Voice Recognition System

Self Organizing Maps

Matthew Paiz

December 8, 2010

1 Introduction

The purpose of this project was to implement a Kohonen Self Organizing Map (SOM) using an FPGA. This system was designed for use in realtime voice recognition applications. Upon initialization, the system will prompt the user to record an event of 'interest'. It will use this reference value to detect the occurrence of other similar events. Throughout its operation, the device will continually make adjustments to its internal weight matrix to increase the accuracy of its predictions.

2 Background

2.1 Artificial Neural Networks

Artificial Neural Networks (ANN) were developed to model the behavior of biological neurons. The model primarily focuses on the signals sent between neurons with dendrites as inputs, and axons as outputs. (Figure 1). Each connection is referred to as a synapse, which allows the propagation of electrical and chemical signals. The mathematical model of a neuron describes the output y, of a neuron when given a set of inputs x, connected using a set of synapses w.

$$y = \lambda\left(\sum_{i}^{l} x_i w_i\right) \tag{1}$$

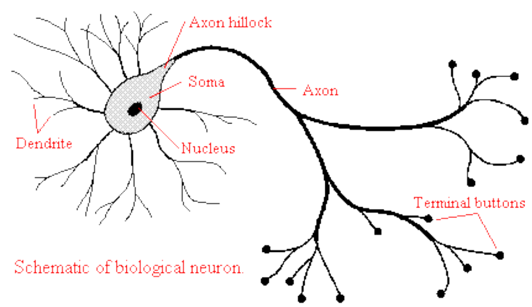

Figure 1: Anatomy of a Neuron

The λ in the equation refers to an activation function (usually a sigmoid). This function models the "firing" and "non-firing" states of a neuron but also serves the purpose of limiting the output of a neuron.

2.2 Kohonen SOM

A Kohonen SOM refers to a type of ANN that uses backpropagation to train a set of weights. This type of network is used to classify an input x into a set number of categories. The value of x is compared to corresponding columns in a weight matrix using the Euclidean Distance Function. The network outputs the corresponding index of the class that most closely resembles the input. Using the vector representation of this class, a learning algorithm is applied to decrease error.

$$w_{ij}(new) = w_{ij}(old) + \alpha\left[x_i - w_{ij}(old)\right]$$

The α value in this equation refers to a learning constant that decreases over time. This gives the algorithm a wide search space initially, and a chance to converge as time passes.

Usually the weight matrix in a SOM is initialized using random values. For this project, we will be initializing the weight matrix with our reference vectors. This will force the network to converge with the appropriate states in the appropriate positions in the weight matrix.

3 Design Overview

The system was implemented using a basic closed-loop control system. The acoustic input to the system is discretized using an Analog to Digital Converter (ADC) before being processed using the SOM. The SOM requires at least two reference inputs be provided to define the two possible classification states. An overview of the design can be seen in Figure 2. The output of this

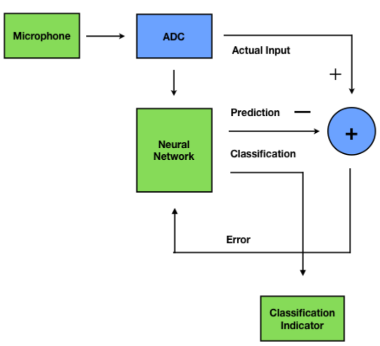

Figure 2: System Design

system is the classification index for the current acoustic input sample. The index is used to designate the class that the input signal belongs to (event of interest or not). Using the corresponding vector representation of the detected class, an error value is calculated and propagated back into the system.

3.1 Hardware Design

The acoustic input is acquired using an analog transducer microphone. The output of this microphone is very small (hundreds of μVs) and must first pass through an amplifier before being sent to the ADC. The amplifier was constructed using a cascaded op-amp design yielding a gain of approximately 10000. Using a summing amplifier as the last stage of the cascaded amplifier, a DC offset of 2.5V was introduced to preserve the negative portion of the signal.

The SOM was implemented using the DE0 FPGA. An FPGA was selected due to its parallel processing capabilities. This alllowed data to be classified while at the same time propogating the error correction values. The design was implemented using the Verilog programming language. It was broken up into several controllers designed to provide essential resources.

- **Main Controller**
  Responsible for initializing and coordinating the behavior of all other controllers. This is the first Verilog module to become active in this design.

- **HCS12 Controller**
  Responsible for providing access to an ADC input, as well as a way to output the weight matrix for debugging.

- **SDRAM Controller** Responsible for providing access to the DE0's 8MB SDRAM chip. This controller is responsible for initializing and encoding read/write commands from other controllers. In addition, it is also responsible for periodically refreshing the SDRAM's memory.

Using the debug information collected from the HCS12, it was possible to perform a number of characterizations on the system.

4 Characterization

This section provides information about the characterization process of this project. This process involved selecting the appropriate parameters for the different project submodules. The parameters are discussed individually.

4.1 ADC Characterization

It was important to characterize the appropriate sampling frequency and resolution for the ADC chip. The sampling frequency was selected at 10 KHz to focus primarily on the frequencies of human voice. In addition, an 8-bit resolution was selected to decrease the size of the search domain. This decrease in size would have the benefit of decreasing the SOM convergence time, but it would also decrease the classification capacity of the system. This was determined to not be a problem, due to the fact that only two states were being detected.

4.2 SOM Characterization

Some important considerations when designing a SOM are the selection of data encoding and distance functions. In the standard Kohonen model, the distance between an input vector $\mathbf{x}$, is determined using the Euclidean distance function.

$$D(\mathbf{x}, \mathbf{w_j}) = \sqrt{\sum_{1}^{n}(x_i - w_j)^2} \qquad (2)$$

This function is directly responsible for determining the appropriate category for all input data. To determine if it would be appropriate for this system, MATLAB simulations were conducted. Shown below is a list of samples acquired using the microphone and an 8-bit ADC. The samples were obtained using a magnitude trigger of 2.94V. Using the above values as the

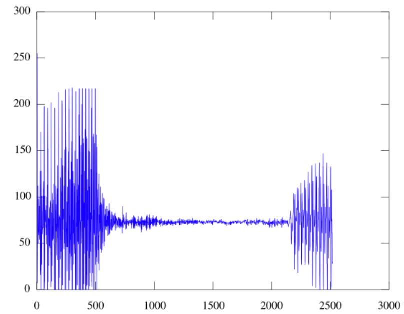

Figure 3: "Testing" Sample

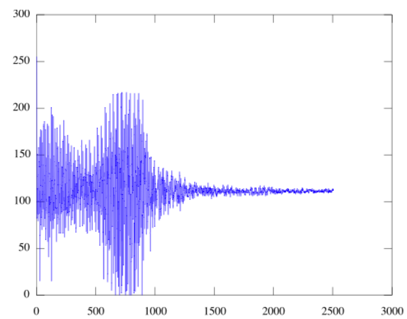

Figure 4: "Hello" Sample

reference inputs for the SOM. It was possible to test the classification of different input samples. This was done using the Euclidean Distance function in MATLAB. Using this function, it was found that the Euclidean Distance could adequately classify most input signals. It is projected that the performance would also increase as the backpropagation was implemented.

5 Results and Conclusions

Due to time constraints, a full working SOM was never delivered. The final system included the ability to sample acoustic data in realtime. The final system did include classification abilities, but a number of errors still existed in the Euclidean Distance implementation. Overall the project could be considered a success, due to the fact that an expandable FPGA framework was developed.