A Gaze-Addressing Control Interface

Using Artificial Neural Networks

Gabriel Baud-Bovy* and George G. Lendaris**
*Graduate Student, Electrical Engineering Department
**Professor of Systems Science & Electrical Engineering
Portland State University, P.O. Box 751, Portland, OR 97207

ABSTRACT

A gaze-addressing communication interface enables a severely disabled person to communicate. A menu of communication objects is displayed on a TV screen, each in a specially designed stimulus field; the person selects an object (e.g., a letter, a command) by gazing at it; an electrode implanted in the visual cortex senses the brain-wave response evoked by the visual stimulus associated with the communication object. Artificial Neural Network (ANN) technology is used to analyze the brain-wave response and identify the gazed-at communication object. The brain-wave responses are deterministic, low amplitude signals embedded in a colored noise with similar temporal and spectral characteristics. Their identification is deemed a challenging task. The result reported in this research are promising (90% of correct classifications with backpropagation ANNs). In addition this research yielded several interesting observations that may be helpful in future work related to such interfaces.

I. INTRODUCTION

The function of a Gaze-Addressing Communication Interface (GACI) is to enable a severely disabled person to communicate. An example situation is the case of amyotrophic lateral sclerosis or of severe cerebral palsy, wherein the sensory systems are intact and the eye movements still under good control. In this situation, the person’s gaze is the most efficient means to transmit information [Sutter, 1992].

An operational scenario is as follows: various communication objects are displayed on a T.V. screen, each in a specially designed stimulus field; the person selects an object by gazing at it (e.g., a letter, a command); an electrode implanted in the visual cortex senses the brain-wave response evoked by the visual stimulus associated with the object; the brain-wave response is analyzed and the object identified; and, finally, an action corresponding to the communication object is executed (e.g. calling the nurse, moving the bed).

The brain-wave responses are deterministic, low amplitude signals embedded in colored noise with similar temporal and spectral characteristics [Uncini, et al, 1990]. An approach using correlation techniques to analyze the brain-wave responses was developed by Sutter, et al, at the Smith-Kettlewell Eye Research Institute, San Francisco [Sutter, 1987, 1991, 1992a,b,c,d], and is currently implemented in their Brain Wave Interface (BRI). This provided the starting point and continuing context of the present research, whose objective is to develop artificial neural network (ANN) processing that affects successful classification of the brain-wave responses in a shorter period of time than the correlation method of Sutter. The neurophysiological aspects of the problem context are only briefly touched upon in this paper, the focus here being on the ANN aspects.

II. PROBLEM CONTEXT

A. Monitor Screen. The monitor screen is divided into 64 fields arranged in an 8x8 array, and rapid variation of the signal comprising the background of a field provides a visual stimulus. The signal variations are designed to yield a different visual stimulus for each field. Each field is labeled via a different communication object (letter, word, command), displayed in small characters inside each field (these remain fixed during the gaze time interval, while variations in the field background generate the visual stimulus).

B. Visual Stimuli. The 64 visual stimuli are deterministic, binary, periodic, orthogonal signals with desirable correlation properties (e.g. any segment of some specified length within a
period is approximately uncorrelated with any other segment of the same length). Each period of a visual stimulus coincides with a particular (binary) modulation pattern called a binary m-sequence [Marmarelis & Marmarelis, 1978; Sutter, 1987, 1991]. For example,

\[ x_1(t) = 0001001101011111 0001001101011111 \ldots \]

is a periodic visual stimulus based on the binary m-sequence of length 15 0001001101011111. Then, the other visual stimuli are obtained by consecutively shifting the first visual stimulus some number of bits, say \( L \) [with \( L=4 \) in the following]:

\[ x_2(t) = 001101011100001 001010111100001 \ldots \]

\[ x_3(t) = 0101111100010101 0101111100010101 \ldots \]

\[ x_4(t) = 111000100110101 111000100110101 \ldots \]

It is possible to verify that such visual stimuli are almost perfectly white.

C. Visual System. Ideally, an improved design for the visual stimulus signals could be obtained if a better model of the eye-brain visual processing system were known. The system has substantial non-linearities, as well as noise sources. Some of the known properties of the visual system that motivate the current design are as follows. The visual stimuli must always be modulated-in-time, since significant alterations of brain-wave patterns are induced by visual stimuli that change rapidly. The symbols (e.g., letters, words) used as communication objects on the screen do not, by themselves, provide the needed visual stimulus. Instead, time-varying signals (such as those defined above) are used to modulate the background in each field, and these provide the visual stimulus. This methodology allows realization of a hierarchical system of menus, with a potentially unlimited number of communication objects, since the objects can be changed without having to redefine the visual stimuli. The brain-wave response is mostly regulated by the visual stimulus at the center of the visual field [Sutter & Tran, 1992b]. This is due to the large amount of the visual system allocated by nature to processing the center of our visual field, where acuity is the greatest. This phenomenon, called cortical magnification, explains how the visual system is able to provide enhanced information about the specific communication object being gazed at, in comparison to adjacent objects within the field of view. Due to the lower contribution of these non-central objects in the field of view, the processing can act as if these contributions are noise on top of the signal representing the object actually being gazed at.

In designing the stimulus signals, Sutter noted that the visual system appears to be overall additive [Sutter, 1992a], thus, the signal-to-noise ratio would not degrade if the brain wave components were uncorrelated. They thus designed the stimulus signals to yield orthogonal brain-wave responses for each stimulus field.

D. Identification of the brain-wave responses. Consider the brain-wave identification process being divided into three stages: preprocessing, classification, and decision stages. Preprocessing could involve a number of operations, e.g., sampling, filtering, feature extraction, etc. In the present research, the brain-wave responses were simply sampled, but not filtered. The signals were probably over-sampled, because there was no reliable means available for optimizing the sampling rate.

Here, we focus on the classification stage, for which we compare performance of the correlation technique (used by Sutter) with performance achieved using ANNs to classify the brain-wave responses. From an abstract viewpoint, the classification stage is simply a box with 1016 input elements (maximum size of the brain-wave response) and 64 output elements (one for each of the fields on the monitor screen). Ideally, when the patient is gazing at the center of a defined field on the monitor screen, the ANN output element corresponding to that field should have a value near 1, while all the other elements should have a value near 0.

The decision stage completes the classification: it converts the real-valued output vector produced by the correlation technique or the ANN into a binary vector. The "desired outputs" would be a set of 64-bit binary vectors, each containing a single 1 (in a different place).

In the present context, it is possible to define a typology of the kinds of errors that can come about, and to introduce a bias in the classification and/or decision stage to decrease the number of errors of one type at the expense of another. This would be appropriate if the "cost" associated with a particular error type is very high.
III. EXPERIMENTS AND RESULTS

Experiments were carried out both, with the correlation technique used by Sutter, and with a feedforward/backpropagation ANN. The objective was to determine if the ANN could be designed to successfully classify the responses based on a shorter segment of the response than is achievable using the correlation technique.

A. Data. The data were provided by E. Sutter, of the Smith-Kettlewell Eye Research Institute. These comprise 99 brain-wave responses that were recorded by an external electrode applied on the scalp of a healthy subject. A binary m-sequence of length 127 was used as visual stimulus. The responses were sampled and quantified—with a maximum length of 1016 samples. We divided the data into two sets: 66 responses for defining the correlation templates and for training the ANNs; 33 responses to evaluate performance.

B. Correlation Technique. This method requires 64 templates—each one serving as an archetypal brain-wave response for one of the 64 different visual stimuli. The first template, T1, was obtained by averaging 66 (of the 99) brain-wave responses. The 63 other templates are derived from the first template by assuming that the templates are all identical except for a shift of time, because the visual stimuli are themselves identical except for a shift of time. The method proceeds by calculating the correlation between the current stimulus signal against each of the 64 templates. Ideally, the template yielding the highest correlation value corresponds to the field being gazed at, and, therefore, to the correct communication object.

Though the following additional step was not carried out by the Sutter group, we found that by normalizing the templates and responses (subtract the mean, and divide by the variance), the classification results of the correlation method improves. [Later in the paper, when we compare results of the ANN vs. the correlation method, we use this improved correlation process.]

C. Modularized ANN Architecture Studied. Early experiments were carried out with a single feedforward/backpropagation architecture, comprising 1016 input elements (maximum length of the brain-wave response), one hidden layer, and 64 output elements. Satisfactory results were not forthcoming from these experiments. Rather than pursue this avenue, since we (the researchers) had apriori information about the structure of the task to be learned by the ANN, it was decided to create an ANN architecture comprising a number of sub-ANNs (modules), fashioned in a way that reflected our a priori knowledge.

The architecture reported here uses a combination of 64 identical sub-ANNs, each having 1 PE in its output layer and each having to identify one of the classes. The desired output value is 1 if the sub-ANN identifies the input as a member of its class, and 0 otherwise. While each of the 64 sub-ANNs has the same architecture and weight values, connections from the input brain-wave response are successively shifted (we assume that a shift in time of the visual stimulus effects a similar shift in time of the brain-wave response).

We tried several sizes for the hidden layer (5, 10, 20, 30, and 50 processing elements). There were no dramatic differences, but 20 indicated a slight advantage. Results for this architecture are shown in Figure 1.

Figure 2 shows the percentage of correct classifications as a function of the length of the brain-wave response for various experiments (representative of all the experiments we performed with the modularized ANN architecture). As can be seen, the ANNs did indeed perform as well as the (improved) correlation technique, but did not demonstrate the hoped-for higher classification success rate based on shorter response segments. [It must be kept in mind, however, that the stimulus signals were optimized for the correlation method; there is still the opportunity to factor ANN requirements into the design of the stimulus signals.]

D. Training sets. It is always necessary to include "appropriately representative" data for any problem context. In the present context, this implies inclusion of responses evoked by each of the 64 visual stimuli into the training set. Because of the limited number of original data available, we assumed that a shift in time of the visual stimulus would produce a similar shift of the brain-wave response, and in this way were able to expand the training set. Training data for class 1 (corresponding to, say, the upper left corner of the T.V. screen) are the original responses (provided by Sutton); training data for class 2 (say for the next field on the T.V. screen) were simulated by shifting the class 1 data by one step; data for class 3 by shifting by two steps; etc.
During training, a module is presented with exemplars of each of the 64 different classes. For its own class, it is trained to yield a 1; for the other 63 classes, it is trained to yield a 0. In order to avoid the typical problem related to a training set containing a 63:1 ratio of patterns requiring a 0 output vs. those requiring a 1 output, the training set was augmented by duplicating the exemplars for the 1-output class 63 times, in order to even out the presentations.

E. Training Process with Short Responses. In some experiments, we trained the ANNs with short segments of the brain-wave responses, to determine whether classification time could be reduced. Since the ANNs have 1016 PEs in the input layer, the shorter responses were simulated by setting to zero the values at the PEs that normally receive the later response values. One set of experiments included training and testing sets of a consistent (shorter) size, and another set of experiments included training and testing sets with mixed sizes. These results were compared, and there are only minor differences between the results for the experiments trained solely on entire responses (Figure 2) and those whose training included shorter responses.

IV. DISCUSSION OF RESULTS

The following are more detailed comments about the experimental results, and a description of two phenomena which, if correctly addressed, could lead to improved classification performance. Similar phenomena could appear in other problem contexts, so there is motivation to understand these phenomena better.

A. Output and Average Output Values. Even though the final classification performance from the ANN and from the correlation technique are similar (Figure 2), there are major differences in the interim results of the two techniques. In Figure 3, for example, it may be noted that the values computed for the 64 classes are very different in each case. Also, the relative difference between the value of the first output element (whose desired output value is 1), and the value of the other output elements (whose desired output value is 0) is usually larger for the ANN than for the correlation technique.

This observation is confirmed by averaging, for each one of the 64 output elements, the 33 values that are associated with the responses in the testing set. A plot of the (64) average values associated with each one of the output elements is represented in Figure 4. As one can see, the average output values from the ANN are closer to the desired output than the average output values of the cross correlation technique. This observation is promising. On the negative side, the ANNs sometimes generate undesirable high values (the "peaks" that may be observed in Figure 3) that will cause the corresponding response to be misclassified (e.g., if the maximum value is used as the basis for making the classification decision). We next discuss how these "peaks" materialize during the training process.

B. Contrast-Enhancement. Figure 5 shows the evolution of the output values as training proceeds. The output values continue to get closer to the ideal during the training process, for up to as many as 15,000 presentations. This seems to contradict the results reported in Figure 1, which show that the classification performance of the ANN reaches its maximum very quickly (90% after 2,000 presentations). This may be explained as a contrast-enhancement type phenomenon. In Figure 5, the data suggest that output values that are already close to the desired outputs stay close (or get closer) at the expense of other values that get worse. In other words, the contrast among the output values is enhanced because the intermediate "gray" values disappear during the training process. At the end of the training process, only the values close to 0 (the "white" values) or close to 1 (the "black" values) remain. The improvement in the average output value during the training process is due to the fact that most of the small deviations from the desired output values (the "gray" value) disappear. However, the classification performance does not change since the misclassifications are mainly due to the "black" or "white" values. The clear conclusion is that just increasing the length of the training process will not improve performance of these experiments. We believe, instead, that it will be necessary to increase the size of the training set to obtain better results.

C. Nearby Misclassification. The term 'nearby misclassification' refers to the fact that misclassification occurs more often among the "early" classes, e.g. between classes 2 and 10 (see Figure 6). A partial explanation for this situation is to observe that the visual stimuli inducing
the brain-wave responses belonging to these classes are "close in time" to the visual stimulus corresponding to class 1 (there is smaller lag between x1(t) and x2(t) than between x1(t) and x3(t), etc.). But unfortunately, there is an asymmetry present: there are fewer misclassifications between classes 56 and 64 than between classes 2 and 10. There must be an additional (as yet unknown) reason to explain why the misclassifications occur mainly between classes 2 and 10.

We designed several experiments that specifically address this issue. The backpropagation algorithm was modified by linking the error evaluated by the criterion function to the class of the response. We tried various combinations of error factors. Unfortunately, the results of these experiments are not conclusive (between 87.9% and 88.8% correct classification). Nevertheless, we believe this approach has potential. It provides an alternative way to introduce expert knowledge into the ANN training methodology, by allowing one to choose the amount of "resources" allocated to the data in the training set. A simpler problem context would certainly be desirable to assist in understanding this approach better, both empirically and theoretically.

Finally, we offer the observation that the misclassifications due to the correlation technique and ANN technology (cf. Figure 6), albeit different, are caused by the same responses. This suggests that the misclassifications may well be due to attributes of the responses rather than to some peculiar characteristic of the methods used. We are not in a position, however, to discard these responses as outliers, because we do not yet have enough information about the characteristics of typical responses.

V. CONCLUSION

The guiding motivation for the present research was to improve the speed of classification in an existing gaze-addressing communication system (BRI) that was developed at the Smith-Kettlewell Institute at San Francisco by Dr. E. Sutter. In particular we explored the potential of ANN processing to replace the correlation technique currently used in the BRI. The initial objective was only partially achieved. Even though we definitely established that the ANN technology can perform at least as well as the (improved) correlation technique, none of the ANN experiments we created performed substantially better than the (improved) correlation technique.

One must notice however that this research was conducted under the severe handicap of having only 99 brain-wave responses (66 used as for the training set, and the remaining 33 as the testing set). While it is well known that fieldable results are dependent upon using large amounts of training data in contexts such as the present one, nevertheless, performance as good as that demonstrated here on less extensive data portends well for the fuller experiments. Future work in this area will require a (significantly) enhanced set of data, and will require expertise in neuro-physiology as well as in ANN technology. Despite the handicap of a paucity of data, the experiments brought a better understanding of the problem context, yielded several interesting observations, and helped to define lines of future research.

VI. REFERENCES


Personal Communications
Sutter E.E. (1992) "The BRI: Communication Through Visually Induced Electrical Brain Responses"
Sutter E. & D. Tran (1992a) "Communication Through Visually Induced Electrical Brain Responses"
Sutter E. & D. Tran (1992b) "The Field Topography of ERG Components in Man: 1. The Photopic Luminance Response"
Sutter E.E. Pevehouse B., Barbaro N. (1992) "Intra-Cranial Electrodes for Communication and Environmental Control"
Figure 1. ANN % correct classification on Train Set (dashed line) vs. on Test Set (solid line) as function of number of train cycles. [expt. x96]

Figure 2. ANN (solid line) vs. Correlation (dashed line). % correct as a function of response length used, ANN trained on full length responses.

Figure 3a. 3-D representation of ideal case, when all test responses are classified correctly. x-axis: 64 output elements; y-axis: 33 responses in the test set; z-axis: value given at output elements.

Figure 3b. ANN after 15,000 presentations (expt. x96).

Figure 3c. Correlation responses to test set (normalized).

Figure 4. ANN (solid line) vs. Correlation (dashed line): Average output values (expt. x96)

Figure 5. Evolution of outputs (expt. x96) at different stages of training.

Figure 6a. Classify brain-wave responses via (normalized) correlation method [full responses, and maximum method used in decision stage; x-axis: the 64 templates; y-axis: the test set of 33 responses.

Figure 6b. Classify brain-wave responses via ANN (expt. x96) [maximum method used in decision stage].