Final Report

Project Title:

Information Theoretic Extraction of EEG Features for Monitoring Subject Attention

PI. Jose C. Principe

Program Manager: Dr. Mark Motter

September 2000

Abstract

The goal of this project was to test the applicability of information theoretic learning (feasibility study) to develop new brain computer interfaces (BCI). The difficulty to BCI comes from several aspects: (1) the effective data collection of signals related to cognition; (2) the preprocessing of these signals to extract the relevant information; (3) the pattern recognition methodology to detect reliably the signals related to cognitive states. We only addressed the two last aspects in this research. We started by evaluating an information theoretic measure of distance (Bhattacharyya distance) for BCI performance with good predictive results. We also compared several features to detect the presence of event related desynchronization (ERD) and synchronization (ERS), and concluded that at least for now the bandpass filtering is the best compromise between simplicity and performance. Finally, we implemented several classifiers for temporal pattern recognition. We found out that the performance of temporal classifiers is superior to static classifiers but not by much. We conclude by stating that the future of BCI should be found in alternate approaches to sense, collect and process the signals created by populations of neurons. Towards this goal, cross-disciplinary teams of neuroscientists and engineers should be funded to approach BCIs from a much more principled view point.

I- What is a Brain Computer Interface?

A Brain computer interface (BCI) is a device that allows a person to control its environment by generating specific brain wave patterns. Using the term "thoughts" to describe the mental activity that produces the brain wave patterns, we can say that the goal of a BCI is to control the external environment directly with thought. There have been at least four types of experimental paradigms to accomplish this goal:

- the generation of event related potentials (ERPs) by appropriate external stimulation (computer displays),
- imagination of movement (hands or feet),
- engagement in mental activity of different nature (e.g. math versus music),
- biofeedback

In all these cases, the electroencephalogram (EEG) was utilized as the variable that accessed and quantified the mental state. However, they differ in detail. For instance, the ERPs are faint, fast

transients related to the processing of expected stimulus (P 300) [6] [12], [14], while the imagination of movement is translated in synchronization and desynchronization in the alpha and beta frequency bands over the motor cortex [2], [9], [15], [21], the math versus music tests attempt to find global changes in the EEG power spectrum over the hemispheres [1], and finally the biofeedback approach instructs the patient to create alpha bursts [14]). From all these signals the one that has been shown more reliable is the alpha and beta changes over the motor cortex [15].

BCIs can be used in a variety of situations. Quadriplegics or patients suffering from amyotrophic lateral sclerosis or any other type of the so called locked-in-syndrome can use a BCI. Human operators in highly instrumented environments (such as fighter pilots) can benefit from BCIs to complement their hands.

The difficulties of creating useful BCIs are mostly derived from the low signal-to-noise ratio of the EEG signals and the intra and inter subject variability. This creates a very low bit rate channel (mostly binary decisions) that has hindered applicability. Presently, the groups working in this area have focused their work in the locked-in-syndrome cases, where the subjects have all the time in the world to make decisions. Even in this area, other methods using eye blinking or gaze have shown practical advantages. Hence, we conclude that much more work is needed in the design of effective signal interfaces to the brain. This project does not attempt to create better signal collection methodologies. It assumes that the available signals are the EEG channels collected over the motor cortex areas, and the goal is to make as reproducible as possible BCIs by testing several signal processing and machine learning algorithms.

II- Experimental Setup

We utilized the data collected in Dr. Gert Pfurtscheller Laboratory in Graz, Austria for our studies, since we would like to obtain a straight comparison for our processing algorithms with his own results. This group uses the event related synchronization (ERS) and event related desynchronization (ERD) in the alpha and lower beta bands over specific areas of the motor cortex. For instance, imagination of hand movement produces an ERD pattern focussed on the contralateral hand area. In some subjects the ERD is followed by a ERS. We will be using this phenomenon to distinguish between YES/NO commands, where YES is associated with an ERD/ERS at the left motor area

and NO is an ERD/ERS at the right motor area. Since we want a real-time system, no averaging will be performed (which, as is well known improves the SNR by a factor equal to the square root of the number of trials).

The subject is seated in front of a monitor and tries to control his mental activity in accordance with cues given on the monitor. Each trial last 8 seconds, and they are separated by a short 0.5 sec break. Forty trials make a run, and a session is 4 runs. Each run consists of 20 left trials and 20 right trials in randomized order. Each trial has a period of fixation to a cross that appears in the screen (3 sec), and a beep occurs (after 2seconds) to signal the occurrence of a cue (left or right arrow) for 1.25 second. The subject is instructed to imagine movement of the corresponding left or right hand for 4 seconds. Feedback in the form of a growing rectangle proportional to the output of the classifier is also displayed in the screen.

Data was collected from 6 sessions of three young healthy subjects (male college students) g3, g7 and i2. They were selected because they display three levels of performance (very good, moderate and poor respectively). Four electrodes were placed 2.5 cm anterior and posterior to C3 and C4 to obtain two bipolar EEG channels over the left and right motor areas. The signal was sampled at 128 Hz, with an anti-aliasing filter at 30 Hz. The trials were visually inspected for artifacts and these trials removed. Table I shows the number of trials used per subject (trial 1 is used to set the feedback). Due to the subject variability, the data was segmented individually per subject. as shown below the subject identification.

Table 1: Data used for the tests

Session	g3 (3-7 sec)	g7 (4.5-7.25 sec)	i2 (3.5-7.6.75 sec)
2	159	154	139
3	156	157	153
4	160	157	160
5	160	138	152
6	152	152	131

III- Feature Selection

The knowledge of the ERD and ERS shows that the alpha (9-13 Hz) and low beta band (20-24 Hz) contain the information for ERD and ERS (see Figure 1). We used a 5th order Butterworth IIR filter and integrated for 1 sec to obtain signal power. Downsampling to 8 Hz is done next. Hence using the a priori knowledge about the task we can quantify the EEG activity with a 4 feature vector (alpha and beta band powers at the left and right motor areas). The first important questions that we asked were: what additional features can we obtain from the data? is this feature set optimal? To rate the feature sets we decided to use an information theoretic measure.
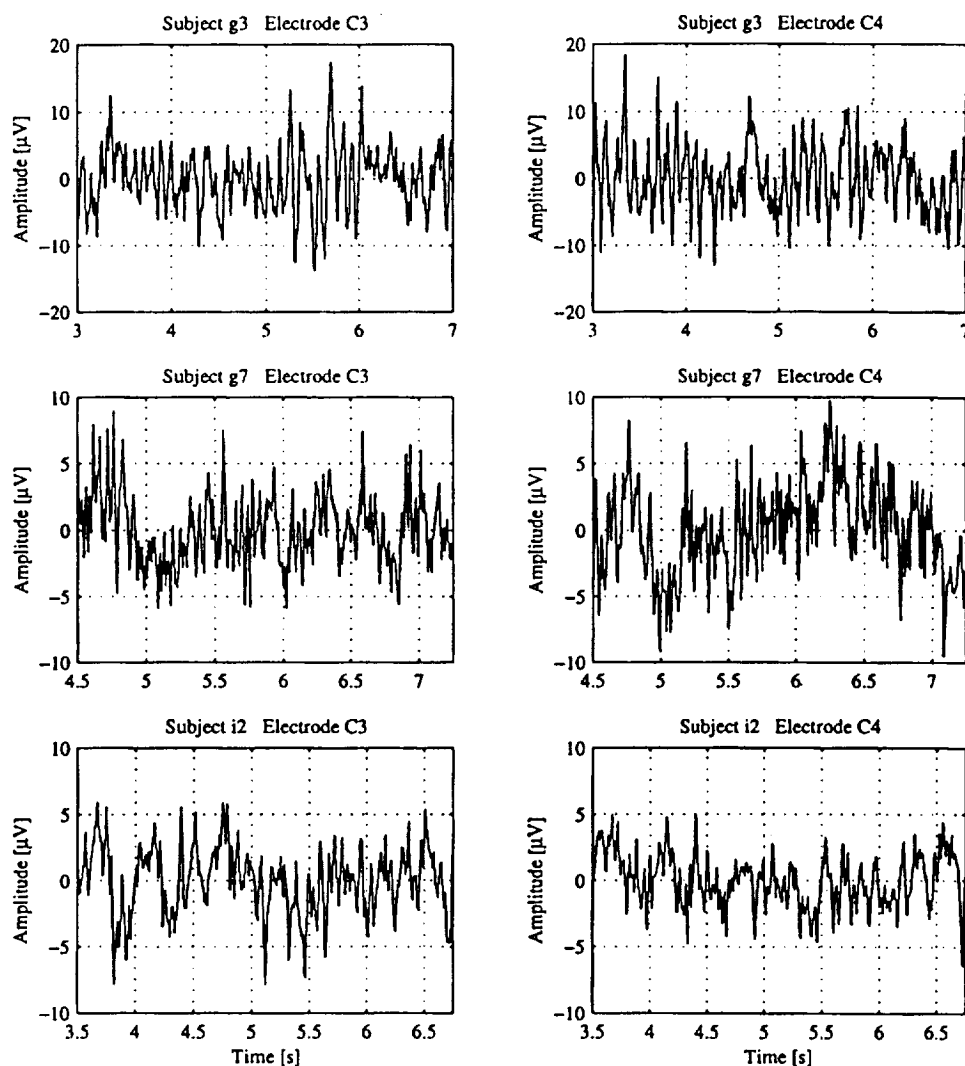
Figure 1. Examples of ERD and ERS in the EEG.

III-1 Distance Measures

We started with the well known Bhattacharyya distance [5] between two distributions $p_1(x)$ and $p_2(x)$ given by

$$D_B = \sqrt{P_1 P_2} \int p_1(x) p_2(x) dx$$

where P_1 and P_2 are the a priori probabilities of each event (here equal to 0.5). When the pdfs are Gaussian the integral can be computed in closed form as

$$D_B = 0.5 e^{-\mu(0.5)} \qquad \mu(0.5) = 0.125 (m_1 - m_2) \left(\frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} (m_1 - m_2) + 0.5 \ln \frac{\left| \frac{\Sigma_1 + \Sigma_2}{2} \right|}{\sqrt{|\Sigma_1| + |\Sigma_2|}}$$

where m and Σ are respectively the mean and covariance matrix of each multidimensional Gaussian distribution. The Bhattacharyya distance in our experiments is used as a measure of the separation between the left and right hand movements, hence, as a measure of performance in the BCI. After computation of the Bhattacharyya distance (BD), we can predict performance of the classifiers and select the subjects in advance with a pre-trial. If a subject gives a low BD, we know that he or she is unable to provide good discrimination and the task of developing a BCI will be very difficult. On the other hand, for pattern recognition the BD also provides a way of ranking features for the classifier.

Early results showed that the estimation of these covariance matrices for our four dimensional small size data sets were meaningless (due to inaccuracies in the estimation of the covariance). We therefore propose here a new method where the MEDIAN operator is used instead of the mean. This is known in the digital signal processing literature this is called rank filtering [3]. We therefore adopt the same definition for D_B but substitute the mean operator by the median (for both the mean m and covariance Σ). The advantage of the median Bhattacharyya distance (MBD) over BD is that the estimations will be much more robust.

We would like to improve on this distance measure even further, but we could not do it in this

grant due to lack of time. Our idea is to use the Information Potential [16] to estimate the distance between the two data sets. In fact we believe we can even use directly this measure to project the data to a subspace where classification can be accomplished.

III-2 Alternative Features

There are many other alternative EEG features besides alpha and beta band powers that we can utilize in the design of a BCI. Now that we have a measure of quality (the (MBD)) we can implement them and compare. In this research we utilized the following

AR coefficients: We computed a 6th order autoregressive (AR) model using the RLS (recursive least square) algorithm. The information is contained in the coefficients, so we performed an averaging over 16 time steps. This feature set is 12 dimensions.

Hjorth parameters: Hjorth [10] in a very influential set of papers proposed the activity, mobility and complexity as three time domain descriptors of EEG activity. Activity is simply the variance of the signal, the mobility the ratio of the variance of the first derivative of the signal over the variance of the signal, while the complexity is the ratio of the mobilities of the first derivative and of the signal itself. The integration is done over 1 sec, with a first order recursive integrator. The values for complexity, mobility and complexity were averaged over 16 time steps, providing a six dimension vector.

Principal Component Analyzer: The goal is to discover the direction in space where the signal has the largest power. We can do this very simply by implementing a temporal PCA network trained with the Sanger's rule [19]. This network is effectively an optimal filter bank where the square of the outputs are the eigenvalues of the time autocorrelation matrix [17]. In this case we choose an embedding of 16 (16 delays which gives 16 PCA components). Once again the outputs were smoothed with a window of 16 samples. We end up with 32 features with PCA. However, since the signal is noisy, we can not take simply the first 2 features of each decomposition (for a total of 4 features from both channels), because the largest component will very likely represent the background EEG activity, not the signal of interest. The selection of the best components requires therefore some form of metric, and we decided to use the 2 projections per channel that produced the best MBD.

IV- Results

IV- 1 Preprocessors

At this point we can compare these 4 preprocessors for the BCI using the MBD. Figure 2 shows the MBD as a function of time (during the trial) for the three subjects and using our 4 preprocessors.
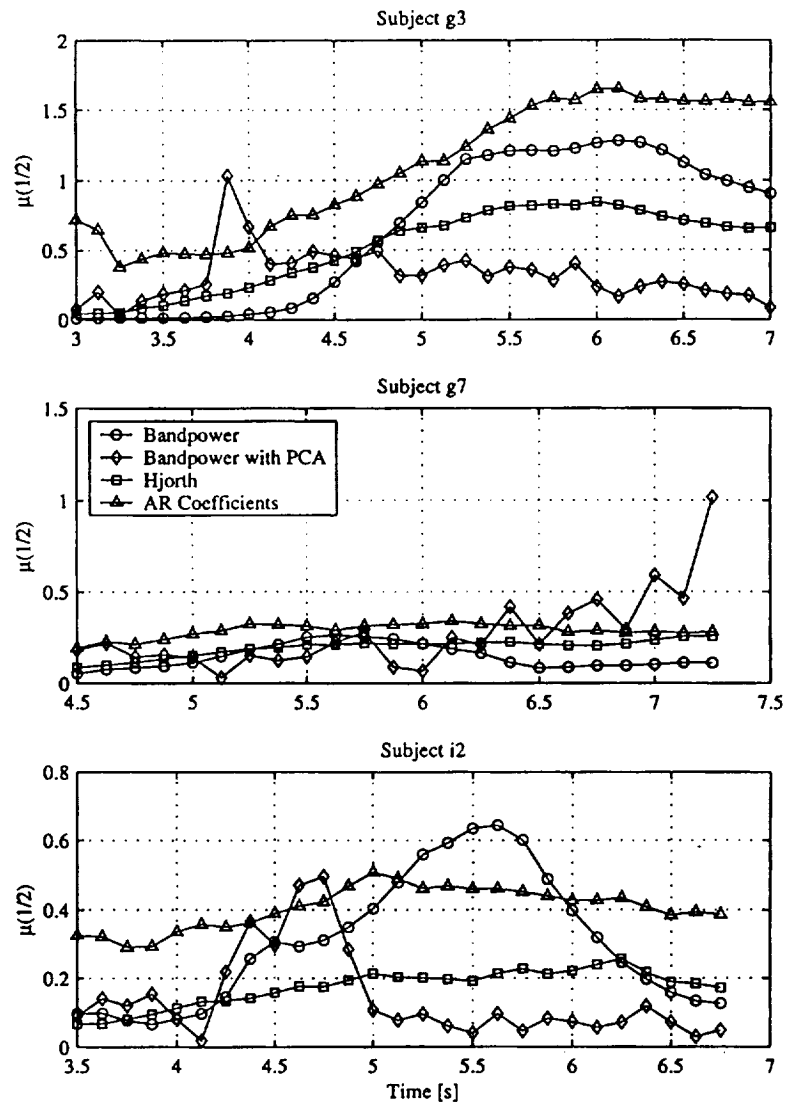

Figure 2. MDB across windows for the 3 subjects and different preprocessors.

We can immediately see several interesting things in these plots: first the three subjects offer 3 different performances as computed by the MBD. Subject g3 provides the largest separation and

hence will yield the best classification results. Subject i2, yields an intermediate value, while subject g7 is the poorest. We can also see that the four preprocessors preserve differently the separability between the two experimental conditions (right and left hand movements) for the same subject. Overall the alpha and beta filters are judged the best due to the simplicity of the method. We were disappointed with the optimal filter bank and think that more work is needed to find out why we can not supplant the other methods. But we run out of time for further experimentation.

There are also other sets of features that we would like to experiment with, in particular the ones found in independent component analysis (ICA) [11]. ICA can be computed with our present algorithms of the information potential, but we thought we should first address more conventional processing to establish baselines. Further work in this area is also warranted.

One important question we did address was how good is the MBD as a predictor of performance for classification of brain states. With the best neural network (discussed in the next section), we created the following table that shows the classification accuracy and the upper bound estimated from the MBD.

Table 2: Estimated and actual error rates

subject	MBD upper bound (%)	error rate of ANN (%)
g3	13.3	8.3
g7	38	31.9
i2	25.4	22.1

As we can see the MBD is a good estimator of performance for classification in BCI. For a more detailed analysis we plot in Figure 3 the error for the neural network as a function of window location to qunatify better the correlation with the Bhattacharyya distance.
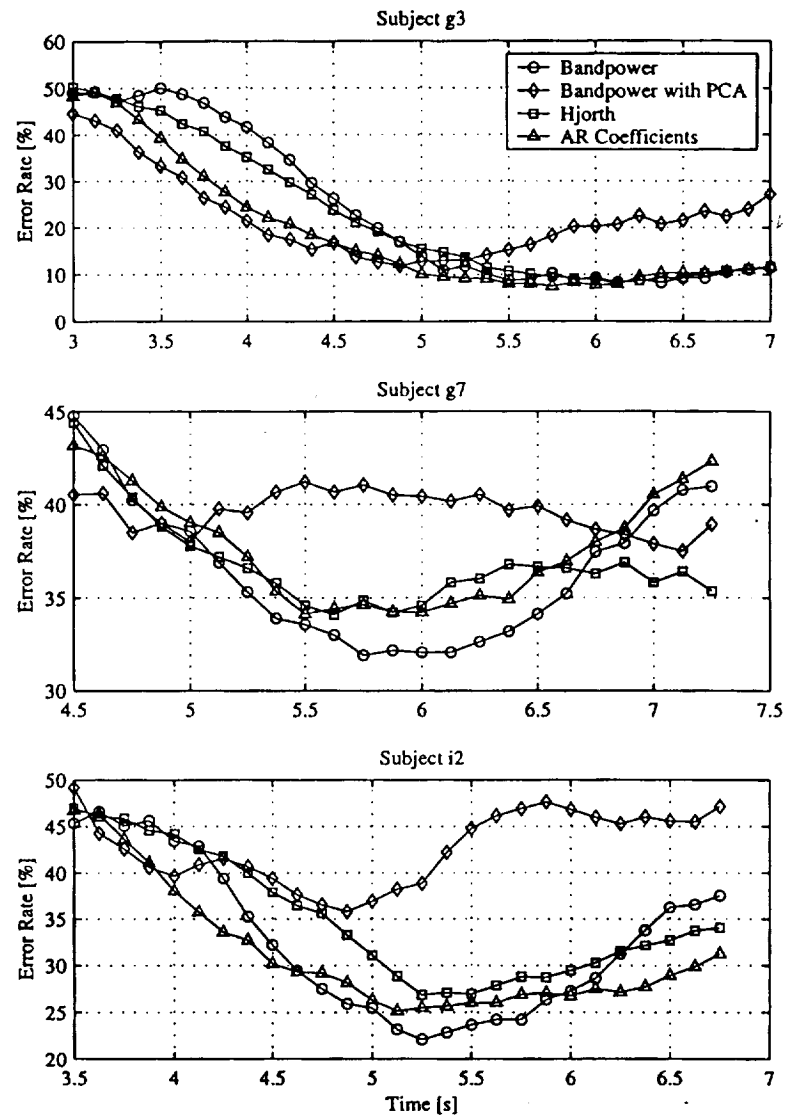
Figure 3. Classification error across windows. Compare with Figure 2.

Notice that this is a very nice time saving feature, because we do not have to go through long data collections, and delve into the experimental selection of parameters and classifier training. In some cases there are simply no good parameters because the subject themselves do not follow the expected population trends.

IV- 2. Classifiers for BCIs

Notice that the task of classifying EEG patterns for a BCI is a nontrivial problem. First, the data is noisy as we already have shown, but the other difficulty is that the signals exist in time, so effectively we are interested in performing temporal pattern recognition. Pattern recognition is normally applied to static patterns [5], so conventional techniques can not be applied directly to BCI. Hence, it is important to compared several types of classifiers with the goal of showing how performance changes with the structure of the classifier.

A very simple trick is to use a time window over the input data or in the first layer of our classifier to effect a time to space mapping [17]. However, the difficulty with this task is to select a good window size that maximizes classification accuracy. The TDNN is such a classifier [20]. We can also just let the classifier find the window that provides the best classification results. The gamma neural network implements this idea [18]. The problem with this approach is that training has to be done over time (BPTT) since the neural network became recurrent and the gradients exist in time.

One of the difficulties of temporal pattern recognition faced in training dynamic neural networks is what is the desired response over time. We have studied this problem some time ago [22] for spike detection, and concluded that the desired signal should be 1 when the transient occurs. When the event is not present the system should be trained with a don't care condition. We tested several (simple prediction or random values in the desired response). Here the problem is even more difficult because the ERD and ERS signals alternate in time and we are not certain where they occur. We can have an indirect measure of the difficulty by observing the MBD plots of Figure 2 and noting that the value varies over time. We therefore developed a method to perform classification without a numerical target [7]. Once again the method is based on mutual information but we only had time to scratch the surface. We published a paper with a simple classification problem but it clearly shows promise to improve classification in temporal pattern recognition [7]. More work is needed in this area since the algorithm has an intrinsic compromise between two information forces that we were not able to find for EEG signals. We proceeded with the comparison of more conventional classifiers: a linear discriminant based on Fisher's method, a perceptron, a TDNN, a gamma network, and a TDNN trained with dynamic targets (DT) as proposed in

[8]. The Table shows the error rates we obtained for each subject

Table 3: Error rates for different classifiers (%)

subject	Fisher	MLP	TDNN	gamma	DT
g3	7	6.8	7.1	7.7	4.8
g7	32	34	30.8	34	32.4
i2	22	20	21	20.6	20.8

As we can see from the table, the neural network topology only affects slightly the general trends of accuracy. If a subject is able to create ERD/ERS (as subject g3) then any of the classifiers work reasonable well (50% is random performance). We can however see that for simplicity the linear Fisher discriminator is unbeatable. If we want to achieve the top performance than we have to go to a neural network trained over time, with a special desired signal that is adapted during training. This is rather complicated and does not change the fact that there are subjects simply unable to provide good brain patterns for BCI devices. It does however state that there is more information for classification when we look at the time series (i.e. the signal has repeatable dynamics), but the issue is how to extract the information efficiently.

V- Conclusions

During this research period we have investigated the possibility of implementing BCIs using EEG signals created in the motor cortex when hand movements are imagined. We have set out the stage to continue working in this area by investigating the problem from first principles. We have now the baseline results with respect to which we can compare improvements in the methods for both temporal pattern recognition and feature extraction. Therefore, we are ready to continue our work in the implementation of the newly introduced information theoretic learning algorithms developed at the CNEL.

VI- Publications

Haselsteiner E., Principe J., "Supervised Learning without Numerical Targets", Proceedings of NC'2000, Berlin, 2000.

Another paper is under preparation to be submitted to the IEEE Trans Biomedical Engineering, and one more conference proceedings to be submitted to the neural network community.

VII - Future Research

After working during the past year with the brain signals from motor cortex, we have the following observations:

1- Quantitative improvements to BCI can only come with a more sophisticated data collection and experimental design schemes. As we could see from the experimental design the subject was engaging the full brain to make a binary decision. This seems the real bottle neck of this methodology when electrodes are placed over the scalp. So we consider that the true potential of BCIs is centered in new experimental paradigms, the development of multi-electrode arrays, devices and algorithms that will be able to collect more specific brain patterns. An integral part of this endeavor is to address the neuroscience fundamentals required to understand better how neural assemblies communicate and process information. This is a very important area of research in the frontiers of neuroscience and biomedical engineering, that we can truly name *Neural Engineering*. NASA funding for this area will have a potential large impact in the computational neuroscience and biomedical engineering communities. We, at the University of Florida Computational NeuroEngineering Laboratory and our contacts with the Brain Institute and other colleagues in neuroscience, will be able to conduct the research necessary to make "brain chips" in analog VLSI technology that will be able to process on site the signals collected by multi-electrode arrays. To embark in this direction we will need funding on the order of $100 K per year to establish collaboration with a neuroscience group with the goal of developing the experimental design, the implementation of analog VLSI chips to process brain signals and extract their features.

2- Of course the feature extraction/pattern recognition methodologies we implemented in the first year need improvements. We propose to investigate the following aspects:

- Use the Information Potential and the Integrated Square Error to evaluate the separability between classes, instead of the MBD, which assumes Gaussian distributions.
- Improve the PCA optimal filter bank with an independent component analyzer using the minimization of mutual information.

- Find new features by using "Information Filtering", i.e. projecting the input data into a sub-space that best preserves separability between the two classes.

- Proceed with the classification of temporal patterns without numerical targets.

This work requires the present level of yearly funding ($50K).

References

1. Anderson C. and Sijercic Z., *"Classification of EEG Signals from Four Subjects During Five Mental Tasks. "* Proc. of the Conference on Engineering Applications in Neural Networks (EANN '96), Turku, Findland, pp. 407-414,1996.

2. Barreto A., Vicente L. and Persad *I., "Spatio-Temporal EEG Patterns Associated with Voluntary Motion Preparation",* Proceedings of the 17th Intematonal Conference of the IEEE Engineering in Medicine and Biology Society, pp. 857-858, 1995.

3. Coyle E., "Rank order operators and the mean absolute error criterion", IEEE Trans. Acous. Speech and Sig. Proc., 36 (1):63-76, 1988.

4. Daviss, B., *"Brain Powered",* Discover, May 1994, pp. 58-65.

5. Fukunaga K., Introduction to Statistical Pattern Recognition", Academic Press, 1990.

6. Gevins A., Leong H., Smith M. E. and Du R., *"Mapping cognitive brain function with modern high resolution electroencephalography. "* Trends in Neurosciences, v. 18, 429-436.

7. Haselsteiner E., Principe J., "Supervised learning witout numerical targets - an information theoretic approach", in Proc. Neural Computation NC'2000, Berlin, 2000.

8. Haselsteiner E., "Time series classfication using adaptive dynamic targets", Proc IEEE Workshop Neural Nets for Sig. Proc., pp 243-252, 1999, IEEE Press.

9. Hiraiwa A., Shimohara K. and Tokunaga Y., *"EEG Topography Recognition by Neural Networks, "* IEEE Eng. Med. & Biol. Mag., 9,(3),1990, pp, 39-42

10. Hjorth B., "EEG analysis based on time domain properties", Electroenc. Clin. Neurophys., 29:306-310, 1970.

11. Hyvarinen A. and E. Oja, "Independent Component Analysi:A Tutorial", Neural Networks, in Press.

12. Konger C., Principe J.C., "Neural network classification of event related potentials for the development of a new computer inter face", Proc. IJCNN'90, San Diego, CA, vol II, pp. 407-411

13. Lusted H., and Knapp B., *"Controllong Computers with Neural Signals",* Scientific American, October 1996, pp. 82 - 87.

14. McMillan G. R., Calhoun G. L., Middendorf M. S., Schnurer J. H., Ingle D.F. and Nasman V. T. *"Direct Brain Interface utilizing self-regulation of Steady State Visual Evoked Response (SSVER) ",* Proceedings RESNA 18th Ann. Conf., 1995.

15. Pfurtscheller G. et al., *"Differentiation between finger, toe and tongue movement in man based on 40 Hz EEG,"* Electroenceph. clin. Neurophysiol. 90,1994,456-460.

16. Principe, J., Xu, D., Fisher, J., Information Theoretic Learning, in *Unsupervised Adaptive Filtering*, Simon Haykin Editor, 265-319, Wiley, 2000.

17. Principe J., Euliano N., Lefebvre C., "Neural and Adaptive Systems: Fundamentals through simulations", Wiley, 2000.

18. Principe J., deVries B., Guedes de Oliveira P., "The gamma filters: a new class of adaptive IIR filters with restricted feedback", IEEE Trans. Signal Proc., '93, vol 41, #2, pp. 649-656.ÿÿ

19. Sanger T., "Optimal unsupervised learning in a single layerlinear feedforward neural network", Neural Networks, vol 12, 459-473, 1989.

20. Waibel A., T. Hananzawa, G. Hinton, K. Shikano, and K. Lang, Phoneme recognition using time delay neural networks, IEEE Trans. on Acoust. Speech and Signal Proc., vol ASSP-37, 328-339, 1989.

21. Wolpaw J. et al., *"An EEG-based brain computer interface for cursor control,"* Electroenceph. clin. Neurophysiol. 78 1991, 252-259.

22. Zahalka A., Principe J., "Transient detection using neural networks: the search for the desired signal", in Advances Neural Information Proc. Systems 5, Ed. Hanson, Cowan, Giles, pp. 688-695.