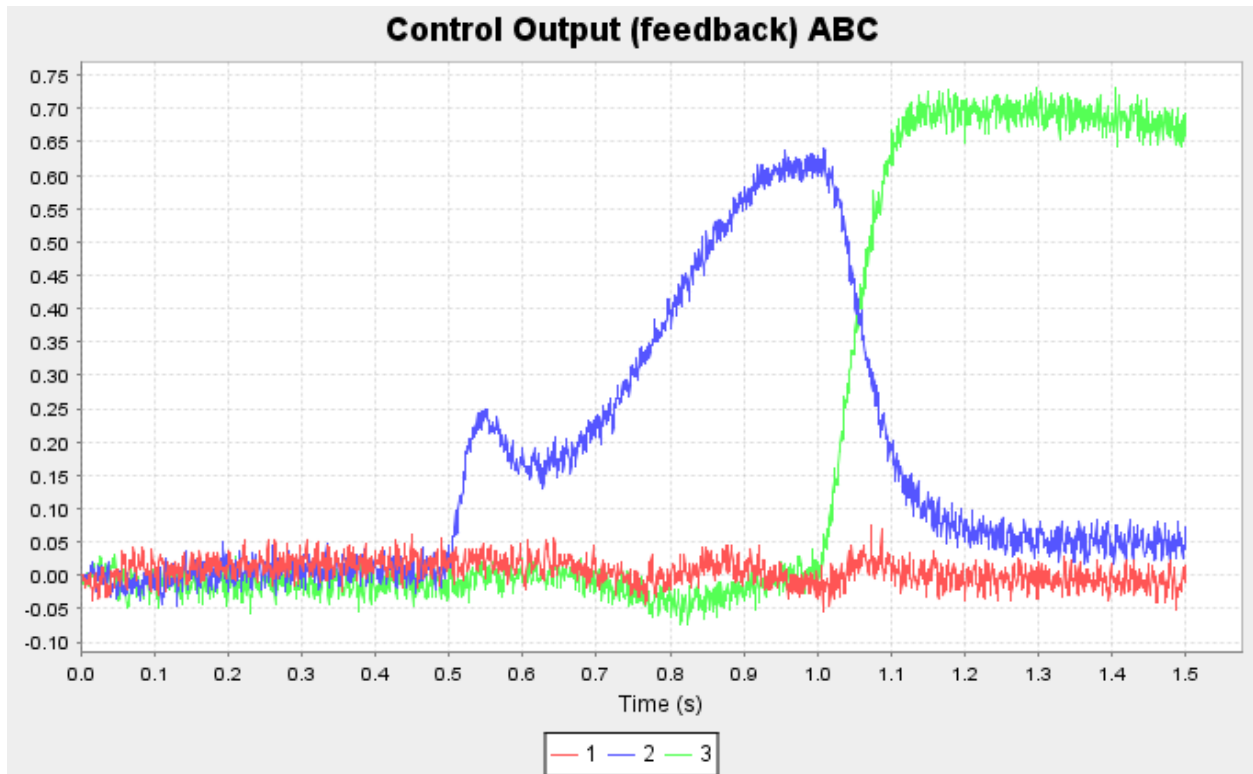
The red line is A, blue line is B, and green line is C.



The red line is A, blue line is B, and green line is C.

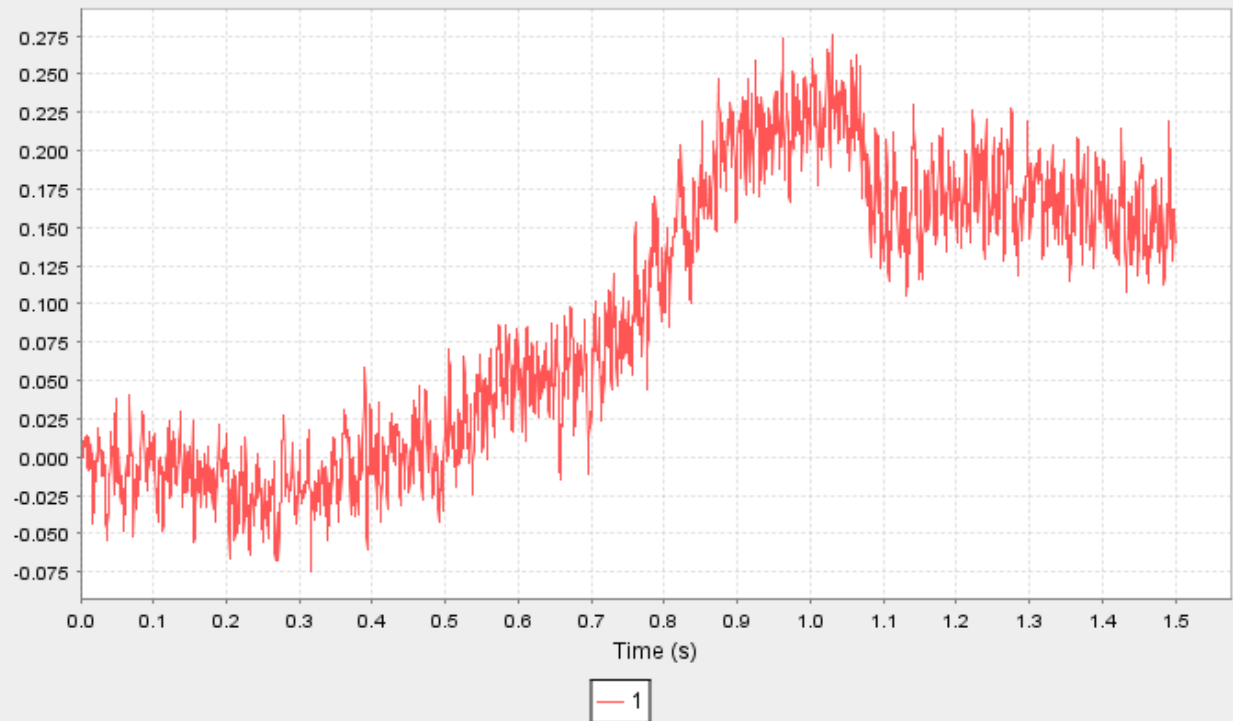


The red line is A, blue line is B, and green line is C, and the yellow line represents the shut-down control element. If the yellow line is ever at zero after the initial input (first 0.5 s), then the system stops responding to input. Essentially, the output from the control is each of the red/blue/green lines multiplied by the yellow line. The output looks like:

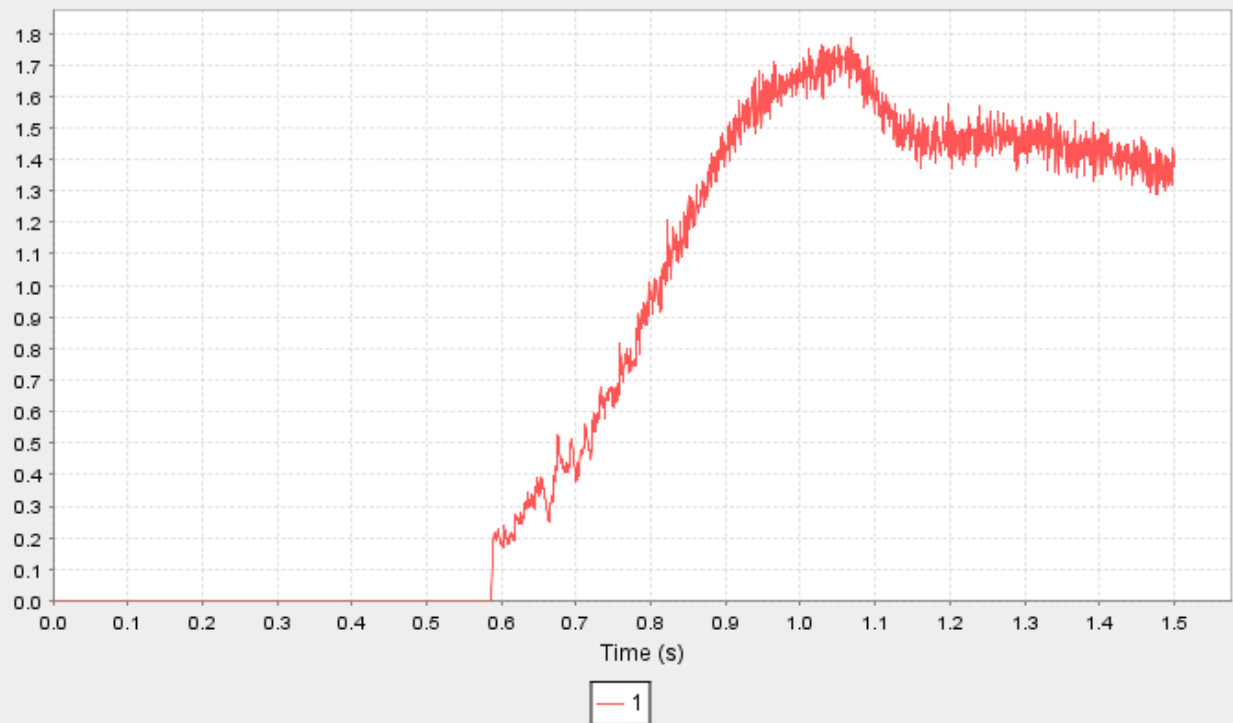


The dimensions compared (multiplied on output) are 1-4, 2-5, 3-6, and their products summed. The output looks like:

**Comparator Output to Threshold ABC**

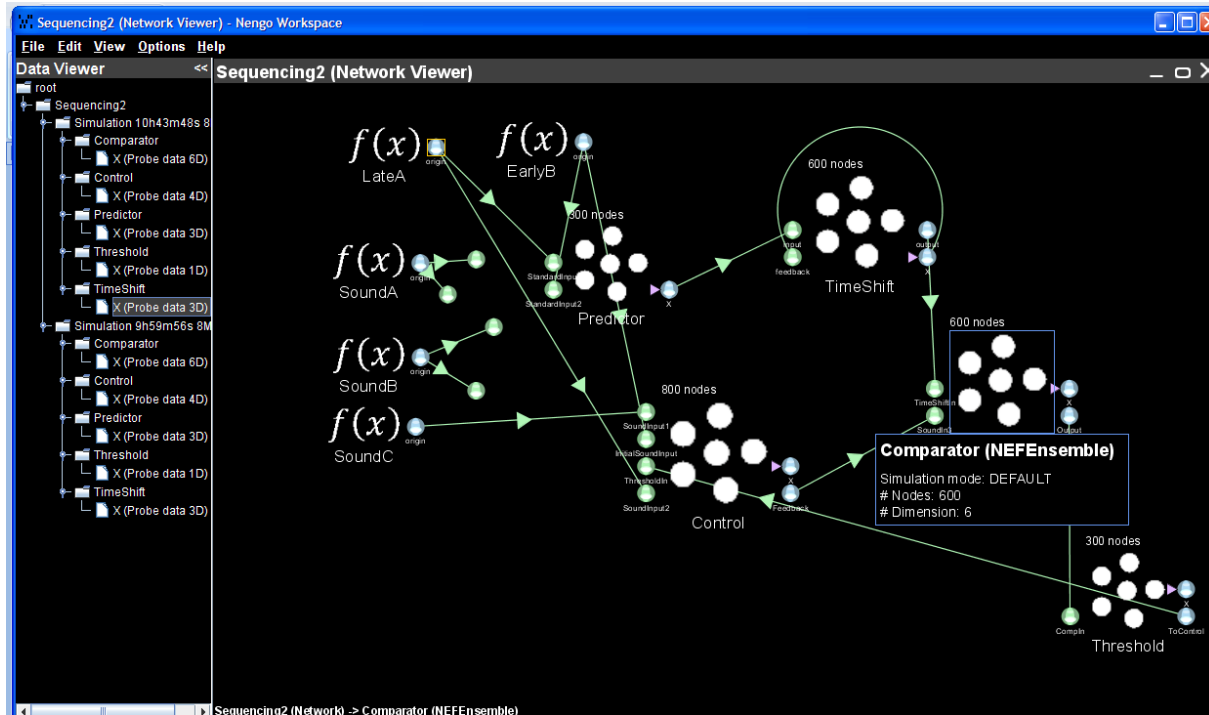


**Threshold Output**



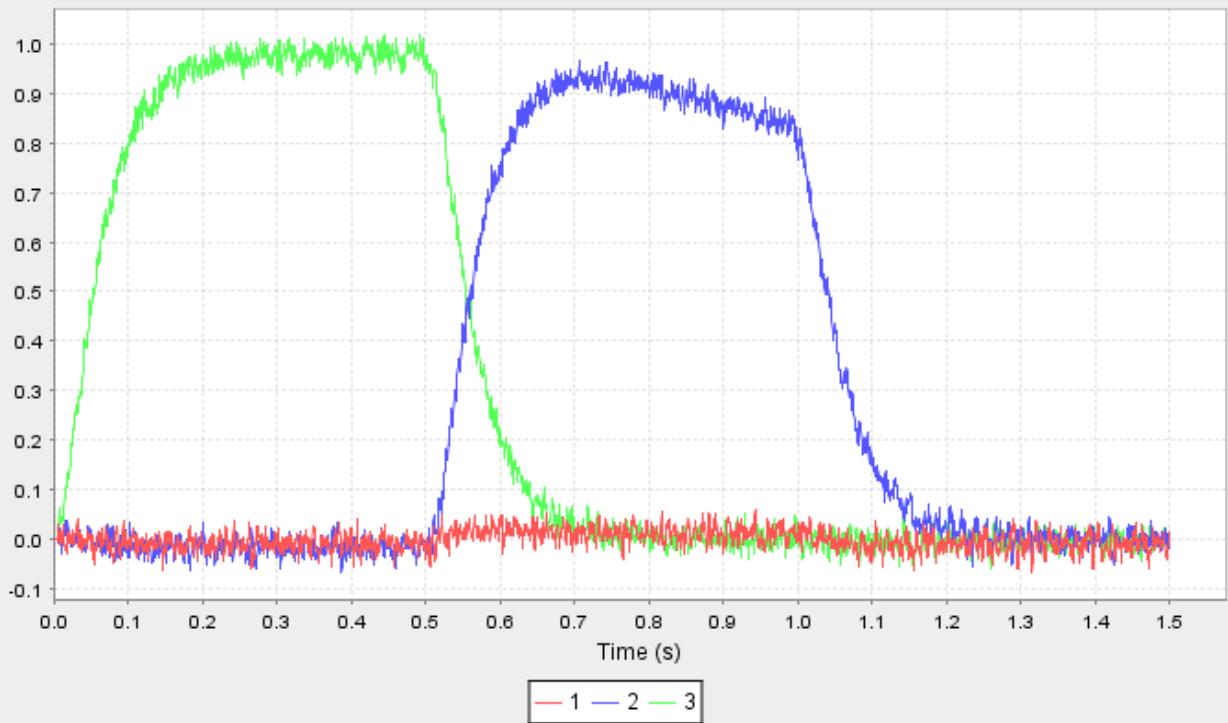
This represents the final output from the system. It is the signal interpreted by the SSNs. As we can see, it is non-zero after the initial sound (A). The tuning curves for this population are from 0.5-1.0 only.

## Running the BAC Sequence

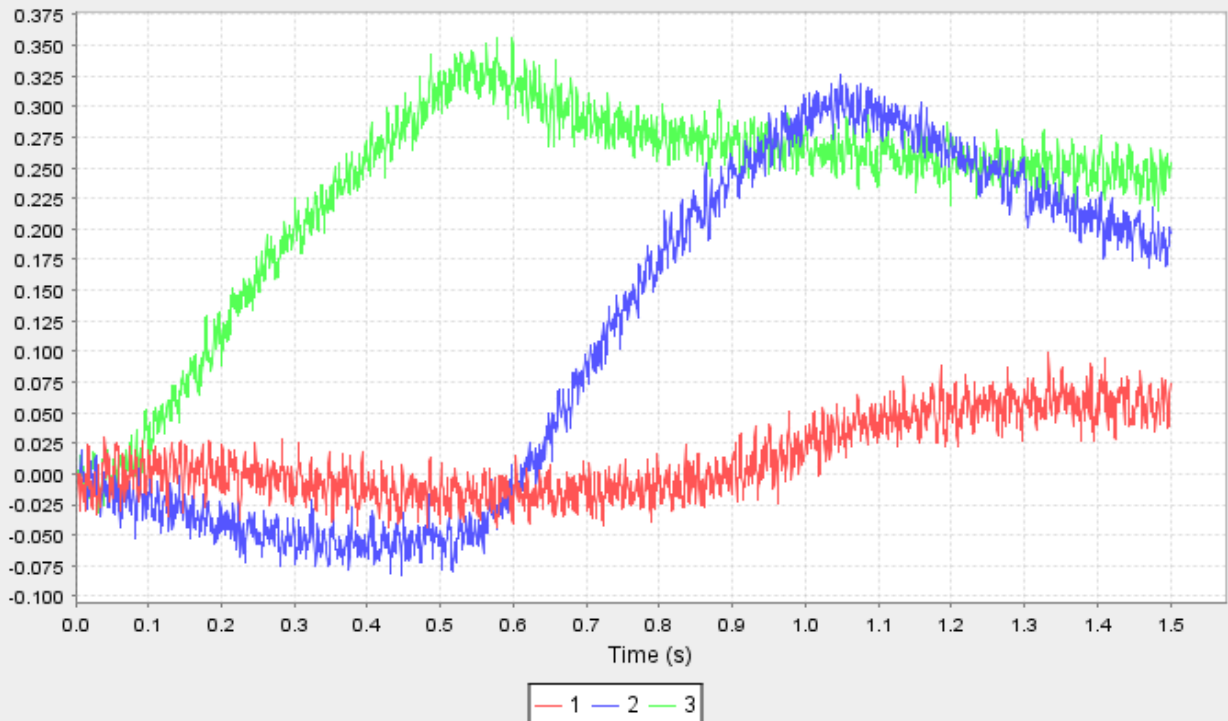


The following activity is seen in each component:

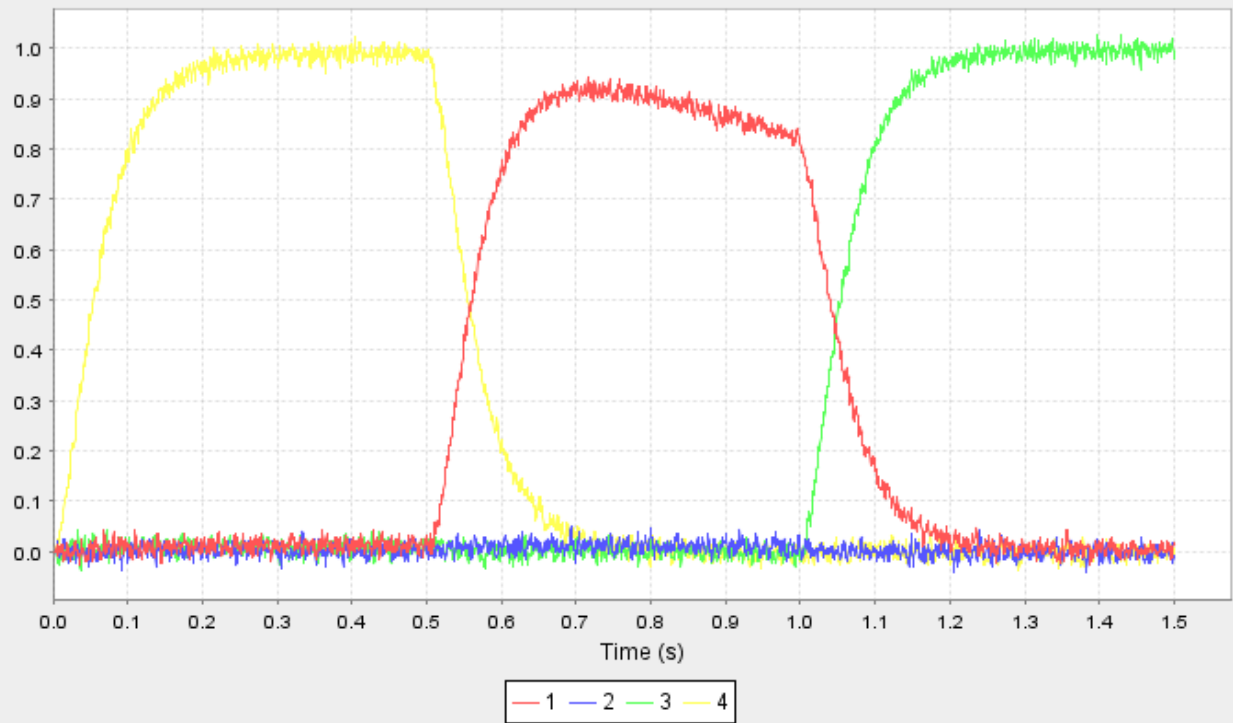
**Predictor Activity BAC**



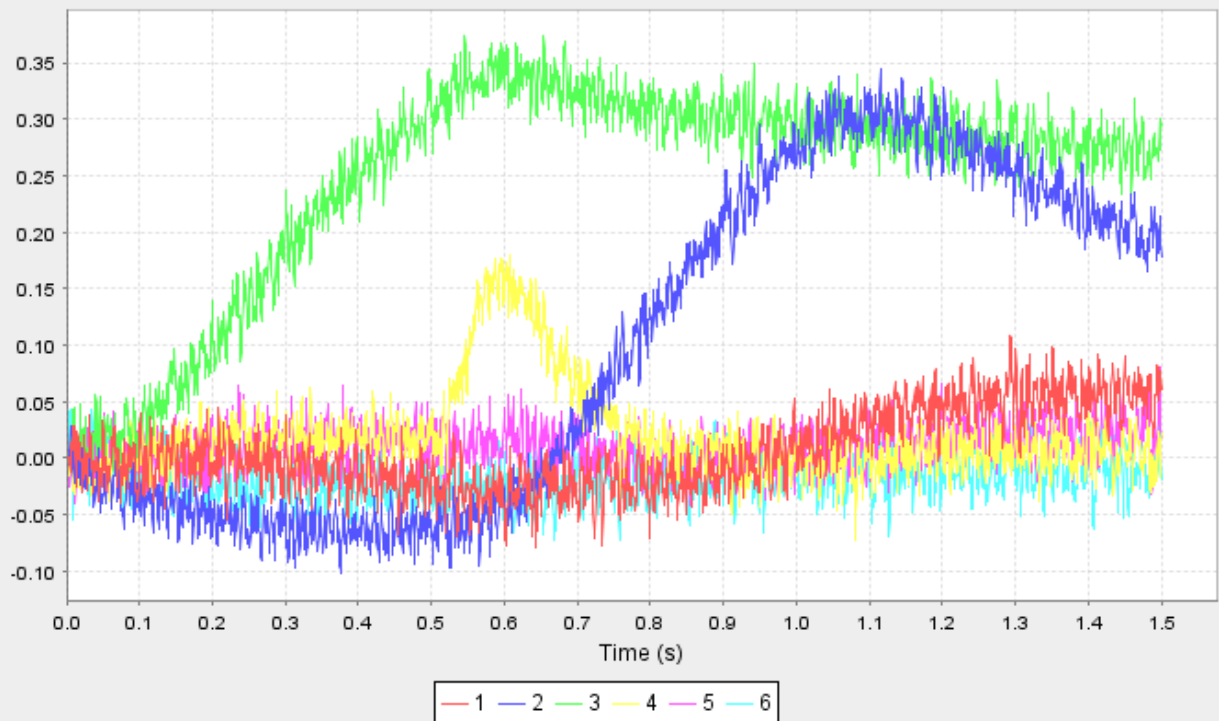
**TimeShift Activity BAC**

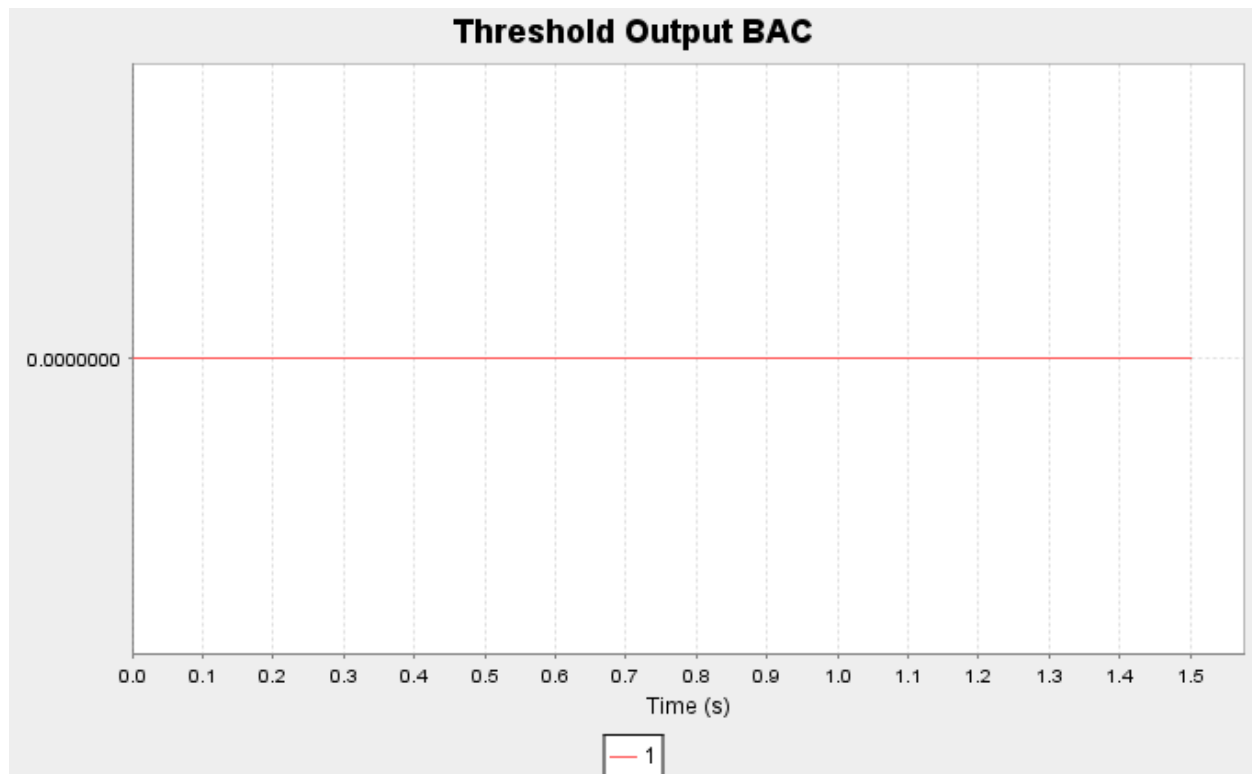


**Control Activity BAC**



**Comparator Activity BAC**

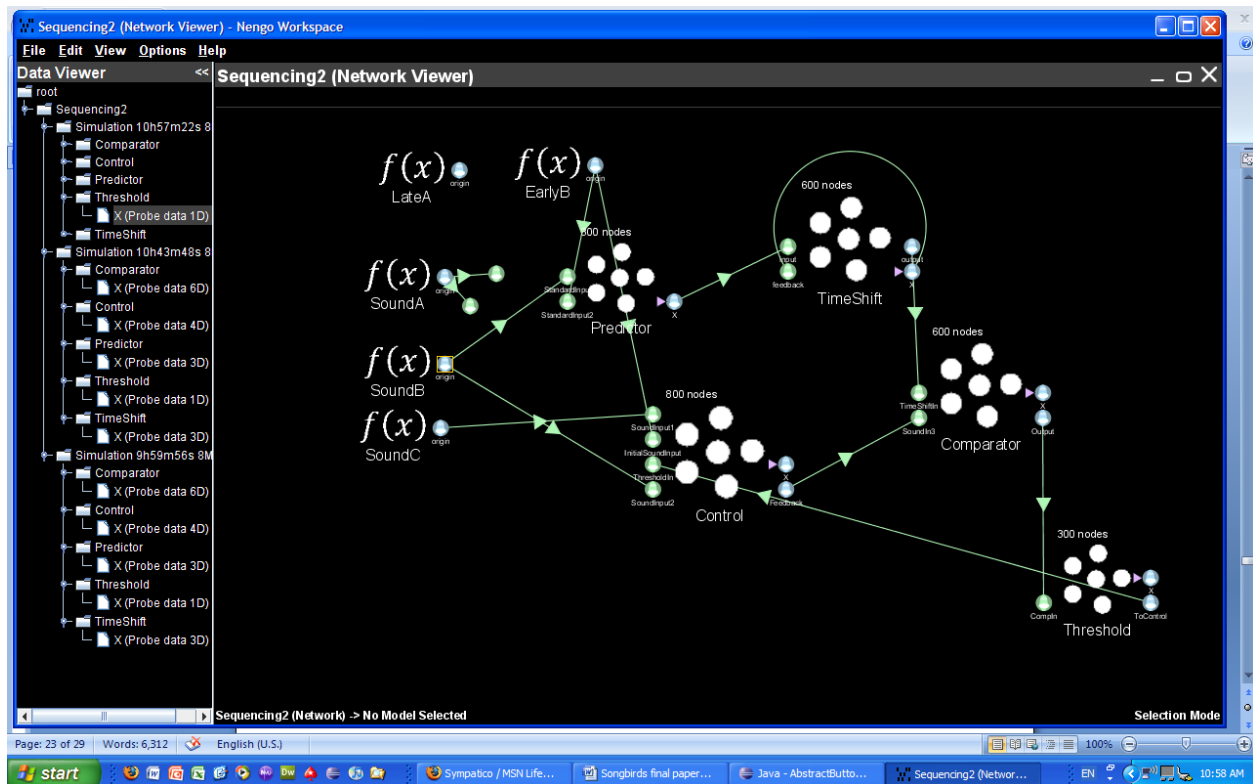




Of course, the final picture of the threshold output is key – there is no activity! This is despite the fact that I allowed EarlyB to provide initial input to the Control element, which is unlikely to occur in HVC. Also notice that while the Predictor and TimeShift neural populations are about as active for BAC as for ABC, the Control network no longer has the yellow (shut down) neurons active and there is significantly reduced activity in the Comparator. Overall, BAC reduces the neural population activity without eliminating it. This is precisely what is seen in the Zebra finch – presenting the reverse BOS reduces overall activity but does not eliminate it.

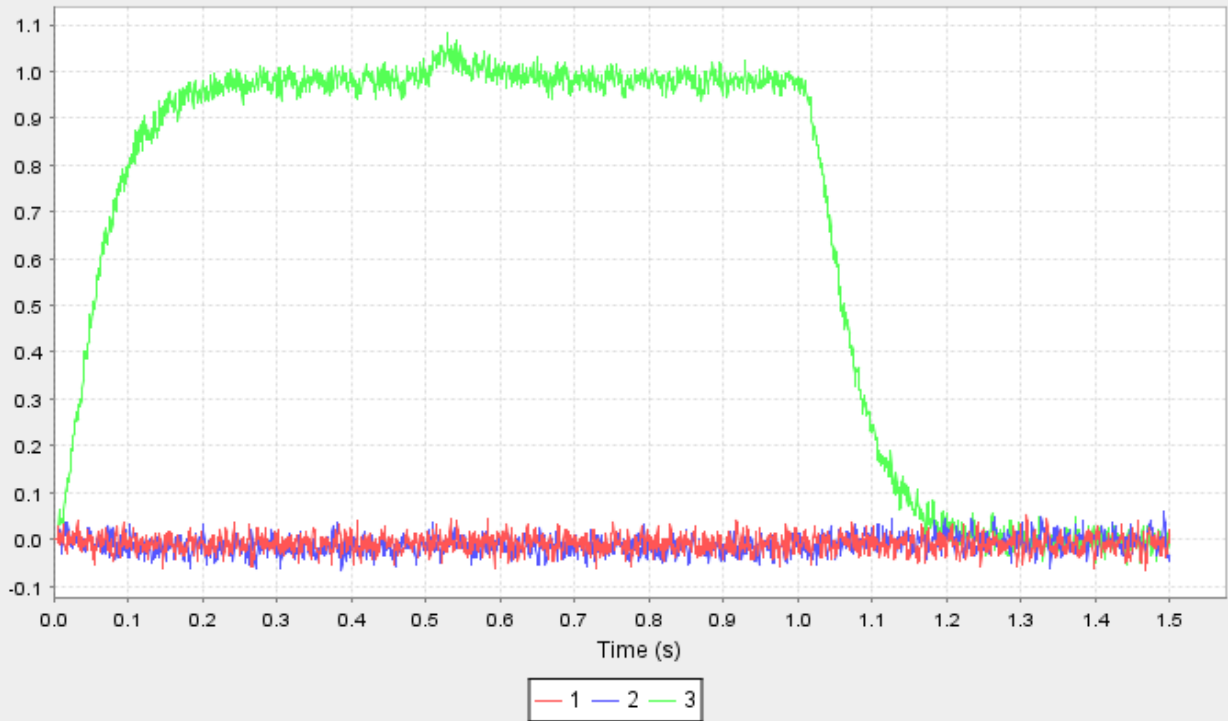


## Running the BBC Sequence

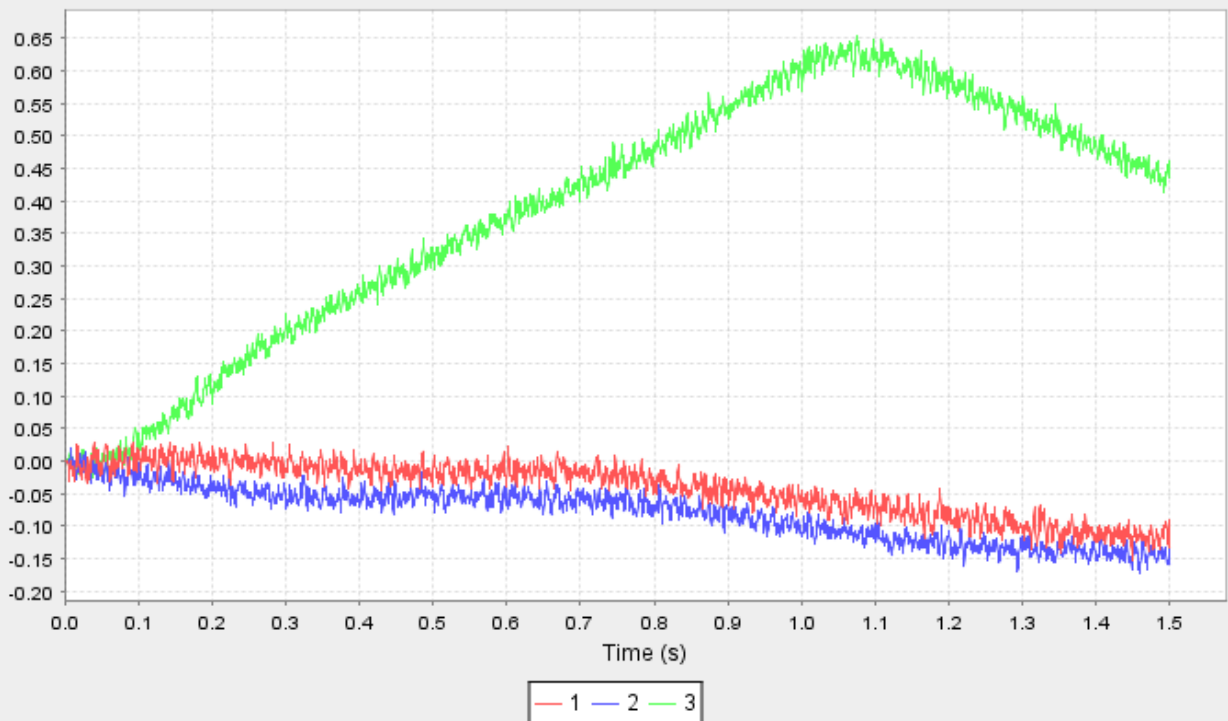


The results are as follows:

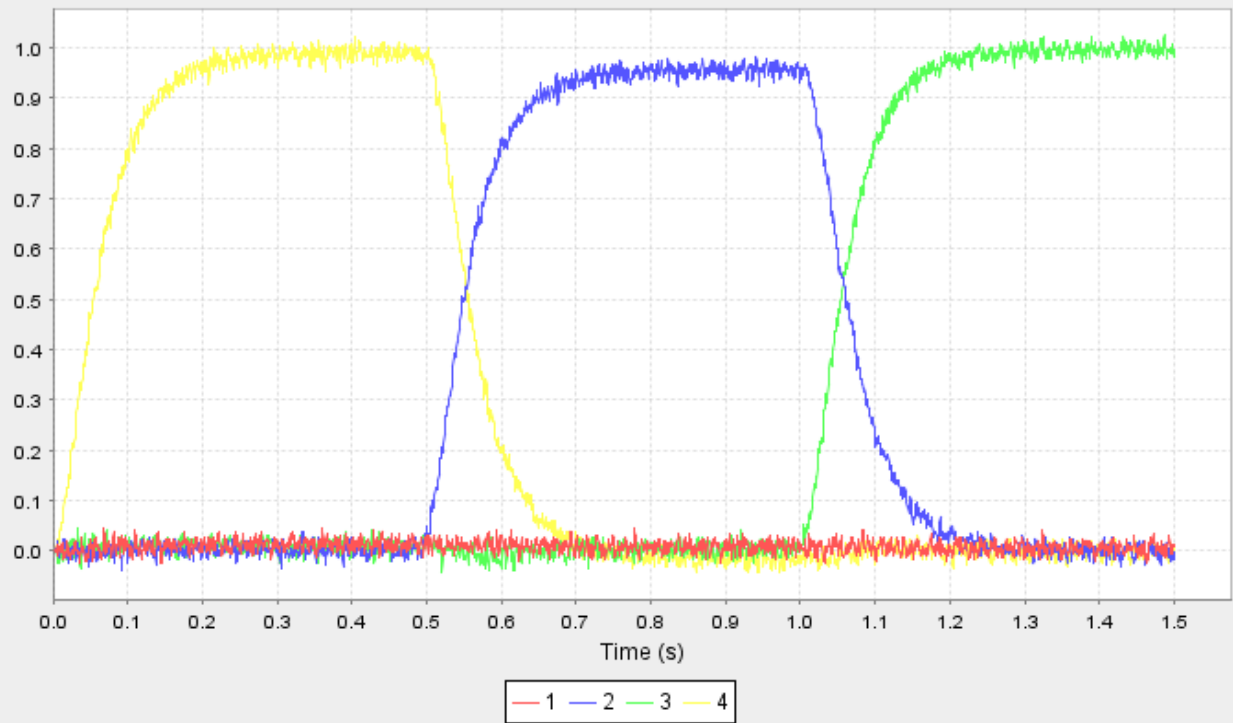
**Predictor Activity BBC**



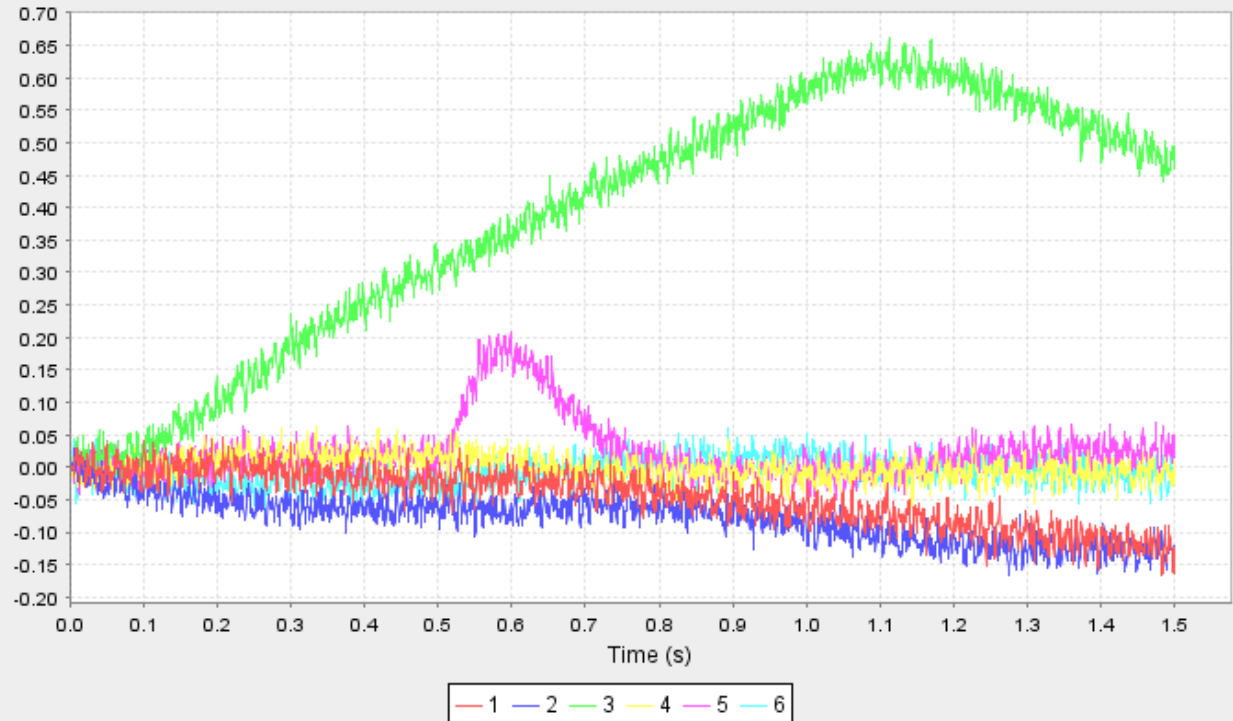
**TimeShift Activity BBC**

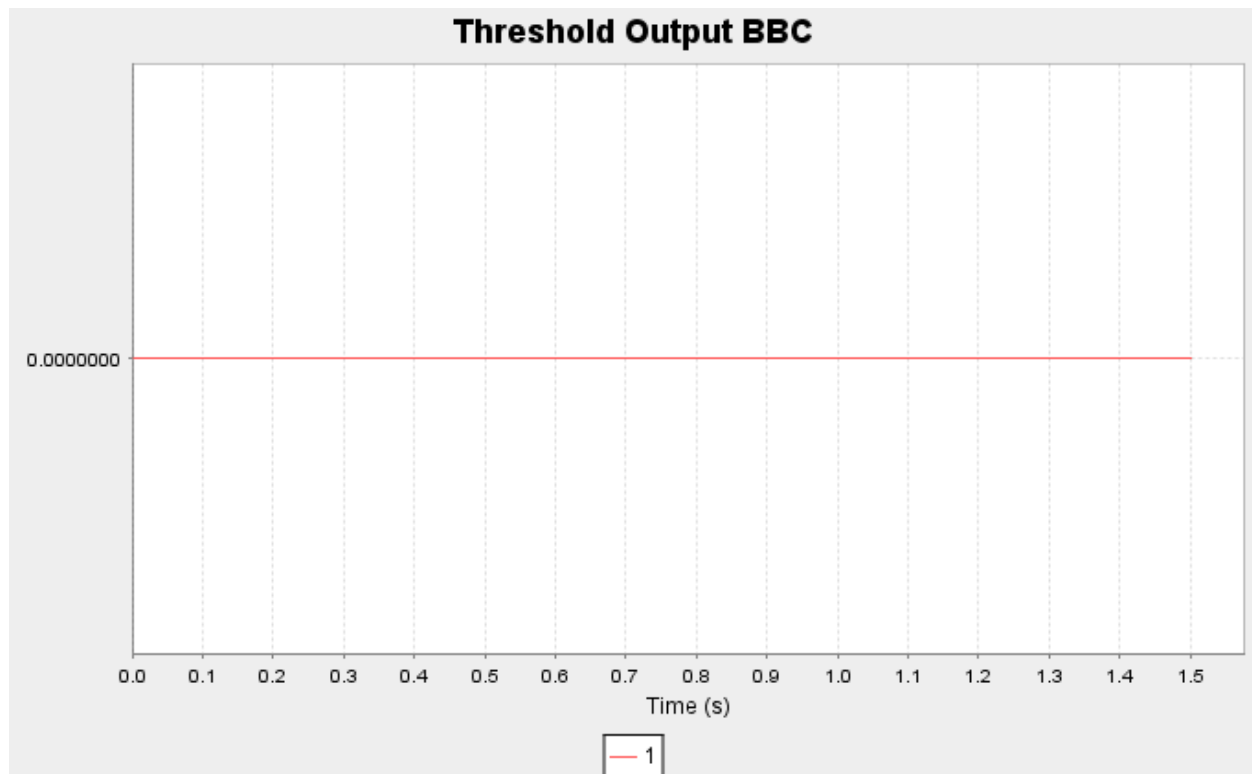


**Control Activity BBC**

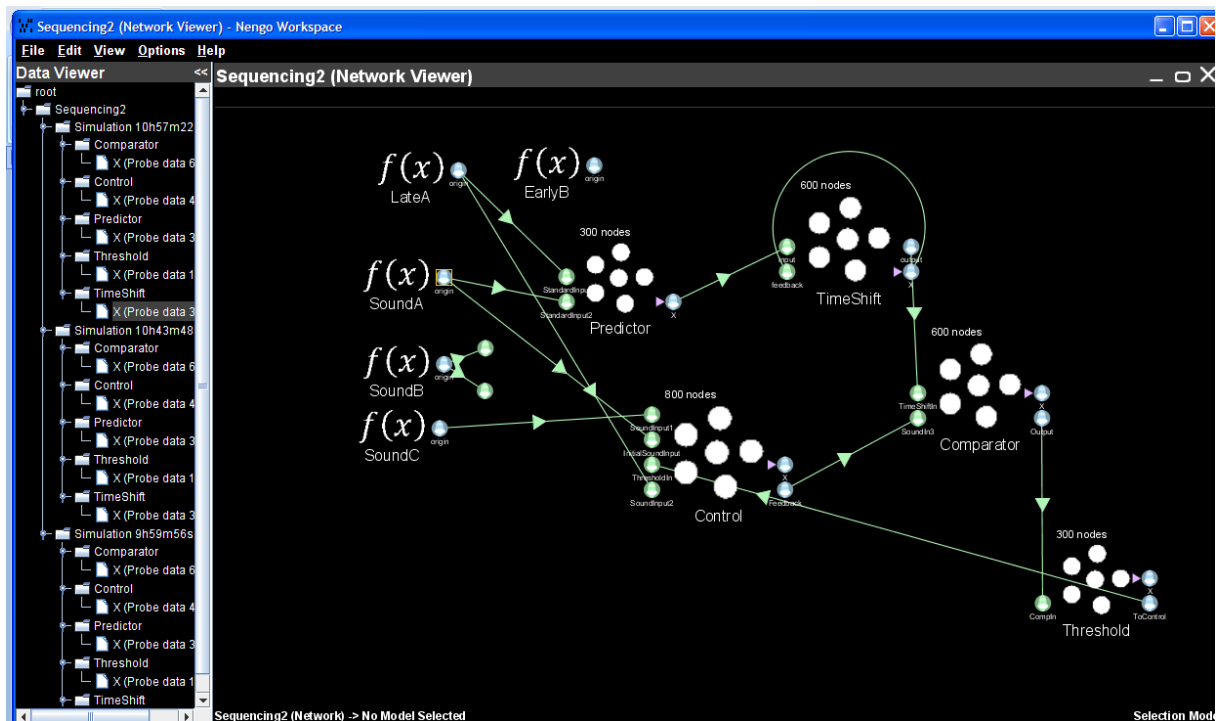


**Comparator Activity BBC**

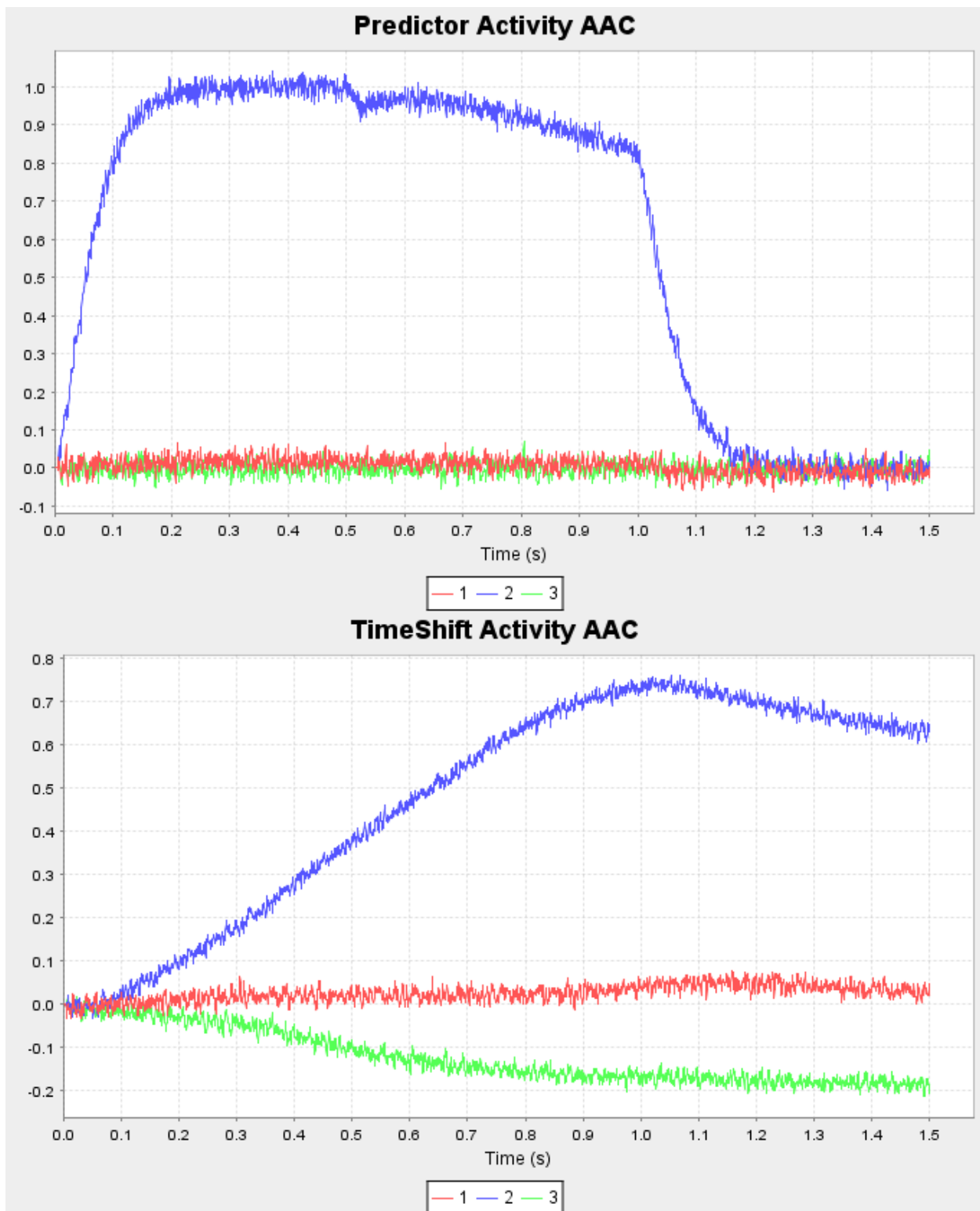




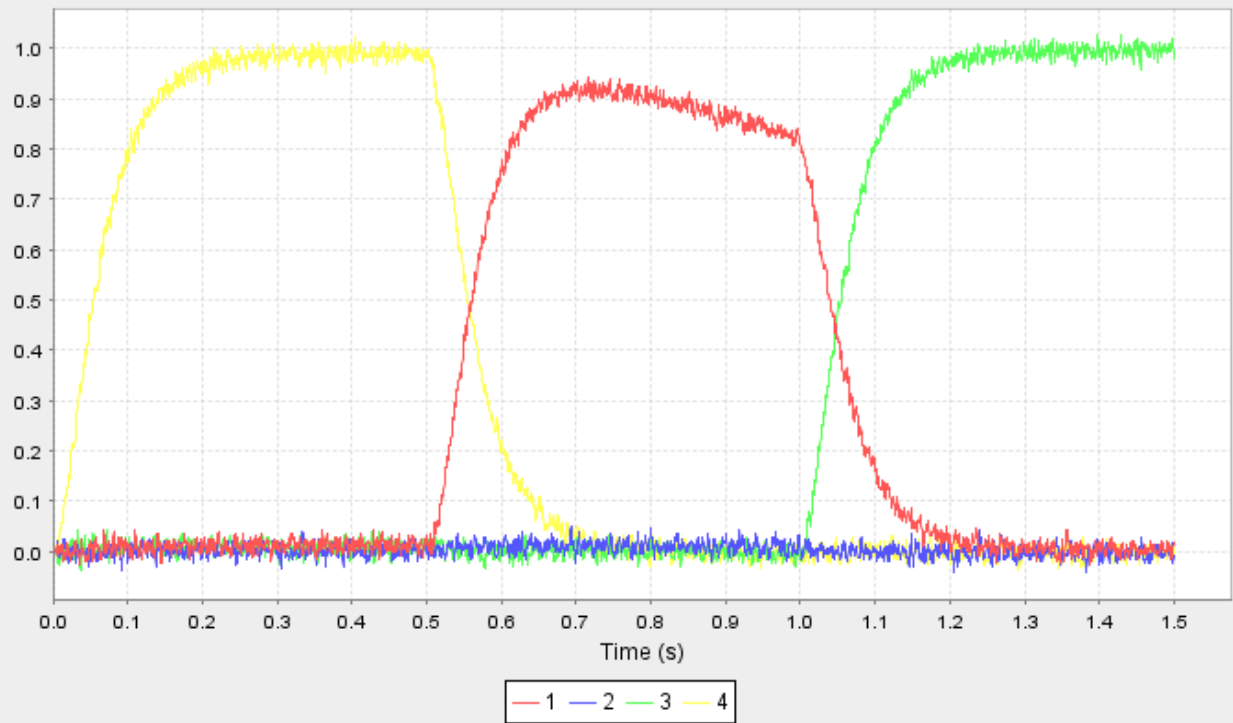
## Running the AAC Sequence



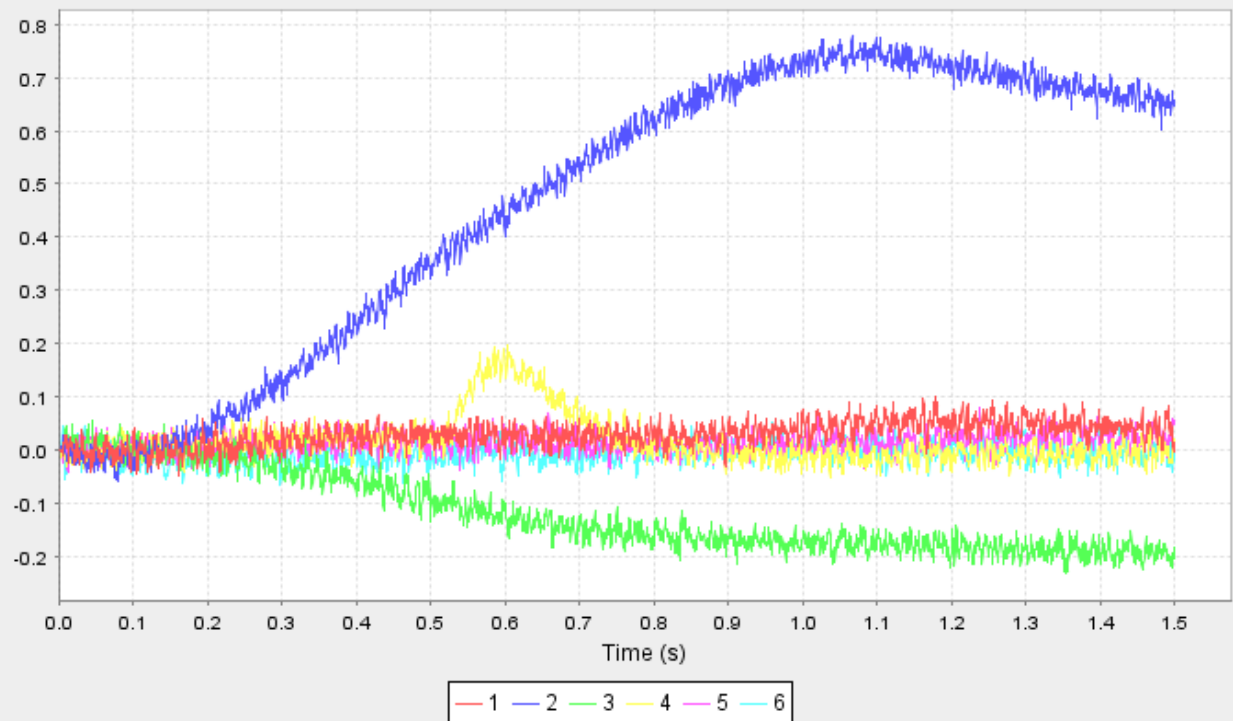
The results are as follows:



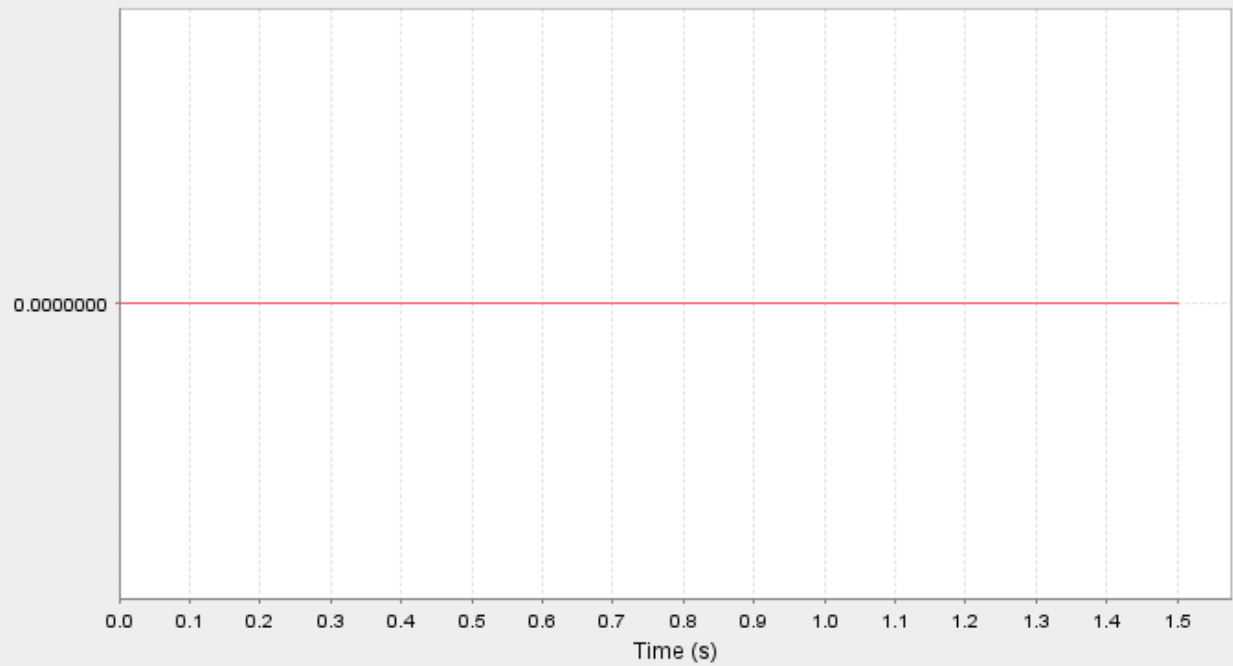
**Control Activity AAC**



**Comparator Activity AAC**



### Threshold Output AAC



— 1

## Summary

We have seen that an NEF can simulate the ability of a neural population to recognize a sequence and respond to any part of it with a precisely timed response (the model only mentioned responding at the end of a sequence, but that was an arbitrary constraint as the SSNs can be adjusted to produce output at any time in response to the Threshold signal). The network created has the neural dynamics experimentally seen in the Zebra finch. It is also able to accurately recognize fairly long sequences without large populations of neurons. Although it was not specifically demonstrated, the sequence is relatively insensitive to time jitter between sequence elements – the feedback constant (0.95 in these simulations) determines this sensitivity.

As discussed in Kazhenikov & Fee (2006) the neurons predict patterns of sound in the song (0.08) by 30-50ms. We see from the dynamics of the TimeShift element that some neurons in that population predict the next syllable (the gradual rise in signal for the next element starts one time step prior to the next syllable being played). Given that a note is approximately 70 ms long, this appears to be further evidence that the model is representing an HVC neural network.

Overall, then, the model appears to be a robust implementation of how HVC could be encoding and interpreting song.

## Bibliography

Beiser, D.G., Hua, S.E., & Houk, J.C. (1997). Network models of the basal ganglia. *Current Opinion in Neurobiology*, 7, 185-190.

Beiser, D.G. & Houk, J.C. (1998). Model of cortical-basal ganglionic processing: encoding the serial order of sensory events. *Journal of Neurophysiology*, 79, 3168-3188.

Blackwell, K.T., Czubayko, U., & Plenz, D. (2003). Quantitative estimate of synaptic inputs to striatal neurons during up and down states *in vitro*. *The Journal of Neuroscience*, 23, 9123-9132.

Dutar, P., Huan, M.V., & Perkel, D.J. (1998). Multiple cell types distinguished by physiological, pharmacological, and anatomic properties in HVC of the adult Zebra finch. *Journal of Neurophysiology*, 80, 1828-1838.

Gabrielli, J.D. (1995). Contribution of the basal ganglia to skill learning and working memory in humans. In J.C. Houk, J.L. Davis, & D.G. Beiser (Eds.), *Models of Information Processing in the Basal Ganglia* (pp. 277-294). Cambridge, MA: MIT Press.



- Glaze, C.M. & Troyer, T.W. (2006). Temporal structure in Zebra finch song: implications for motor coding. *The Journal of Neuroscience*, 26, 991-1005.
- Glaze, C.M. & Troyer, T.W. (2007). Behavioral measurements of a temporally precise motor code for birdsong. *The Journal of Neuroscience*, 27, 7631-7639.
- Gruber, A.J., Solla, S.A., Surmeier, D.J., & Houk, J.C. (2003). Modulation of striatal single units by expected reward: a spiny neuron model displaying dopamine-induced bistability. *Journal of Neurophysiology*, 90, 1095-1114.
- Harrington, D.L. & Haaland, K.Y. (1991). Sequencing in Parkinson's disease. Abnormalities in programming and controlling movement. *Brain*, 114, 99—115.
- Kitano, K., Hideyuki, C., Kaneda, K., Atsushi, N., Masahiko, T., & Fukai, T. (2002). Two-state membrane potential transitions of striatal spiny neurons as evidenced by numerical simulations and electrophysiological recordings in awake monkeys. *The Journal of Neuroscience*, 22, 1-6.
- Kock, G., Costa, A., Bursa, L., Peppe, A., Gatto, I., Torriero, S., Lo Gerfo, E., Salerno, S., Oliveri, M., Carlesimo, G.A., & Caltagirone, C. (2008). Impaired reproduction of second but not millisecond time intervals in Parkinson's disease. *Neuropsychologia*, in press.
- Kozhevnikov, A.A. & Fee, M.S. (2006). Singing-related activity of identified HVC neurons in the Zebra finch. *Journal of Neurophysiology*, 97, 4271-4283.
- Kubota, M. & Taniguchi, I. (1998). Electrophysiological characteristics of classes of neurons in the HVC of the zebra finch. *Journal of Neurophysiology*, 80, 914-923.
- Lewicki, M.S. & Arthur, B.J. (1996). Hierarchical organization of auditory temporal context sensitivity. *The Journal of Neuroscience*, 16, 6978-6998.
- Margoliash, D. & Fortune, E.S. (1992) Temporal and harmonic combination-sensitive neurons in the zebra finch's HVC. *Journal of Neuroscience*, 12, 4309-4326.
- Mooney, R. (2000). Different subthreshold mechanisms underlie song selectivity in identified HVC neurons of the Zebra Finch. *The Journal of Neuroscience*, 20, 5420-5436.
- Mooney, R. & Prather, J.F. (2005). The HVC microcircuit: the synaptic basis for interactions between song motor and vocal plasticity pathways. *The Journal of Neuroscience*, 25, 1952-1964.
- Nenadic, I., Gaser, C., Volz, H., Rammsayer, T., Hager, F., & Sauer, H. (2003). Processing of temporal information and the basal ganglia: new evidence from fMRI. *Experimental Brain Research*, 148, 238-246.
- Nottebohm, F. (2005). The neural basis of birdsong. *PLoS Biol* 3, e164.
- Perbal, S., Deweer, B., Pillon, B., Vidailhet, M., Dubois, B., & Pouthas, V. (2005). Effects of internal clock and memory disorders on duration reproductions and duration productions in patients with Parkinson's disease. *Brain and Cognition*, 58, 35-48.
- Rao, S.M., Mayer, A.R., Harrington, D.L. (2001). The evolution of brain activation during temporal processing. *Nature Neuroscience*, 4, 317-323.
- Smith, J.G., Harper, D.N., Gittings, D., & Abernethy, D. (2007). The effects of Parkinson's disease on time estimation as a function of stimulus duration range and modality. *Brain and Cognition*, 64, 130-143.
- Spiro, J.E., Dalva, M.B., & Mooney, R. (1999). Long-range inhibition within the Zebra Finch song nucleus RA can coordinate the firing of multiple projection neurons. *Journal of Neurophysiology*, 81, 3007-3020.

Stern, E.A., Jaeger, D., & Wilson, C.J. (1998). Membrane potential synchrony of simultaneously recorded striatal spiny neurons *in vivo*. *Nature*, 394, 475-478.

Tripp, B. & Eliasmith, C. (2007). Neural populations can induce reliable postsynaptic currents without observable spike rate changes or precise spike timing. *Cerebral Cortex*, 17, 1830-1840.

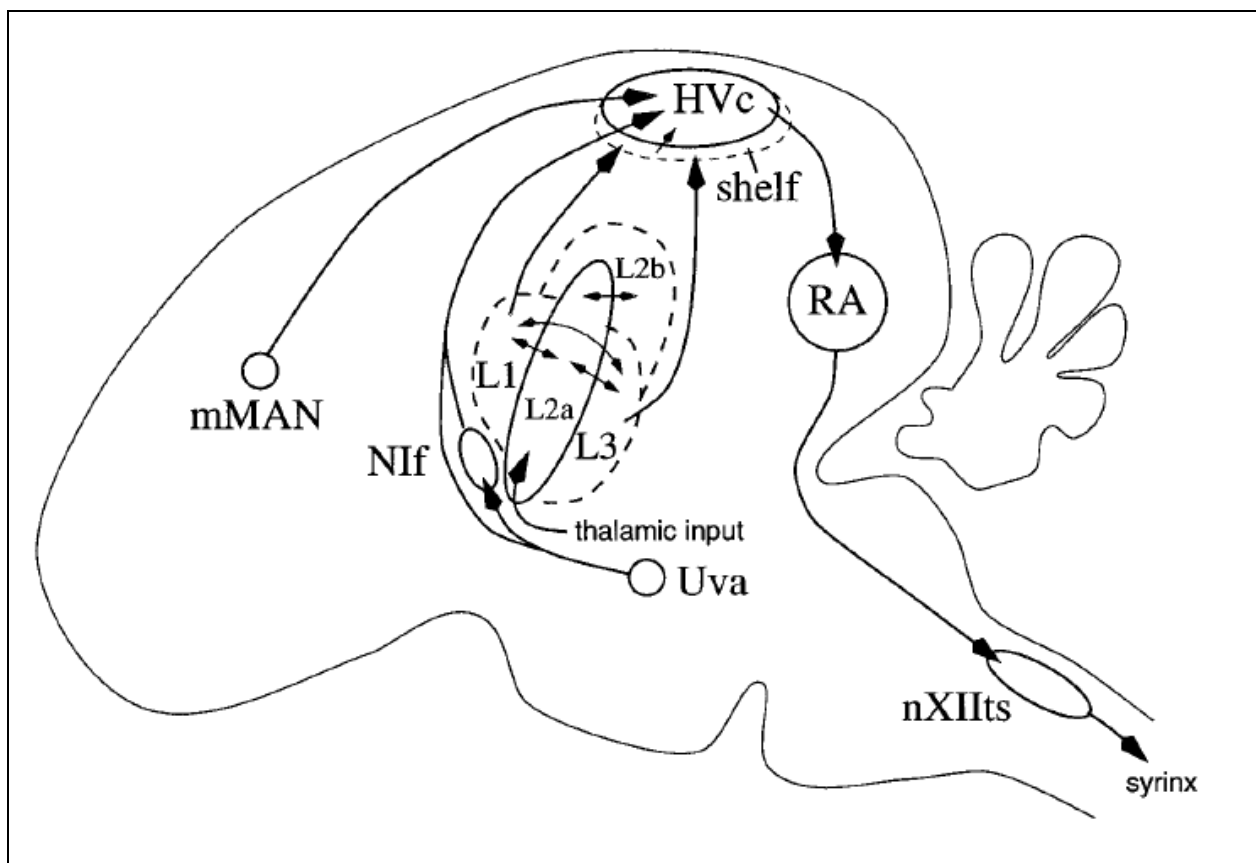
Troyer, T.W. & Bottjer, S.W. (2001). Birdsong: models and mechanisms. *Current Opinion in Neurobiology*, 11, 721-726.

Vates, G.E., Vicario, D.S., & Nottebohm, F. (1997). Reafferent thalamo-“cortical” loops in the song system of oscine songbirds. *The Journal of Comparative Neurology*, 380, 275-290.

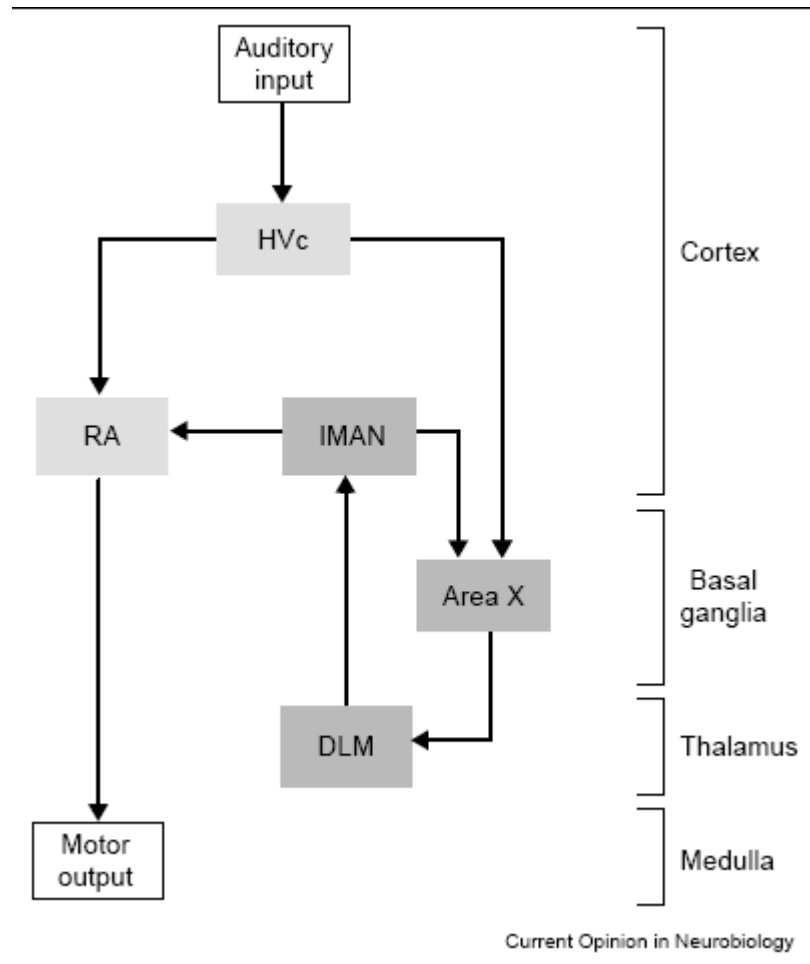
Wilson, C.J. & Kawaguchi, Y. (1996). The origins of two-state spontaneous membrane potential fluctuations of neostriatal spiny neurons. *The Journal of Neuroscience*, 16, 2397-2410.

Yu, A.C. & Margoliash, D. (1996). Temporal hierarchical control of singing in birds. *Science*, 273, 1871-1875.

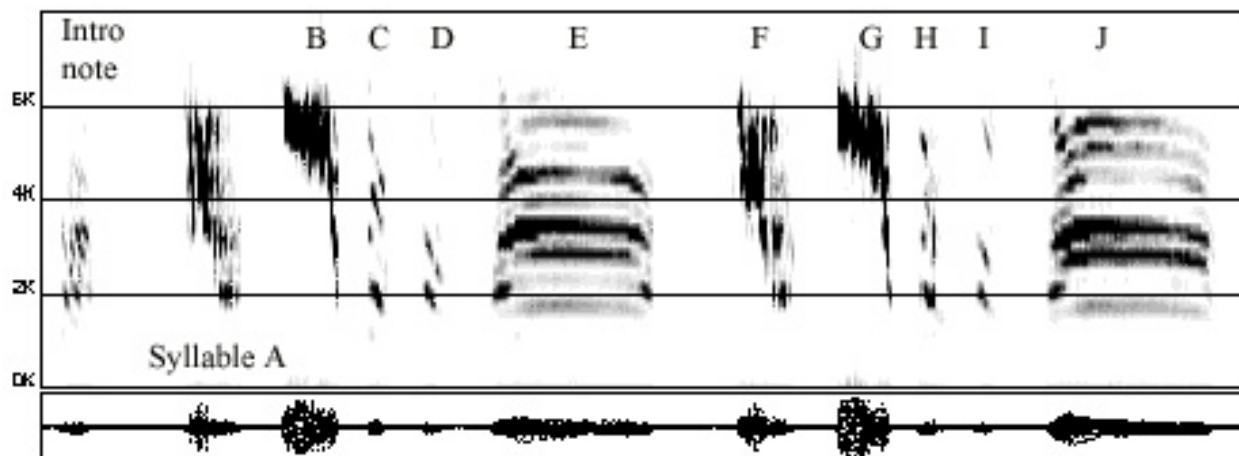
## Figures



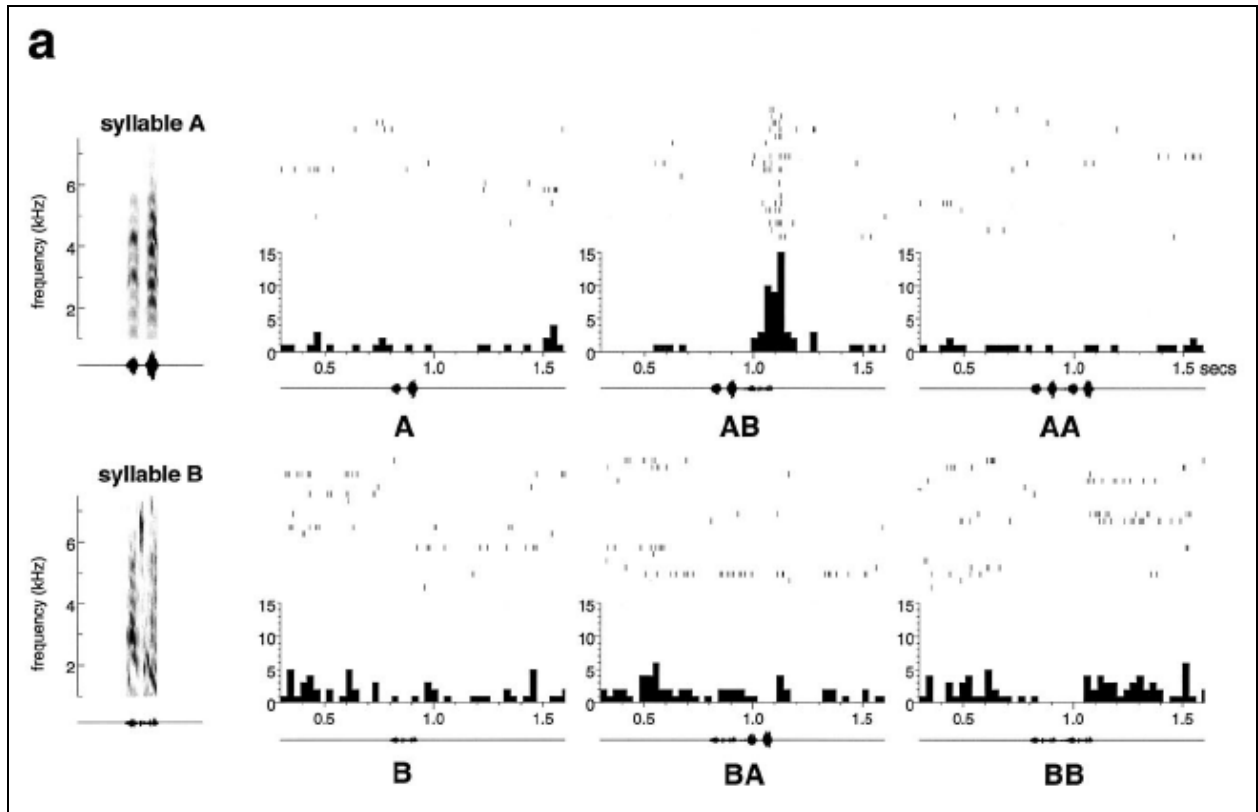
**Figure 1a:** Schematic of songbird brain from Lewicki & Arthur (1996).



**Figure 1b:** Correspondence of songbird brain structure to human structures, from Troyer & Bottjer (2001).



**Figure 2:** Sample of song syllables of Zebra Finch from Lewicki & Arthur (1996).



**Figure 3:** SSNs response to BOS from Lewicki & Arthur (1996).

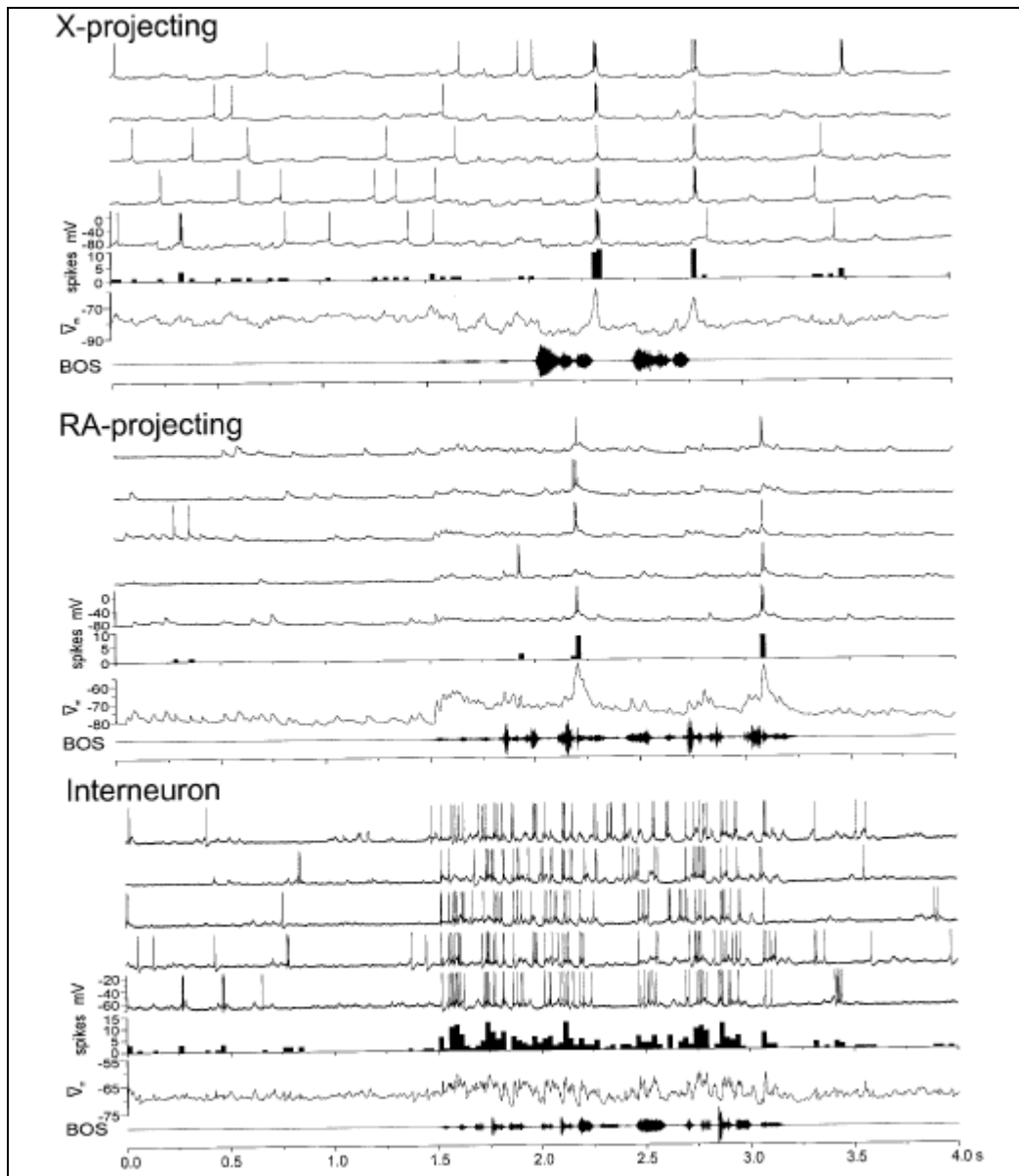


Figure 4: From Mooney, 2000. Three recordings of individual HVC neuronal responses to the BOS (bottom trace). The second last trace in each is the membrane voltage and the third last trace the cumulative number of spikes.