Neural networks in the extraction of mouse ultrasonic vocalizations

Drew Kristensen, Dr. Adam A. Smith (Advisor)
Department of Mathematics and Computer Science, University of Puget Sound

Introduction

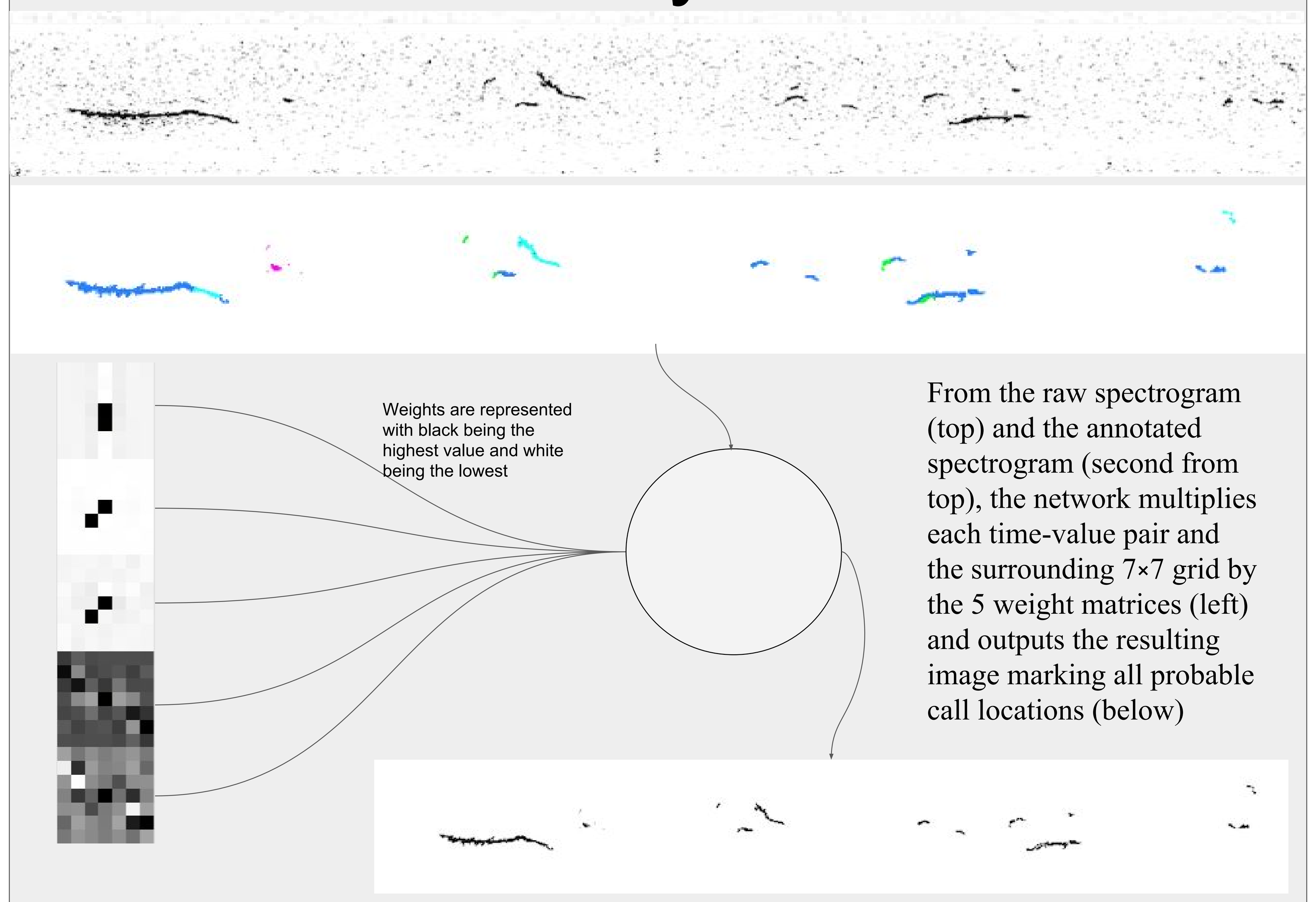
The ultimate goal of this study is to gain a better understanding of the affective state of mice, by analyzing mouse vocalizations. To achieve this goal, this study aimed to distinguish vocalizations from recordings using an artificial neural network. In order to analyze these calls, they must first be identified and characterized. While human technicians are very accurate at these classifications, manually annotating the ultrasonic vocalizations (USVs) is a time-intensive process and might be better if it were automated.

This project relies on neural networks, a machine learning technique that employs weighted inputs with the ability to learn as a method of discerning calls from noise.

Background

A neural network consists of a number of artificial neurons which take weighted inputs in order to calculate a value. Here, our weighted inputs are the weights stored within our neurons, and the time-frequency values found in the spectrogram. In order to do this, we classify every time-frequency value as one of 6 categories; rising, flat, and falling for calls, streak and spike noise, or neither noise nor a vocalization. While a single neuron is capable of only the most simple distinctions, with more neurons, the more complicated the “learning” function can be. By sending the input data to a layer of these neurons, we create a neural network. The output of these neurons is taken as the input of a final neuron, so the weights of the various neurons can be changed to better fit the data. This results in the final classification of the input as “call” or “no-call”. Because the neurons take in many weighted values and calculate a sum on which to base the output on, the various weights for each input value can be changed with each iteration. This re-evaluation serves the purpose of ensuring that the network can accurately converge upon the best values when working through the trial set.

Preliminary Results



From the raw spectrogram (top) and the annotated spectrogram (second from top), the network multiplies each time-value pair and the surrounding 7×7 grid by the 5 weight matrices (left) and outputs the resulting image marking all probable call locations (below)

The artificial neural network (ANN) works by taking a spectrogram generated from a preprocessed audio file. From this spectrogram, a user must first mark the image in locations where either calls or noise are present. The ANN then goes through the spectrogram, time-frequency pair by time-frequency pair, multiplying each value and the surrounding 7×7 grid of these values by a series of weights stored in various matrices within a layer of neurons. If the result of the operation is above a certain threshold, the ANN outputs a positive value for a call. These neurons which hold these weighted matrices can be trained if given an annotated version of a spectrogram in order to better represent the data. If the value output was too small and therefore below the threshold, the difference between the target value and estimated value is passed through the input matrices, allowing each value within them to be altered by an amount relative to their overall weight to the final output. This process also works for values which were higher than the actual values, and thus need training down to lower values. Training allows the ANN to key in on specific aspects of calls that may or may not be apparent to human observers, yet work with high precision.

Motivation

Mice are a fundamental part of research in many scientific fields because we are able to tailor them to fit our needs. These animals provide a great testing medium for a wide range of treatments which need experimentation or discovery. This is possible since mice share many important anatomical and physiological features with humans. Furthermore, we can alter mice to fit our specific needs through careful use of breeding, surgical changes, or alterations to their genetic makeup. These modifications allow us to model specific diseases such as cancer or mental disorders such as autism and schizophrenia. This study’s aim is to create an additional tool to aid the current analysis of a mouse’s affective state done through traditional methods, in order to gain an insight into mice that model these mental disorders.

References

G. P. Lahvis, E. Alleva, and M. L. Scattoni, “Translating Mouse Vocalizations: Prosody and Frequency Modulation,” Genes Brain Behav, vol. 10, no. 1, pp. 4–16, Feb. 2011.

M. Mitchell, An introduction to genetic algorithms. Cambridge, Mass.: MIT Press, 1996.

Mitchell, Tom M. "Artificial Neural Networks." *Machine Learning*. New York: McGraw-Hill, 1997. 81-127. Print.

Smith, Adam A. "Hidden Markov Models and Mouse Ultrasonic Vocalizations." *XRDS: Crossroads, The ACM Magazine for Students XRDS* 21.4 (2015): 48-53. Web.

I would like to extend my sincerest gratitude to the Sherman Fairchild Research Grant for making this research project possible, and to Dr. Adam A. Smith, whose patience and support was much appreciated.