Narendra Joshi
CS591: Neural Networks

Date: December 8, 2012

A Spiking Neural Network for the Classification of Zebra Finch Songs

Introduction:

Classification of bird songs is relevant for several ecological, developmental and behavioral studies of zebra finches. Several studies have examined the changes in the quality of the song of young zebra finches during their growth into adulthood. Most of these studies have only looked at the general features of sound and the broader arrangement of syllables. The most characteristic differences in the songs of young and adult zebra finches have not yet been successfully quantified. A clearer knowledge of this topic would provide more accurate measure of the learning process in zebra finches and also some insights into how the brain processes sounds.

Our study aims to provide some insights into the neural processing of sound by implementing a spiking neural network that can differentiate between the songs of young birds and adult birds. This is done by first extracting selected quantitative features of the songs and then feeding these feature inputs into the network for classification. This study demonstrates that the extraction of good features to represent sound is critical for successful classification and very small changes in neuronal and synaptic properties can drastically change the behavior of a neural network and its ability to process auditory stimulus.

Zebra Finch Auditory System:

When sound signal is received at the ear, hair cells in the cochlea convert the mechanical movement of cochlear fluid generated by the sound to neuronal action potentials. These action potentials are transmitted by auditory neurons from hair cells in cochlea, through the brain stem and basal ganglia, into the cerebral cortex. The higher level processing of information in the sound signal happens in the cortex and several regions of the cortex (called ‘nuclei’) are involved in this process.

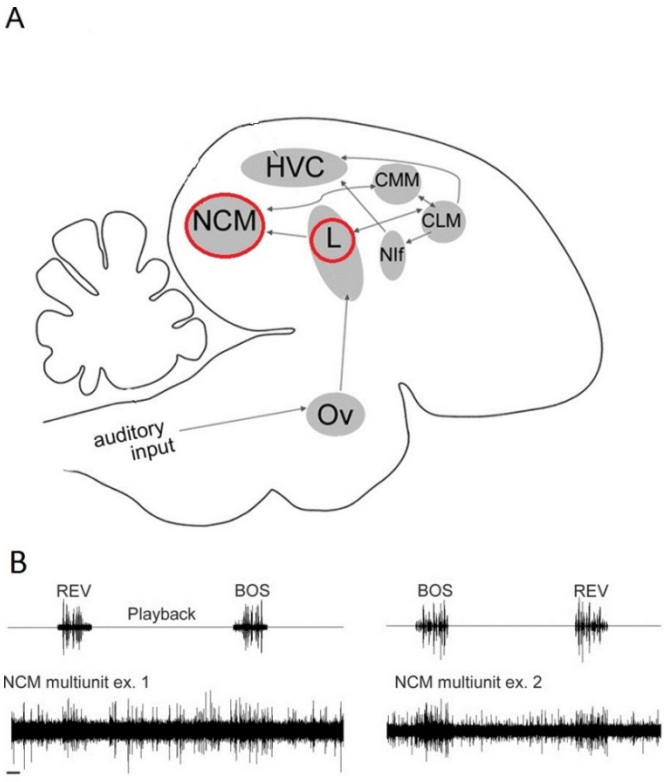


Figure 1: 1A shows a basic outline of the zebra finch auditory system. Notice Field L and NCM, marked with red circles. 1B shows neuronal firing pattern in NCM to BOS (bird’s own song) and REV (reversed BOS). The response to BOS is stronger, with larger number of firings. (Adapted from Healey and Joshi, 2012)

In this experiment we modeled the behavior of two brain nuclei involved in auditory information processing: Field L and NCM. (See Figure 1A) Field L is the main cerebral entry point for all auditory action potentials arising in the ear. It is a 'primary auditory nucleus', which means that not a lot of information processing happens at this point. This nucleus is known to respond robustly to all incoming auditory signals. The other region of interest is called NCM (Caudo-Medial Nidopallium) and it is a 'secondary auditory nucleus', which means that more specific processing of the auditory signal happens in this region.

NCM receives a major fraction of the auditory signals coming out of Field L. Electrophysiological recordings show that NCM in zebra finches responds strongly, with a large number of action potentials, to songs produced by adult males of the same species. (Stripling et al., 1997) However, the response is generally weaker for songs of other species, for random noises and even for non-song sounds of the same species. (See Figure 1B) In this context, the immature song of a juvenile zebra finch is expected to generate fewer action potentials in NCM, since such a song is not as behaviorally relevant as the song of an adult. We used this difference in number of action potentials as the basis for differentiating between the songs of adults and juveniles.

Spiking Neural Network

We created a spiking neural network with one input layer of 30 neurons, one hidden layer of 30 neurons and an output layer with one neuron. Our network was based on the Izhikevich model but a lot of the features of this model were removed to create a simplified network. (Izhikevich , 2006) There are no connections between neurons within a layer. (See Figure 2) The first layer models the behavior of Field L and the second layer acts as the NCM. The first layer receives incoming auditory input in the form of 30 feature components of an audio recording. The inputs cause firing of neurons in this layer and all of these firings are conveyed to the second layer. All neurons in second layer converge on the last neuron which acts as the output neuron.

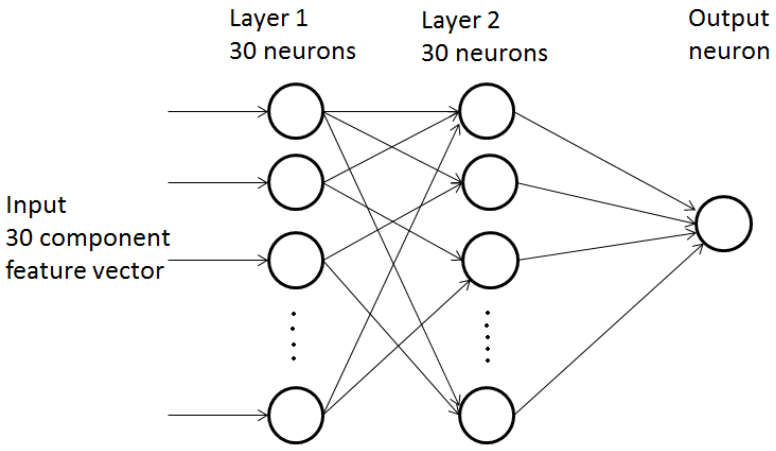


Figure 2: Basic outline of the Spiking Neural Network used in this study

Every neuron in layer1 had 20 synapses with 20 randomly chosen neurons in layer 2. This number was found to be the most optimal for generating good spiking behavior. Using fewer than 20 synapses led to very sparse firing and since in this network the number of spikes resulting from a stimulus input is used as the criterion for classification, a robust firing pattern is essential for any differences between two stimuli to be well represented. When neurons with 10 synapses were used, performance was very poor for both training and testing sets. On the other hand, when the number of synapses per neuron was set to 30 (the maximum possible), the

network quickly reached a local minimum for the training set but performed poorly on the test set.

We used Spike-timing Dependent Plasticity (STDP) as the learning model to train our spiking neural network. Only the synapses between neurons in layer 1 and layer 2 were updated. The initial synaptic weights for these neurons were set to 6 mV. The firing of a neuron in layer 2 immediately after the firing of its pre-synaptic neuron in layer 1 increases the weight of the synapse between the two neurons by 0.2 mV. On the other hand, if a neuron in layer 1 fires after its target neuron in layer 2 has fired, the synaptic weight is decreased by 0.1 mV. These parameters for synaptic weight update were observed to be very sensitive to changes. For example, when the update step was changed from 0.2 mV to 0.3 mV, all of the synapses quickly attained the maximum possible value of synaptic weights. This homogenization of synaptic connections led to a drastic deterioration of the network's ability to differentiate between two distinct input stimuli.

Audio features:

Extraction of audio features is one of the most important steps/decisions for audio classification tasks using neural networks. The size of audio files can easily go into the megabytes, which makes raw audio files very time consuming as well as memory intensive as inputs for artificial neural networks. Therefore, representation of the important features of the audio recording in more simplified forms makes dealing with audio files much more computationally feasible. All of the audio recordings were 1.00 second long clips of either adult song or a juvenile song. We used freely available Matlab codes for extracting quantitative features from audio recordings. (Ellis, 2006; Moddemeijer, 2001; Slaney, 1998)

MFCC (Mel-frequency cepstral coefficients) are the most widely used audio features for speech recognition tasks**.** In this experiment, the sound recording was segmented into short frames of 256 sample points each and then each frame was processed in several stages to obtain 13 numbers representing various features of the sound frame. The numbers for all frames in a recording were averaged to obtain one set of 13 features for that particular recording. Similarly, RASTA (Relative Spectral Transform) features were derived using MFCC features and basically serve to remove the effect of static background noise and other unchanging components of the signal. (Ellis, 2006) Entropy feature has been used in previous studies of zebra finches to monitor developmental changes in quality of a young bird's song. The entropy of a young bird's song decreases over time as it matures into adulthood. Similarly, energy feature gives a representation of the distribution of amplitudes over time in the audio recording.

There are several other features that could additionally have been used. However, it was visually observed that these features provided a decent representation of differences in the songs of adults and juvenile birds. Therefore, in this experiment, we limited ourselves to these four broad categories of features. Other studies have used more features. The decision of which features to include seems to depend highly on the goals of a particular classification task or any other audio processing task.

The entropy values were observed to be generally higher in juveniles as compared to adults. On the other hand, adult songs had higher average energy content. The differences in MFCC features were more difficult to discern. Of the 13 features, the first one tended to have a larger magnitude in adults than in juveniles. Similarly, there were some other general differences in MFCC features and in the RASTA features. The numerical values of the features were multiplied by constant factors to generate numbers in the range of 10 mV. This had to be done to ensure that presentation of these features to the neural network leads to a substantial change in the membrane potential of target neurons.

Properties of the neural network:

We tested the firing properties of the spiking neural network using random voltage inputs. As expected, neurons with large, positive incoming signal tended to fire with a high frequency. Similarly, neurons with strong positive synaptic connections also tended to fire more frequently. On the other hand, neurons with negative input voltage and negative incoming synaptic weights were always silent. Resting potential of all neurons was set to -65 mV. When a neuron fires a spike, its membrane potential gets reset to -65 immediately. The number of neurons in each layer didn't have any major effect on firing patterns. However, as noted earlier, changing the number of synaptic connections had very noticeable changes in firing behavior. Networks with sparse connections tended to have low frequency of firing. The most active synapses fired an action potential almost every millisecond but a big proportion of neurons fired only after a delay of a few milliseconds. "Bursting" like pattern was observed occasionally whereby a lot of neurons fired at the same time followed by widespread suppression of activity for a few milliseconds.

A pattern of synaptic connections was generated at the beginning of the experiment and this connection was maintained throughout the training and testing phases. Only the synaptic weights were readjusted during training phase. During testing, both the synaptic connections and synaptic weights remained unchanged and simply the firing patterns were observed.

The neural network was trained on training datasets using STDP for several epochs. After each epoch, a file containing the updated synaptic weights was saved and this file was reloaded each time the network was active. During training, the network was tested every 5 epochs using the testing datasets to check the effect of training on the performance of the network. We used 'error rate' as a measure of the performance of the neural network. Error rate is basically the proportion of inputs that was classified incorrectly and it has values ranging from 1 to 0.

As noted earlier, the number of spikes generated in layer 2 of the neural network by a particular song input was used to determine whether the song is of an adult or of a juvenile. It was observed that most adult songs resulted in more than 20 spikes during training when the synaptic weights were being updated. Similarly juvenile neurons tended to cause slightly fewer spikes. Once this observation had been made, the STDP learning process was adjusted to magnify the difference in firing rates for the two kinds of stimuli. The value of 20 spikes was thus used as a "cutoff" to separate adult songs from juvenile songs.

Experiments and results:

As expected, synaptic update using STDP was observed to have a substantial effect on the rate of firing of neurons. We observed a difference of about 50% (~20 vs. ~30 spikes) in the number of spikes per 'activity session' for the 30 neurons in second layer depending on whether or not synaptic weights were being incremented using STDP, i.e. if STDP was introduced in a network, it tended to increase the number of spikes by 50% within a few epochs.

For each test, we trained the neural network for 150 epochs unless we were specifically looking for the effect of the length of training. We did not observe any significant improvement in training error rate when more than 150 epochs, up to 5000, were used. (See Appendix Figure 1) Over the course of training, the error rate had a fairly consistent pattern of change, with an initial rapid drop between epoch 1 and 10 followed by a similarly rapid increase in error around epoch number 15. This elevated error rate slowly decreased over the next 100 or so epochs to stabilize around the value of 0.3. (See figure 3) This value of 0.3 for error rate was observed repeatedly over several trials and seems to be a lower limit to the error rate for this network.

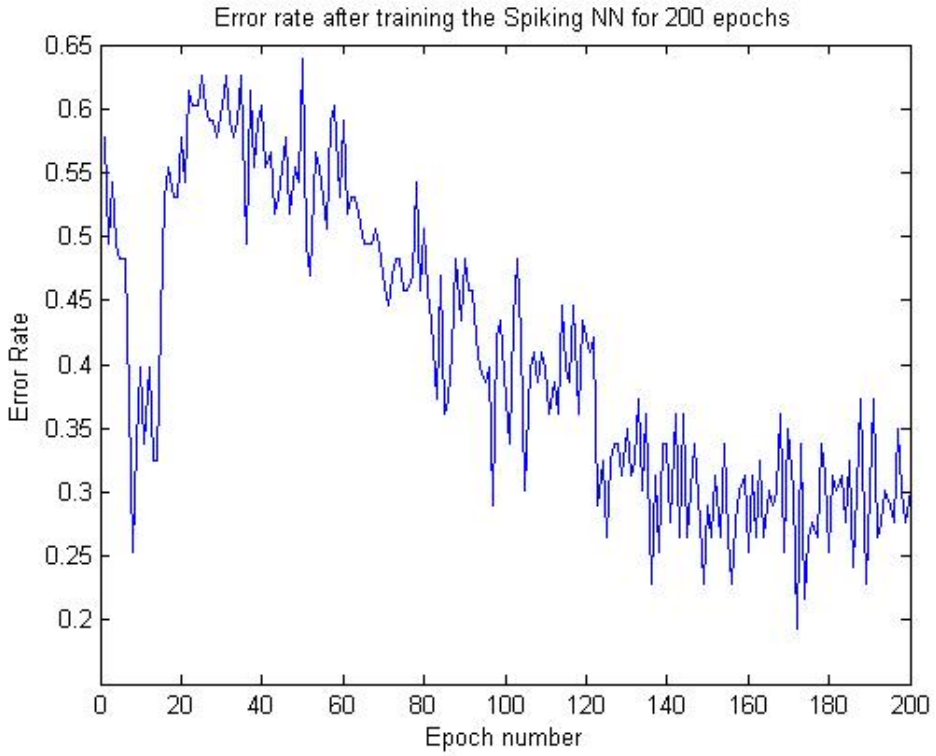


Figure 3: Error rates at different stages of training a spiking neural network for 200 epochs

The results in Figure 3 were observed with synaptic weight limits of +20 mV and -50 mV, which means that the modification of the initial synaptic weight of 6 mV using STDP could only lead to a new values within the range of +20 mV and -50 mV. When the test set was run through this trained network, the error rate was usually much higher, around 0.5. On some instances, the test error rate was as low as 0.28, however, such instances were quite rare.

During the course of several trials, it was observed that the upper and lower limits of synaptic weights had a big influence on how the error rate evolved for both, the training set as well as the test set.

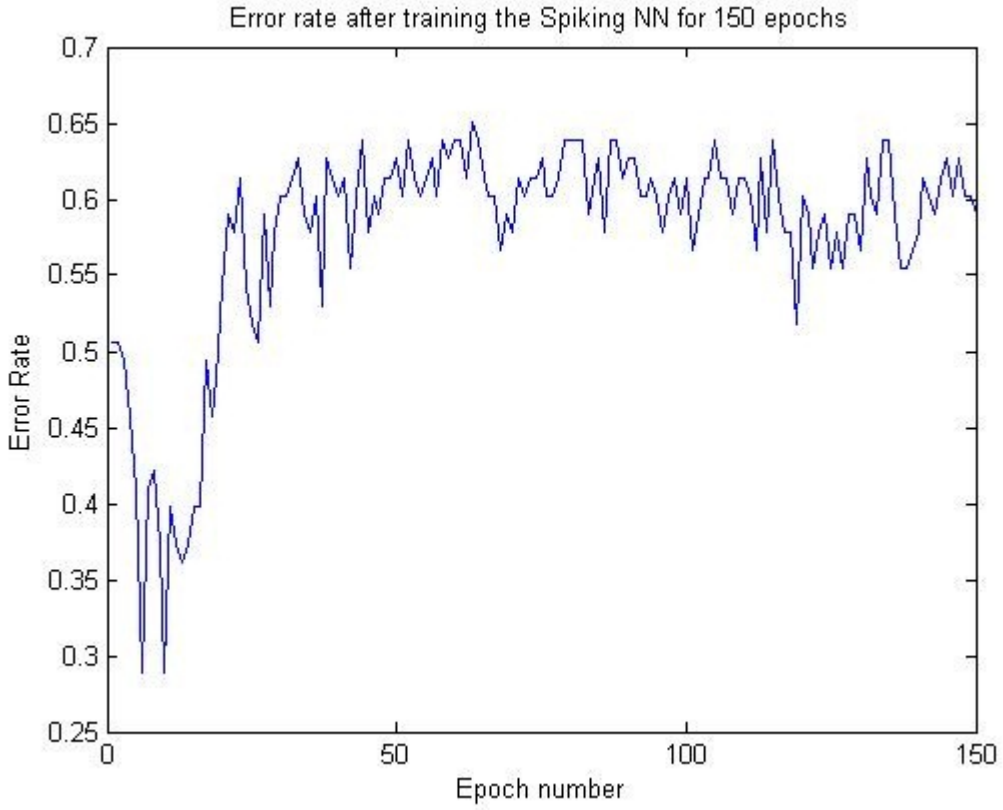


Figure 4: When the possible range of synaptic weights was set to +50 to -50, the error rate for training set showed a very different behavior as compared to figure 3. The error rate was initially around 0.5, decreased momentarily to about 0.3 and rose back to about 0.6. However, unlike in the cases where synaptic weight was confined to a range of +20 to -50, the error rate did not decrease in subsequent epochs. Surprisingly, this led to an improved performance in test set, with the fully trained network having an error of 0.25 in this particular instance.

As shown in Figure 4, when we changed the limits of synaptic weights to +50 mV to -50 mV, the error rate for training set had a completely different behavior. The error rates remained consistently high after around epoch 20, unlike in Figure 3. When maximum possible synaptic weight was limited to 20 mV (as in Figure 3), the best performance of trained networks consistently hovered around 0.5. However, increasing the upper limit of possible synaptic weights to 50 mV (as in Figure 4) substantially increased the performance of trained neural network on the test set to an error rate of as low as 0.275. Surprisingly, however, this manipulation of synaptic weight limits led to almost doubling of the error rate on the training set from an average of around 0.3 to around 0.6.

A comparison of the results in figures 3 and 4 indicates that avoiding local minima is very important for optimal performance of this spiking neural network. To check if the

surprisingly high error rate for the training sets could be lowered, we readjusted the rules for how the network classifies the input stimuli.

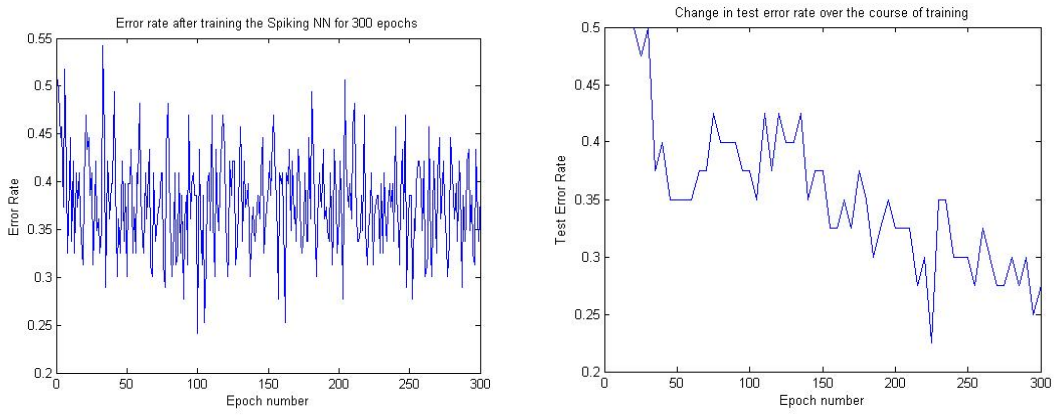


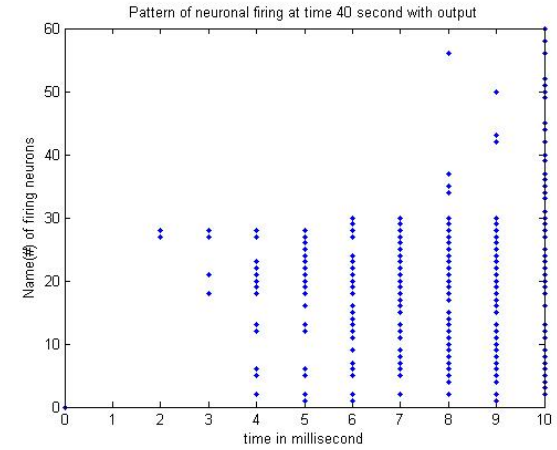
Figure 5: Error rates for training set decrease substantially when the cutoff is changed from a fixed value of 20 spikes to a re-adjustable parameter. Figure on the right shows that the error rates for test set keeps decreasing until 300 epochs. However, longer training sessions showed that the error rate mostly stays around the value of 0.25.

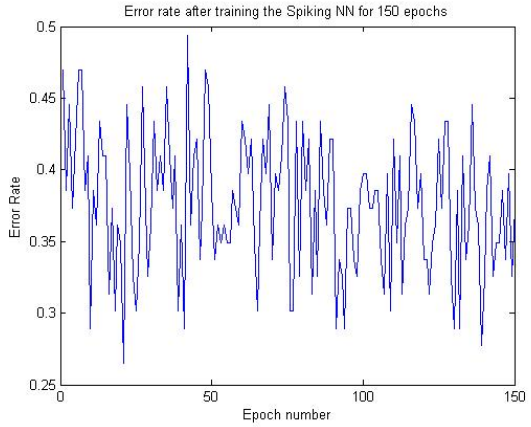
Most synaptic weights get updated to reach a value close to the upper limit (+50 mV for most of the experiments). At some synapses, the weight gets reduced to the lower limit (-50mV for most of the experiments). Overall, there is a variation of synaptic weights spread between the two limits, with about 70% of the synapses having a positive weight.

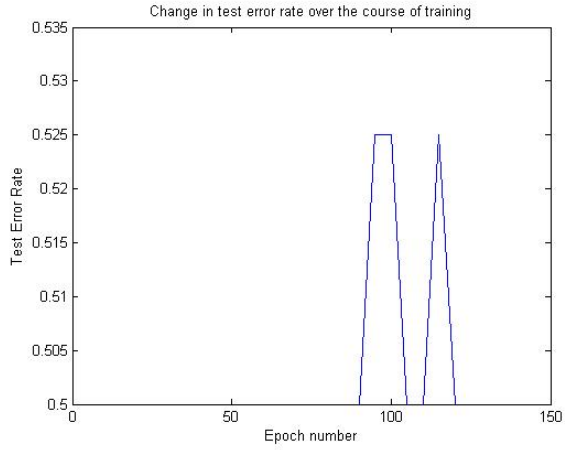
Using limits other than +50 mV to -50 mV changed the behavior of the spiking neural network drastically. When the limits were set at +10 mV to -10 mV, there was a higher final error rate of about 0.4 for training set. The test set showed even poorer performance, with an error rate of 0.5 that changed little throughout the course of training. The values of synaptic weights are still fairly well spread within the two limits. The smaller synaptic weights mean that the neurons now require more incoming signals to reach spiking threshold. This lengthens the average ‘waiting time’ for the neurons, which in turn means that the overall firing rate of the neural network is lower. This lowered firing rate was clearly evident in ‘firing plot’ showing overall number of firings for a particular dataset. (See figure 6)

Using limits larger than +50 mV to -50 mV did not improve the performance of the neural network but it didn’t diminish the performance either. (see Apendix, Figure 2)

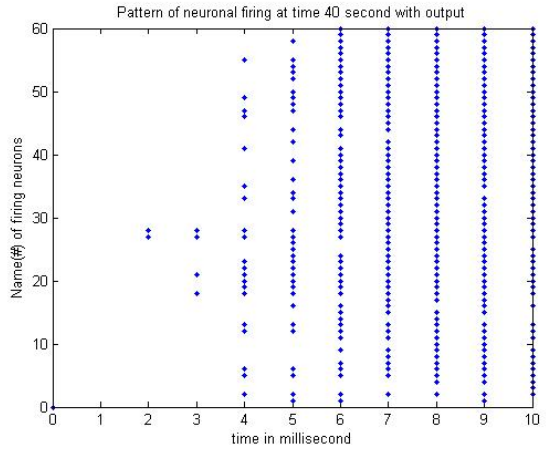
Synaptic weight limits: +10 mV to -10 mV

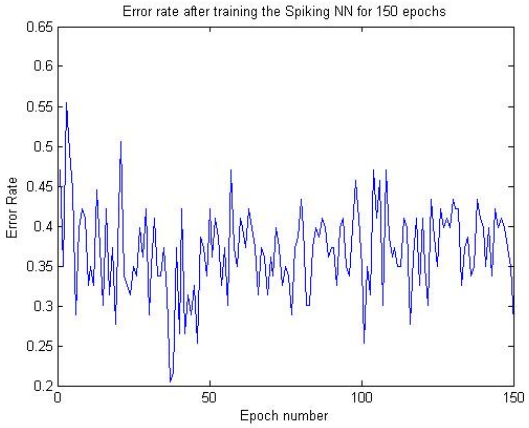






Synaptic weight limits: +50 mV to -50 mV





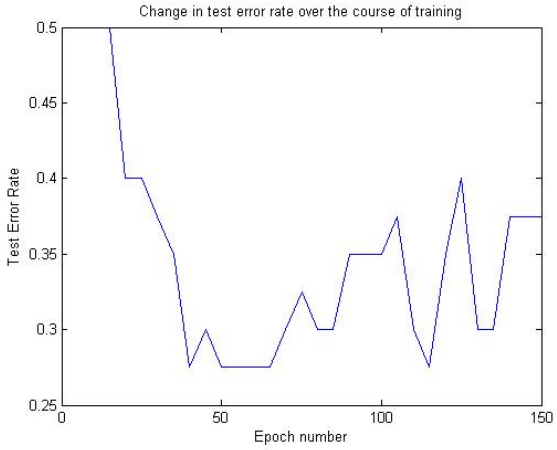


Figure 6: The narrower range of synaptic weights (+10 mV to -10 mV) results in a sparser spiking pattern as compared to the wider range (+50 mV to -50 mV), see top two graphs. The error rates during training are comparable for both ranges (see middle two graphs). However, there is a substantial difference in error rates for test sets (bottom two graphs), with the narrower range performing very poorly but the wider range still gives error rates as low as 0.3.

Limitations and future work:

As indicated by our experiments, even relatively small changes in some of the variables in our spiking neural network had drastic effect on the behavior of the network. However, we could also test the effect of other variables that were not manipulated in this study.

First of all, more hidden layers could be used to check whether the number of layers has any significant consequences for firing patterns in the neural network. Also, more layers would allow us to model the other auditory nuclei in the brain of zebra finches, thus making our model more biologically relevant.

A firing threshold was not specifically used in the current network. The membrane potential had to reach 30 mV through accumulation of incoming voltages in order for an action potential to fire. We could incorporate a firing threshold to see its possible implications for network behavior. A firing threshold lower than 30 mV would be expected to result in higher spiking rates.

Finally, we could also try using different resting potentials other than -65 mV.

Conclusions

The best performance of our spiking neural network was an error rate of roughly 0.25, which means that about 75% of the input stimuli were classified correctly. Several features of the network were readjusted but none led to better than 75% correct classification. However, we observed that the network responded very strongly even to small change in its properties, which gives us confidence in its ability to differentiate between different kinds of signals. The quality of the numerical features extracted from audio clips to generate inputs for the neural network might be acting as the limiting factor in the network's performance. A better method of feature extraction will hopefully lead to a more efficient classification of adult and juvenile zebra finch songs.

References:

Ellis, Dan. PLP and RASTA (and MFCC, and inversion) in matlab using melfcc.m and invmelfcc.m. [Online]. Available: http://labrosa.ee.columbia.edu/ matlab/rastamat/ . (2006)

Izhikevich, Eugene M. "Polychronization: Computation with spikes." Neural computation 18.2 (2006): 245-282.

Moddemeijer, Rudy. Matlab library of Rudy Moddemeijer. http://www.cs.rug.nl/~rudy/matlab/ . (2001)

Remage-Healey, Luke, and Narendra R. Joshi. "Changing Neuroestrogens Within the Auditory Forebrain Rapidly Transform Stimulus Selectivity in a Downstream Sensorimotor Nucleus." *The Journal of Neuroscience* 32.24 (2012): 8231-8241.

Slaney, Malcolm. "Auditory toolbox." *Interval Research Corporation, Tech. Rep* 10 (1998): 1998.

Stripling, Roy, Susan F. Volman, and David F. Clayton. "Response modulation in the zebra finch neostriatum: relationship to nuclear gene regulation." *The Journal of neuroscience* 17.10 (1997): 3883-3893.

Appendix:

Energy feature was calculated using a simple formula in Matlab:

Energy = (1/(length(wavFile))) * sum(wavFile.^2),

where wavFile is the one-dimensional representation of amplitude samplings over the course of the recording.

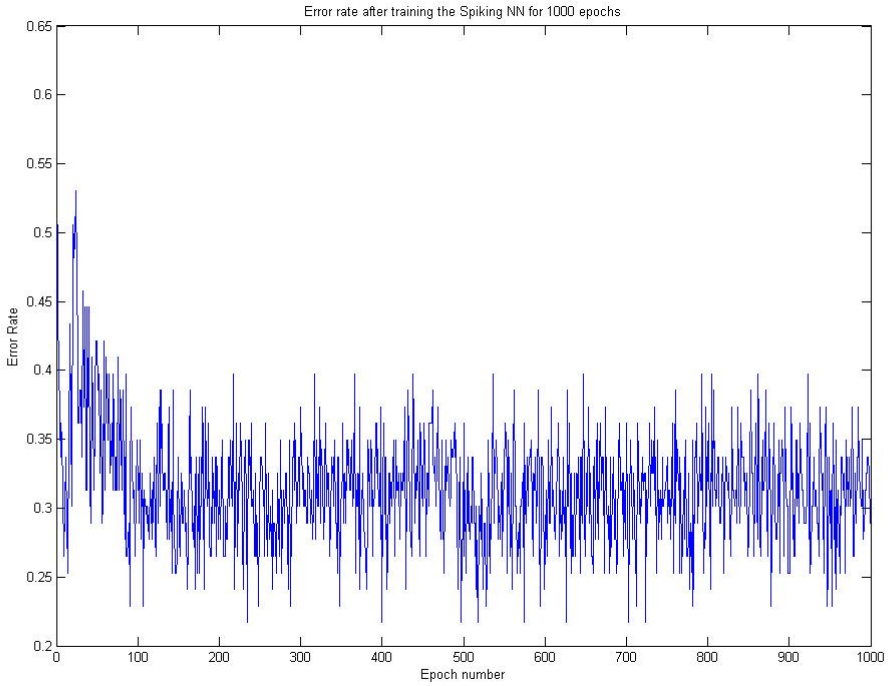


Figure 1: Result of training a network for 1000 epochs. The usual behavior of an initial dip followed by an abrupt rise in error rate and then a gradual decrease is observed. There is no noticeable improvement in error rate after about 150 epochs.

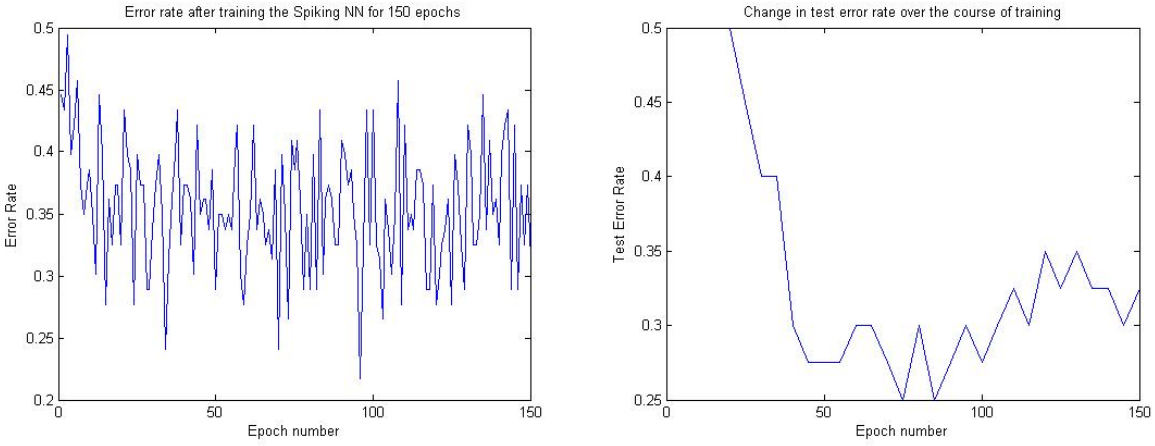


Figure 2: Using a very wide range of synaptic weights (+80 mV to -80 mV) doesn't result in any noticeable improvement in performance for either training set or test set as compared to the range (+50 mV to -50 mV). However, it is noteworthy that the performance doesn't get worse either.