HABITUATION BASED NEURAL CLASSIFIERS FOR SPATIO-TEMPORAL SIGNALS

Bryan W. Stiles and Joydeep Ghosh

Department of Electrical and Computer Engineering,
University of Texas at Austin,
Austin, Texas 78712

ABSTRACT

Based on the habituation mechanism found in biological neural systems, novel dynamic neural networks are proposed for recognizing temporal patterns. The specific task considered in this paper is the classification of whale songs from passive SONAR data, but the networks are also readily applicable to other temporal pattern recognition problems. The fact that the networks designed operate dynamically is important, because it makes the goal of real time data analysis possible.

1. INTRODUCTION

For temporal data, static neural networks, such as the MLP, are untenable because they can produce only a fixed mapping between their current inputs and outputs regardless of the surrounding temporal context. The most common way to recognize a temporal pattern with such a network is to expand the input vector to contain a temporal window which is large enough to include most of the contextual clues needed to recognize each pattern. This type of network is referred to as a time delay neural network (TDNN)[1]. Unfortunately, TDNNs are often unsuitable when a long temporal window is required or the window is of variable length.

While studying biological phenomena that can encode temporal information, we encountered a particularly simple and well understood phenomenon known as habituation. Primarily, habituation is a means by which biological neural systems ignore repetitive, irrelevant stimuli; but aside from this primary function, it has also been suggested to be a means of encoding temporal information [2]. Researchers in neurophysiology have developed mathematical models of habituation [3] [4] [5]. From the model cited in [3], we have designed preprocessing units for use in a spatio-temporal classification network. The model equation is shown as follows:

$$W_i(t+1) = W_i(t) + \tau_i(\alpha_i(1 - W_i(t)) - W_i(t)I_i(t)) \quad (1)$$

Here, $W_i(t)$ is the preprocessed input at time t, and $I_i(t)$, the unpreprocessed input. In [3] $W_i(t)$ represents a synaptic strength, and $I_i(t)$ the activity of the presynaptic neuron, but because our designs use habituation as a preprocessing step, the variables are redefined accordingly. The parameters, τ_i and α_i affect the rate at which habituation occurs, thereby determining the temporal resolution and range of the information obtained.

This work was supported in part by an NSF grant ECS 9307632 and ONR contract N00014-92C-0232. Bryan Stiles was supported by an MCD fellowship and the Du Pont Graduate Fellowship in Electrical Engineering.

2. MATHEMATICAL ANALYSIS

It is the purpose of this section to develop insight into what type of information is encoded by habituation. Due to the fact that habituation is nonlinear, it may be very difficult to analyze for a wide range of inputs. For this reason we will concentrate on the ability of habituation to encode information about a simple pulse function input. We will later show that even with this simplistic assumption, useful properties of habituation can be derived. The input function to be used is defined by the following equation.

$$I(t) = \begin{cases} 1 & \text{if } t_0 \le t < t_0 + \Delta t \\ K & \text{otherwise} \end{cases} \quad (2)$$

The habituation value, W(t), is assumed to have reached an equilibrium value prior to the start of the pulse at $t = t_0$. The consequences of this assumption will be discussed in detail later.

From Equation 1 it is simple to determine the equilibrium value of $W(t)$ when $I(t) = K \; \forall t$.

$$W_{eq}^K = \frac{\alpha}{\alpha + K} \quad (3)$$

If t_0 is assumed to be large, then $W(t_0) \simeq W_{eq}^K$. In the following equation, Equation 1 is expanded to get $W(t_0 + \Delta t)$ in terms of $W(t_0)$.

$$\begin{aligned} W(t_0 + \Delta t) &= I_{hab}^{\Delta t}(t_0 + \Delta t - 1)W(t_0) + \\ &\quad \alpha\tau \sum_{j=0}^{\Delta t - 1} I_{hab}^{j}(t_0 + \Delta t - 1) \end{aligned} \quad (4)$$

Here the term $I_{hab}^x(t)$ is simply a notational shorthand which is used to make the equations simpler and easier to read. It is defined by the following two equations.

$$I_{hab}^x(t) \equiv \prod_{i=t-x+1}^{t} (1 - \alpha\tau - \tau I(i)) \quad (5)$$

0-7803-2431-5/95 $4.00 © 1995 IEEE

$$I^0_{hab}(t) \equiv 1 \tag{6}$$

Substituting Equation 2 and Equation 3 into Equation 4 yields a simplified form for $W(t_0 + \Delta t)$.

$$W(t_0 + \Delta t) = W^1_{eq} + (W^K_{eq} - W^1_{eq})\phi^{\Delta t} \tag{7}$$

$$\phi \equiv 1 - \alpha\tau - \tau \tag{8}$$

It is important to notice that some error is introduced here by assuming that $W(t_0)$ is equal to W^K_{eq}. This error approaches zero as t_0 approaches infinity with $I(t) = K$ for a long time prior to t_0. This assumption means that some of the conclusions which may be drawn from Equation 7 cannot be generalized to inputs which have multiple pulses without introducing some error.

It is apparent that $W^1_{eq} \le W(t_0 + \Delta t) \le W^K_{eq}$ so long as $\alpha\tau + \tau \le 1$. When this additional restriction is placed on α and τ an interesting definition can be made. Let A_I be the ratio between the decrease in W(t) due to the pulse and the range of W(t).

$$A_I \equiv \frac{W^K_{eq} - W(t_0 + \Delta t)}{W^K_{eq} - W^1_{eq}} = 1 - \phi^{\Delta t} \tag{9}$$

We can also determine the half-life of habituation, H_h. This is the pulse length required to achieve $A_I = .5$; that is to decrease $W(t)$ by half its range.

$$H_h = \frac{-\log 2}{\log \phi} \tag{10}$$

So far we have determined that so long as $\alpha\tau + \tau \le 1$, the values of $W(t)$ produced in response to the input pulse are bounded between W^K_{eq} and W^1_{eq}. Initially for $t < t_0$, $W(t)$ is assumed to be equal to W^K_{eq}. During the pulse it decays geometrically with rate, ϕ, toward the W^1_{eq} asymptote. The longer the pulse the closer $W(t)$ approaches its lower bound at W^1_{eq}. Values of $W(t)$ for $t_0 < t < t_0 + \Delta t$ can be calculated by substituting $t - t_0$ into Equation 7 in place of Δt. In order to calculate $W(t)$ for all values of t, the only remaining step is to determine W(t) for $t > t_0 + \Delta t$. By substituting Equation 7 into Equation 4 the following equation for $W(t_0 + \Delta t + i)$ can be determined for all positive integer values of i.

$$W(t_0 + \Delta t + i) = W^K_{eq} - (W^K_{eq} - W(t_0 + \Delta t))\phi^i_K \tag{11}$$

$$\phi_K = 1 - \alpha\tau - \tau K \tag{12}$$

Another value of interest is H_d, the halflife of dishabituation. This is the amount of time required, after the pulse has passed, for $W(t)$ to rebound halfway back to its original equilibrium value, W^K_{eq}.

$$W(t_0 + \Delta t + H_d) \equiv \frac{W(t_0) + W(t_0 + \Delta t)}{2} \tag{13}$$

By making use of Equations 4 and 7 and simplifying, the following equation can be derived for H_d.

$$H_d = \frac{-\log 2}{\log \phi_K} \tag{14}$$

The values, H_d and A_I, are important because they can be used along with the current habituation value to determine how long ago the last input pulse was observed.

$$i = \frac{\log(W^K_{eq} - W(t_0 + \Delta t + i)) - \log(W^K_{eq} - W^1_{eq}) - \log A_I}{log \phi_K} \tag{15}$$

This equation, however, can only be used if K is a constant known value and Δt is a known value. The latter assumption can be relaxed if $A_I \simeq 1$, i. e. when Δt is sufficiently large. Then the time of occurrence of any pulse with $\Delta t \ge \Delta t_{min}$ can be estimated by leaving out the $\log A_I$ term in Equation 15. If the value of K varies, for example due to background noise, then the value H_d becomes important in determining the error in estimating i due to the variation in K. For large values of i the error in estimating i becomes large because the $\log(W^K_{eq} - W(t_0 + \Delta t + i))$ term approaches $-\infty$ and any variation in W^K_{eq} becomes more and more important.

As long as K is constant, i can be approximated without knowledge of A_I or Δt. The approximation for i is given in Equation 16, and Equation 17 gives the maximum possible error in the approximation.

$$i \simeq i_{app} = \frac{\log(W^K_{eq} - W(t_0 + \Delta t + i)) - \log(W^K_{eq} - W^1_{eq})}{log \phi_K} \tag{16}$$

$$i_{err} \equiv i_{app} - i \;=\; \frac{\log(Weq^K - W(t_0 + \Delta t_{min}))}{\log \phi_K} - \frac{\log(W^K_{eq} - W^1_{eq})}{\log \phi_K} \tag{17}$$

Desired values of A_I, H_d, and Δt_{min} can be used to calculate α and τ.

Obviously, there are some limitations in the single habituator model. The amount of information which is obtainable for even a simple pulse input model is limited to the time of the most recent pulse. Even this small amount of information can be powerful, however. For example suppose each input presented to a habituated MLP is an indication of the occurrence of some specific feature in the signal. With the single habituator model the most recent time of occurrence of each input feature can be determined. The order in which these features were observed can, of course, also be determined. The habituated MLP does not have as much local temporal information about its inputs as a TDNN, but at the same time it does not have hard limits on the length of its memory. Still the amount of temporal information encodable by a single habituator is limited. In order to overcome this limitation it is necessary to use multiple habituators for each input.

When using multiple habituators, it is difficult to analytically determine input information from habituation values, because of the necessity of solving simultaneous nonlinear equations. However, it is relatively simple to show that information is gained by using two habituators. Once again, we will consider the same single pulse input model given in Equation 2. With two habituators it is possible to determine the length of a pulse, Δt, along with the time at which it occurred. First i is calculated from one of the habituators by using Equation 16 and thus taking the approximation $A_I = 1$. The α and τ parameters for the first habituation unit must be chosen so that this approximation is valid for some Δt_{min} which is less than the pulse lengths in which we are interested. If arbitrary pulse lengths are desired take $\Delta t_{min} = 1$. Once we have i we can determine A_I for the second habituator and then Δt as follows.

$$A_I = \frac{W_{eq}^K - W(t_0 + \Delta t + i)}{\phi_K^i (W_{eq}^K - W_{eq}^1)} \tag{18}$$

$$\Delta t = \frac{\log(1 - A_I)}{\log \phi} \tag{19}$$

Several conclusions can be drawn from this analysis. For one thing habituation does encode some temporal information. Also it seems that habituation is particularly good at encoding the time of occurrence of the most recent input pulse. Such information can in fact be encoded using only a single habituator for each input. This kind of information is particularly useful if each input is an indicator of the occurrence of a particular event or feature. A single habituation unit per input system can encode the order of occurrence of the most recent examples of each such feature. A TDNN could not encode such information unless all of the features in question occurred within the time window of the TDNN. The fact that habituation can also provide temporal information for less restricted forms of inputs is demonstrated empirically in the following section. It is, however, interesting to note that the minke whale song data set is also somewhat pulselike in nature. Certainly habituation can outperform TDNNs in situations where the inputs are constant for a large number of presentations, because the habituation values can contain information about the input before the string of constant values, but a finite window length TDNN cannot. With the habituation unit such information decays to zero over time, but with a TDNN it goes to zero as soon as the TDNNs time window is filled with constant inputs.

3. EXPERIMENTAL RESULTS

In order to ascertain the usefulness of habituation as a means of encoding temporal information, several experiments are performed using various habituated and unhabituated MLPs and TDNNs. The experiments performed utilize two different data sets. One of these sets consists of four different minke whale songs. The other contains two types of whale cries and two types of porpoise whistles. For conciseness, the minke whale data set is referred to as data set 1 (DS1) and the other set is called data set 2 (DS2). Table 1 lists the number of signals per class for the training and test subsets of each data set. Table 2 lists the average number of feature vectors per signal for each signal class. Each feature vector is 8 dimensional and denotes signal energy in 8 frequency bands [6]. The classes 1-4 in data set 1 are obviously not the same as the classes 1-4 in data set 2, but the have been tabulated as if they were, in order to create compact tables.

Our investigation demonstrates that MLPs and TDNNs with habituation preprocessing units outperform unhabituated TDNNs of similar complexity on the data sets examined. A typical result is shown in Figure 1, where the percent correctly classified for four different networks is plotted as a function of the number of hidden units. The experimental networks shown are the HTDNN, a TDNN with habituation preprocessing, and the HMLP, an MLP with habituation preprocessing. A TDNN and an MLP without habituation preprocessing are also shown for comparison. Each TDNN shown had a five sample time window. The results were generated using DS2. The results for DS1 were similar, with the habituated MLP outperforming the unhabituated MLP and TDNN, by 34 percent and 11 percent respectively. All results were determined after the presentation of each new input feature vector rather than at the end of each signal.

Table 1: Number of Instances of Each Signal Type

Class	I	II	III	IV
DS1 train set	5	5	5	5
DS1 test set	10	10	10	10
DS2 train set	5	5	3	3
DS2 test set	5	5	3	2

Table 2: Average Number of Feature Vectors Per Signal

Class	I	II	III	IV
DS1 train set	242	182	254	187
DS1 test set	233	177	254	174
DS2 train set	76	61	70	63
DS2 test set	67	65	82	85

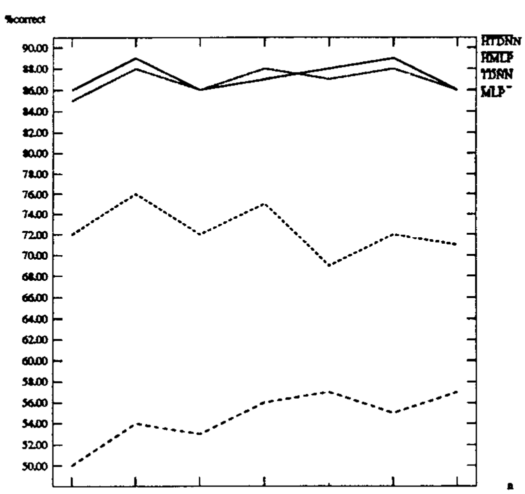

Figure 1: Comparison of Habituated and Unhabituated Networks

In our experiments the parameters α_i and τ_i are set to constant values such that $\alpha_i = A$ and $\tau_i = T$ for all inputs, I_i. The habituated networks achieve peak performance within a narrow range of T values. However, the dependence on A is not as significant, and habituated networks consistently outperform networks without habituation for a wide range of values of A and T. Figure 2 demonstrates the effect that varying T has on the mean square error of a habituated MLP on DS2. Figure 3 demonstrates the effect of varying A. The mean square error for an unhabituated MLP on the same data set is 0.135. The MSE for an unhabituated TDNN with a five sample time window is 0.083.

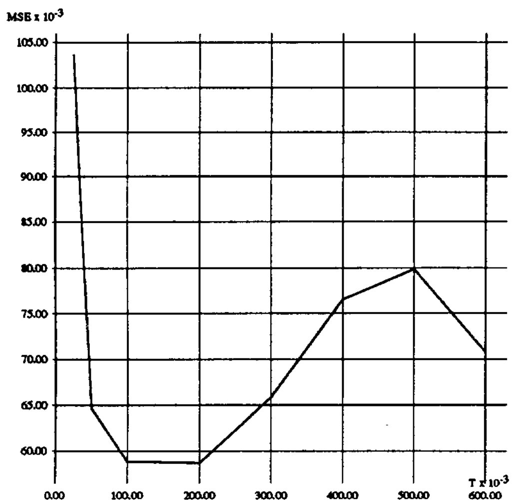

Figure 2: Performance of HMLP with varying T and A=0.2

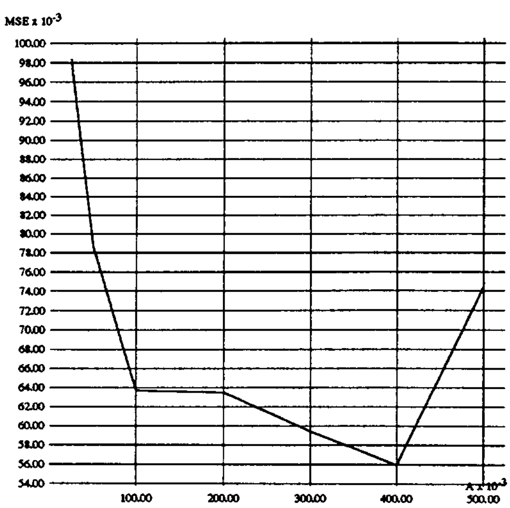

Figure 3: Performance of HMLP with varying A and T=0.05

For a more detailed account of the experiments performed see [7].

4. CONCLUSIONS

The theoretical analysis presented demonstrates that habituation can encode meaningful temporal information for simple types of input. Specifically, the order of last occurrence of a set of Boolean events can be established with a single habituator per input network. Empirically, single habituation unit per input networks are shown to outperform unhabituated MLPs and TDNNs for the two sonar data sets examined. Neither of these data sets is limited to Boolean events, so the range of inputs for which habituation is useful is obviously not limited to the simple case which is analyzed.

The habituated MLP and TDNN dramatically outperform the unhabituated MLP and TDNN for the data sets examined. The habituated MLPs have two major advantages over unhabituated TDNNs. First, they have access to long term information which TDNNs do not. Secondly, they have fewer trainable parameters. For data sets in which long term temporal information is unnecessary, the primary advantage of habituated MLPs is lost. Even in this case, however, habituation can be used to advantage. Because of the reduction in complexity for habituation as compared to time windowing, a habituated TDNN may be designed with similar performance and less complexity than an unhabituated TDNN. The reduction in complexity is important because it results in faster training and better generalization.

5. REFERENCES

[1] Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., and Lang, K. "Phoneme Recognition Using Time Delay Neural Networks," *IEEE Transactions on Acoustics, Speech, and Signal Processing.* ASSP-37, pp 328-339, 1989.

[2] Robin, D., Abbas, P., and Hug, L. "Neural Responses to Auditory Patterns," *J. Acoust. Soc. Am.* Vol. 87 part 4, pp 1673-1682, 1990.

[3] Wang, D. and Arbib, M. "Modeling the Dishabituation Hierarchy: the Role of the Primordial Hippocampus," *Biol. Cybern.* Vol. 67, pp 535-544, 1992.

[4] Byrne, J. H. and Gingrich, K. J. "Mathematical Model of Cellular and Molecular Processes Contributing to Associative and Nonassociative Learning in *Aplysia*," *Neural Models of Plasticity*, pp 58-70, 1989.

[5] Bailey, C. and Kandel, E. "Molecular Approaches to the Study of Short-Term and Long-Term Memory," *Functions of the Brain.* pp 98-129, Clarendon Press, Oxford, 1985.

[6] Ghosh, J., Deuser, L., Beck, S. "A Neural Network Based Hybrid System for Detection, Characterization, and Classification of Short-Duration Oceanic Signals," *IEEE Journal of Oceanic Engineering*, Vol. 17, No. 4, 1992.

[7] Stiles, B. W., "Dynamic Neural Networks for the Classification of Oceanographic Data," Master's thesis, Department of Elec. Eng., Univ. of Texas at Austin, 1994.