A Dynamic Channel Selection Strategy for Dense Array ERP Classification

Srinivas Kota[1], *Student Member, IEEE*, Lalit Gupta[1], *Senior Member, IEEE*, Dennis Molfese[2]
and Ravi Vaidyanathan[3]

Abstract-The goal of this paper is to introduce a new strategy to accurately classify event-related potentials (ERPs), recorded using dense electrode arrays, into predefined brain activity categories. The challenge is to exploit the enhanced spatial information offered by dense arrays while overcoming the significant increase in the dimensionality problem introduced by the large increase in the number of channels. These conflicting objectives are achieved by introducing a spatio-temporal array model to observe the dense array ERP amplitude variations across channels and time, simultaneously. To account for latency variations and EEG noise in the array elements, each spatio-temporal element in the array is initially modeled as a Gaussian random variable. A two-step process which uses the Kolmogrov-Smirnov test and the Lilliefors test is formulated to select the array elements which have different Gaussian densities across all ERP categories. Selecting spatio-temporal elements that fit the assumed model and also statistically differ across the ERP categories not only ensures high classification accuracies but also decreases the dimensionality significantly. The selection is dynamic in the sense that selecting spatio-temporal array elements corresponds to selecting ERP samples of different channels at different time-instants. Each selected array element is classified using a univariate Gaussian classifier and the resulting decisions are fused into a decision fusion vector which is classified using a discrete Bayes classifier. By converting an inherently multivariate classification problem into a simpler problem involving only univariate classifications, the dimensionality problem which plagues the design of practical multivariate ERPs classifiers is circumvented. Consequently, classifiers can be designed to classify the ERPs that are unique to an individual without having to collect a prohibitively large ERP data set from him/her. The application of the resulting dynamic channel selection-based classification strategy is demonstrated by designing and testing classifiers for eight subjects using ERPs from a Stroop color-test and it is shown that strategy yields high classification accuracies. Finally, it is noted that because of the generalized formulation of the strategy, it can be applied to various other problems involving the classification of multivariate signals acquired from multiple identical or multiple heterogeneous sensors.

Index Terms- decision fusion, dense electrode arrays, dimensionality reduction, dynamic channel selection, event-related potentials, spatio-temporal modeling.

S Kota, and L. Gupta are with Department of Electrical and Computer Engineering, Southern Illinois University, Carbondale, IL 62901 USA. (Corresponding author to provide ph: 618-453-7032; Fax: 618-453-7972; email: lgupta@siu.edu).
D. Molfese is with Developmental Neuropsychology Laboratory at the Birth Defects Center, University of Louisville, Louisville, KY 40292, U.S.A.

R. Vaidyanathan is with Bristol Robotics Laboratory in the University of Bristol, Bristol BS8 1TR, U.K. and the US Naval Postgraduate School, Monterey, CA, 93943.

I. INTRODUCTION

The primary objective of the work described in this paper is to introduce a new strategy to accurately classify event-related potentials (ERPs), recorded using dense electrode arrays, into predefined brain activity categories. ERPs are brain responses that are time-locked to the onset of an external event such as the presentation of an audio or a video stimulus. The accurate classification of ERPs is of the utmost importance because ERPs are used extensively in numerous human cognition studies and in clinical evaluations. Due to practical issues related with the data acquisition methods, lack of concentration, discomfort, and fatigue, it may be impossible to collect enough ERPs from a single subject to exceed the dimension of the ERP vector which then presents problems in accurately estimating the parameters for the design of multivariate ERP classifiers. The recent introduction of dense array networks such as the Geodesics Sensor net [1] to record electroencephalograms (EEGs) and ERPs offers a significant increase in the spatial resolution over the International 10-20 recording system [2]. However, the large increase in the channels also increases the dimensionality-related problems in the design and evaluation of practical ERP parametric classifiers. This paper introduces a new ERP classification strategy that is capable of exploiting information offered by the increased spatial resolution while simultaneously circumventing the dimensionality problem. The main advantage of the new strategy is that ERP classifiers can be designed without having to collect a prohibitively large ERP data set. Consequently, practical classifiers can be designed, even for single subjects, using a practical number of their own ERPs so that accurate ERP based assessments can be made regarding the clinical disorder that is unique to the subject.

The use of dense electrode arrays for recording EEGs and ERPs has increased significantly. References [3] through [14] are only a miniscule sample of dense array ERP based studies that focus on memory [3], emotion [4],[5], learning [6],[7],[8],[9], language [10],[11], reading[12], attention [13], and anxiety [14]. A dense electrode array is typically a system consisting of 64, 128, or 256 scalp electrodes, whereas, the International 10-20 specifies a maximum of 19 electrode sites. Because of the increase in the spatial resolution, it is possible to make more accurate inferences of the underlying brain activity in EEGs and ERPs. However, the biggest

challenge facing dense-array brain-waveform research is the task of analyzing the vast amount of data collected by the large number of channels. Analyzing all dense array data in real-time is impossible and even off-line analyses is quite formidable. Therefore, it is not unusual for researchers, using prior knowledge, to preselect a small subset of dense-array channels in which they most strongly expect the brain activity of interest to occur or to simply partition the channels into groups and represent the brain activity in each group by the group average. In many analysis studies the grouping may be intentional, for example, to study the effects over different hemisphere regions [5],[11],[13]. However, the true benefits offered by the large number of spatially distributed dense-array channels are not exploited if the grouping of the channels is strictly to simplify the analysis or for dimensionality reduction. The traditional statistical analyses methods such as analysis of variance, principal components, and factor analysis pool the ERPs collected from a large group of subjects with an assumption that the ERPs of the subjects in the group share similar features [11],[13],[15]-[19]. The goal of these analyses studies are clearly aimed at establishing within-group similarities and across-group differences in brain activity in large populations.

In general, the procedure for designing any parametric pattern classifier is to collect an ensemble of patterns from each pattern category, partition the ensembles into a training set and a test set, use the training set to estimate the classifier parameters, and use the test set to evaluate the classifier performance. For multivariate patterns, the mean and covariance are the most useful parameters estimated from the training set. For example, the covariance matrix is needed to determine the Gaussian discriminant function and to determine the principal components [20]. The accuracy of the estimates clearly depends on the number of vectors in the training set. If the number is less than or equal to the dimension of the pattern vector, the estimated covariance matrix is singular and is, therefore, not invertible. Moreover, only a subset of the total number of principal components (eigenvectors of the covariance matrix) can be determined. Therefore, the number of vectors in the training set needed to estimate the covariance matrix must exceed the dimension of the pattern vector. Even better covariance estimates are obtained by using a much larger number of vectors in the training set.

ERPs are multivariate signals with dimensions that are a function of the sampling rate and the duration. For example, an ERP of a given channel sampled at 250Hz over a one-second duration will have a dimension of 250 (number of time samples). The covariance matrix of the ERPs of the channel will, therefore, be of dimension 250×250. Consequently, the training set for each ERP channel classifier must have at least 251 ERPs for the 250×250 covariance matrix to be non-singular and to be able to determine the 250 eigenvectors. Furthermore, ERP vectors are also needed to test and evaluate the performance of the classifiers. Collecting such a large number of ERPs to design and evaluate the performance, especially from a single individual, is quite difficult in practice. This dimensionality problem is exacerbated significantly when attempts are made to fuse the ERPs of multiple channels in order to design data fusion classifiers that aim to exploit complementary information from several ERP channels [21]. For example, fusing 250-sample ERPs from 256 dense array channels will result in a data fusion vector which will have an enormously large dimension of 64,000. The obvious solution is to this problem is to select a subset of the dense array channels. The problems with this approach include determining: a suitable criterion for channel selection, how many channels should be selected, and how to fuse the information from the selected channels while satisfying the dimensionality requirement for multivariate classifier development. Furthermore, selecting a small set of channels defeats the main purpose for using dense arrays.

A possible approach to overcome the dimensionality problem is to design the classifier with a large ERP data set collected from a group of subjects. This approach is valid if the focus is on classifying the ERPs of individuals belonging to different population groups. However, in many clinical diagnostic and assessment applications, it is more meaningful to analyze the ERPs of a single patient. For example, ERPs are often used in the cognitive rehabilitation of patients with brain injuries [22]-[25]. In many cases, it may not be possible to collect ERP data from large groups of patients with the similar injuries. Moreover, it cannot be expected that the ERPs of patients with different severities and types of brain injuries will have similar features. Similarly, comments can be made for the cognitive rehabilitation of patients with post-stroke aphasia. Many attention and learning-related problems are also unique to individuals and cannot be generalized across a large group of individuals. The application, therefore, of commonly used ERP techniques will be more effective if only the ERPs of each individual, rather than of group of individuals, are used in the assessment. From a classification point of view, pooling ERPs of a large group of subjects will increase the risk of over-generalization which has the well-known effect of increasing the probability of classification error [20].

Our approach to overcoming the dimensionality problem is to develop a generalized dimensionality reduction strategy to extract only useful discriminatory information from the dense-array channels. The strategy formulated in this paper exploits the fact that different channels reflect different aspects of the brain activity and although the amount of data generated in dense array channels is large, only a small fraction of the data contains useful discriminatory information. That is, all channels do not necessarily carry useful information at all sampling instants. It may also be possible that some channels do not carry any useful information at any time. Similarly, at some sampling instants, none of the channels may carry useful information. The key, therefore, is to simultaneously exploit the spatially enhanced information offered by the dense array channels and to decrease the dimension by excluding channel-time combinations that do not contribute any useful classification information. In order to accomplish this, a spatio-temporal array model is introduced to represent the dense array brain activity simultaneously across channels and across time. With the help of this model, a two-step process which uses the

Kolmogrov-Smirnov test and the Lilliefors test is formulated to determine those spatio-temporal elements in the arrays which have different Gaussian densities across all of the brain activity categories. What is unique about this approach is that a model is assumed for the spatio-temporal elements and only those elements that fit the model and are statistically different across the categories are selected for classification. This is unlike the usual approach in which a model is assumed for the data and then applied to the data whether it fits or does not fit the data completely or partially. For example, due to its analytical tractability, the feature vector is often assumed to have a multivariate Gaussian density in numerous pattern classification problems.

It is noted that selecting the elements of the spatio-temporal array can be viewed as dynamic channel selection because selecting spatio-temporal array elements correspond to selecting the samples of different channels at different time-instants. The selected elements are classified using univariate Gaussian classifiers and the resulting decisions of the univariate classifiers are fused into a decision fusion vector which is classified by a discrete Bayes classifier. The multivariate classification problem is, therefore, converted into a simpler problem involving only univariate classifications and the dimensionality problem is circumvented by having to estimate univariate parameters instead of multivariate parameters. A normalization step is also included to decrease the trial-to-trial inter-class variability due to amplitude and slope variations in single-trial ERPs. In order to demonstrate the application of the dynamic channel selection strategy, classifiers are designed and tested for 8 subjects using 256-channel ERPs from a Stroop-color test. Results are presented for classifiers designed for each individual subject and for classifiers designed using the ERPs of the 8 subjects pooled into a single group.

II. MULTI-CHANNEL ERP CLASSIFICATION

ERPs are difficult to classify because they are embedded in the ongoing background EEG with signal-to-noise ratios (SNRs) typically less than –5 dB. Moreover, components of ERPs experience trial-to-trial variations due to latency shifts. These inter-trial variations increase the intra-class variability and also the overlap between the ERP classes. As a result of the increased overlapping, the probability of misclassification increases. Additional difficulties are introduced by artifact contaminations in ERP ensembles which are typically due to eye blinks, eye movements, muscle activities, and power line noise [26]. Analysis and classification are generally conducted on ERPs averaged over a large number of single-trials because signal averaging improves the signal-to-noise ratios (SNRs) of ERPs. The improvement in classification accuracy through signal averaging has been shown systematically in [27].

Several techniques based on fusing a small set of discriminatory features into a feature vector have been developed to classify ERPs. Examples of features derived from the ERPs of individual channels, among many others, include autoregressive (AR) model parameters [28],[29] peak and latency measurements [30],[31], PCA [15],[19],[32] and wavelets [15],[18],[33]. The features are classified using

discriminant analysis [19],[30],[34] neural networks [15], [28],[34],[35], decision trees [9],[35], and support vector machines [34]. Fusing a small set of discriminatory features will certainly decrease the dimensionality problem, however, there is a loss of information during feature generation [36] and the classification accuracy for a given problem is dictated by the choice and discriminatory quality of the features. Furthermore, fusing the features from all dense array channels into a single feature vector will result in a feature vector with a large dimension even if the feature set extracted from each ERP channel is small. For example, fusing 10 AR model parameters of each channel of a 256-channel dense array will result in a feature vector having a dimension equal to 2,560. Our previous studies focused on formulating strategies that were aimed at exploiting the complementary information from multiple channels through information (data and decision) fusion techniques [21],[27]. The goal of these multi-channel information fusion strategies was to obtain results that were significantly better than the results that are possible using any one channel. The data and decision fusion classifiers were, however, multivariate classifiers which required the total number of single-trial ERPs in the training set to be larger than the dimension of the ERPs. The strategy developed in the following sections of this paper is aimed at overcoming this dimensionality limitation.

III. SPATIO-TEMPORAL MODELING

Dense array or multi-channel brain activity can be analyzed by observing the amplitude variations of each channel over time or by observing the amplitude variations across the channels at different time- instants. Figure 1 is introduced to help visualize the amplitude variations across channels and time simultaneously. The horizontal axis k, $k = 1,2,...,K$, represents discrete time and the vertical axis m, $m = 1,2,...,M$ represents the channels, where K and M are the number of sampling instants and the number of channels, respectively. Each row, therefore, is an ERP of dimension K. Each one of the M columns can be used to generate a brain activity map across the channels at a given sampling time-instant. The resulting array of single-trial brain activity, across all channels and all time, can be represented by an array $Z(m,k)$ in which the single-trial ERP of channel m can be represented as $Z_m(k)$ and the brain activity at time k across all M channels by $Z_k(m)$. The array $Z(m,k)$ will be referred to as a single-trial spatio-temporal ERP array. The dimension of $Z(m,k)$ is $M \times K$ which will be typically quite large for dense arrays, for example, the array will have dimension 256×250 if the number of channels and time-instants are 256 and 250, respectively.

We first introduce a strategy to convert the multivariate classification problem into a set of univariate classification problem by modeling each element $z(m,k)$ of the single-trial spatio-temporal array in the following manner:

$$z(m,k) = s(m,k) + l(m,k) + e(m,k) \qquad (1)$$

where $s(m,k)$ is the signal component and $l(m,k)$ and $e(m,k)$ are the amplitude components at (m,k) due to

3

latency shifts and background EEG noise, respectively. It is assumed that $s(m,k)$, $l(m,k)$ and $e(m,k)$ are uncorrelated. This univariate ERP model is similar to the widely used ERP vector model which assumes that a brain-response to an external stimulus is the sum of the ERP and EEG components with the 2 components being uncorrelated. It is also well-known that the ERP components experience latency shifts which are random variations in the positions of the peaks and valleys. The changes in the amplitudes of each spatio-temporal element are, therefore, due to the EEG as well as the latency. The univariate model in Equation (1), which describes the brain activity of a channel m at a given time-instant k, will be used to formulate the dynamic channel selection classification strategy described in the next section.

IV. DYNAMIC CHANNEL SELECTION CLASSIFICATION STRATEGY

The goal of the dynamic channel selection based classification strategy is to determine and select only those elements in the spatio-temporal arrays which carry the most useful discriminatory information for classification across the different brain activity classes. The complete strategy which consists of dynamic channel selection, univariate classifier design, and decision fusion, is described next.

Dynamic Channel Selection

Each spatio-temporal element $z(m,k)$ is a random variable which can be characterized by its probability density function (PDF); therefore, in order to facilitate correct classification, we propose to select those elements that have different probability density functions across <u>all</u> C classes. The Kolmogrov-Smirnov (K-S) test can be used to determine whether two data sets are from the same probability distribution [37]. In order to test across all C classes, the multi-class test K-S test is formulated as:

$$\hat{F}(m,k) =$$
$$\begin{cases} 1 & if \quad p_j(m,k) \neq p_i(m,k), all \ j \neq i; \ i,j = 1,2,...,C \\ 0 & otherwise \end{cases} \quad (2)$$

where $p_i(m,k)$ and $p_j(m,k)$ are the PDFs of $z(m,k)$ under classes i and j, respectively. That is, the K-S test is applied in a pairwise fashion with the total number of pairwise tests equal to $C(C-1)/2$. The array $\hat{F}(m,k)$ is a "K-S mask" which selects the statistically different elements of $Z(m,k)$. That is, if

$$\hat{Z}_c(m,k) = Z_c(m,k)\hat{F}(m,k); m = 1,2,...,M; k = 1,2,...,K, \quad (3)$$

where $Z_c(m,k)$ spatio-temporal ERP array of brain activity class $c = 1,2,...,C$, then, $\hat{Z}_c(m,k)$ will have non-zero elements that are statistically different across all C classes. Note that the product of the two arrays in Equation (3) is the array product and not the matrix product. Having identified the elements with different PDFs, the next task is to assume a model for the PDF of those elements.

In Equation (1), we will assume that $l(m,k)$ and

$e(m,k)$ are zero-mean Gaussian random variables with variances $\sigma_l^2(m,k)$ and $\sigma_e^2(m,k)$, respectively, so that

$$n(m,k) = [l(m,k) + e(m,k)] \quad (4)$$

is a zero-mean Gaussian random variable with variance $\sigma_n^2(m,k) = [\sigma_l^2(m,k) + \sigma_e^2(m,k)]$. From the assumptions, it follows that

$$z(m,k) = s(m,k) + n(m,k) \quad (5)$$

is a Gaussian random variable with mean $s(m,k)$ and variance $\sigma_n^2(m,k)$. The validity of the Gaussian assumption for each non-zero element of $\hat{Z}_c(m,k)$, across all C classes, can be evaluated by using the Lilliefors test for normality [37], as follows:

$$\tilde{F}(m,k) = 1 \ only \ if \ \hat{F}(m,k) = 1 \quad AND$$
$$p_i(m,k) \sim N(\bullet,\bullet); \quad all \ i; i = 1,2,...,C \quad (6)$$

where $p_i(m,k) \sim N(\bullet,\bullet)$ is the notation used to denote that $p_i(m,k)$ is Gaussian. Then, each non-zero element in the array

$$\tilde{Z}_c(m,k) = Z_c(m,k)\tilde{F}(m,k); m = 1,2,...,M; k = 1,2,...,K \quad (7)$$

will be Gaussian and will be statistically different across the C classes. That is, the array $\tilde{F}(m,k)$ is, therefore, a "mask" which identifies the spatio-temporal elements that have different Gaussian PDFs across the classes. Figure 2 is an alternate form of viewing the selection of the elements from the original array $Z(m,k)$. This simplified figure shows that channels E_7 is selected at time 1, channels E_2, and E_3, are selected at time 2, channels E_{14}, E_{11}, and E_5 are selected at time t_j, E_8 is selected at time t_{j+1}, E_4 is selected at time K etc. The selection, therefore, is dynamic in the sense that the samples of different channels are selected at different time-instants according to the non-zero elements in mask $\tilde{F}(m,k)$.

Univariate Classifier Design

Single-trial ERPs may be classified directly, see [38] for notable work in single-trial classification, or may be classified by first averaging r single-trial ERPs in order to improve the SNR [21],[27]. Increasing r improves the SNR of the averaged ERPs and the performance can, therefore, be expressed as a function of the averaging parameter r. ERPs averaged over r trials are referred to r-ERPs and single-trial ERPs are referred to as 1-ERPs. Using the notations introduced earlier in this section, dense array ERPs averaged over r trials can, therefore, be represented by an array $Z^r(m,k)$ in which the mth row ERP averaged over r trials is $Z_m^r(k)$. The array $Z^r(m,k)$ will be referred to as a r-averaged ERP spatio-temporal array. Similarly, $\tilde{Z}^r(m,k)$ is the r-averaged spatio-temporal ERP array consisting of the dynamically selected averaged elements. A univariate Gaussian classifier can be developed to independently classify

the dynamically selected (non-zero) elements $\tilde{z}^{r}(m,k)$ of a test array $\tilde{Z}^{r}(m,k)$. In general, the univariate Gaussian discriminant function for class c is given by

$$g_c[\tilde{z}^{r}(m,k)] = -\ln \tilde{\sigma}_c^{r}(m,k)$$
$$-[\tilde{z}^{r}(m,k) - \tilde{\mu}_c^{r}(m,k)]^2 / 2[\tilde{\sigma}_c^{r}(m,k)]^2 + \ln P_c \quad , (8)$$

where $\tilde{\mu}_c^{r}(m,k)$ and $[\tilde{\sigma}_c^{r}(m,k)]^2$ are the mean and variance of $\tilde{z}^{r}(m,k)$ under class c, respectively, and P_c is the *a priori* probability of class c. A test element $\tilde{z}^{r}(m,k)$ is assigned to the discriminant function that yields the highest value. That is, the test sample $\tilde{z}^{r}(m,k)$ is assigned to a category d given by

$$d = \arg \max_c \{g_c[\tilde{z}^{r}(m,k)]\} \quad (9)$$

Decision Fusion

The decisions of the univariate Gaussian classifiers can be combined into a decision fusion vector and the final decision of the brain activity class can then be made by classifying the decision fusion vector. Let L be the number of non-zero elements in $\tilde{Z}^{r}(m,k)$. Consequently, the number of univariate classifiers and the number of decisions will be L. Let

$$D_L = \overset{L}{\underset{j=1}{\nabla}} d_j, \quad (10)$$

where ∇ represents the concatenation operation and d_j, $j = 1,2,...,L$ is the decision of the jth univariate classifier. That is, $D_L = (d_1, d_2,...,d_L)^T$ is the decision fusion vector formed by concatenating the independent decisions, in any fixed order, of the L Gaussian univariate classifiers. The decision fusion vector D_L is a discrete random vector in which each element can take one of C values. Let the PDF of D_L under class c be $P(D_L / c)$. Then, it can be shown that the Bayes discriminant function for class c can be written as [21]

$$g_c(D_L) = \sum_{j=1}^{L} [\delta(d_j - 1)\ln(p_{j,1/c}) + \delta(d_j - 2)\ln(p_{j,2/c})$$
$$+ ... + \delta(d_j - C)\ln(p_{j,C/c})] + \ln P_c \quad (11)$$

where

$$\delta(x - a) = \begin{cases} 1 & if \quad x = a \\ 0 & if \quad x \neq a \end{cases} \quad (12)$$

and

$$p_{j,a/c} = P(d_j = a / c), \ j = 1,2,...,L \quad (13)$$

is the probability that $d_j = a$ when the true class is c. The final decision c_L^{*} resulting from decision fusion is given by

$$c_L^{*} = \arg \max_c [g_c(D_L)] \quad (14)$$

The training phase of the entire dynamic classification strategy is summarized in Figure 3 in which $\{Z_c(m,k)\}$ represents the single-trial training sets of all C classes. The training set $\{Z_c(m,k)\}$ is randomly partitioned into two mutually exclusive sets $\{Z_c^{A}(m,k)\}$ and $\{Z_c^{B}(m,k)\}$. The top-half of Figure 3 shows the channel selection using the 2 masks and the estimation of the parameters $\tilde{\mu}_c(m,k)$ and $[\tilde{\sigma}_c(m,k)]^2$ of the selected elements using the set $\{Z_c^{A}(m,k)\}$. These parameters are used to determine the discriminant function of the Gaussian classifier in the bottom-half of the figure. The set $\{Z_c^{B}(m,k)\}$ is used to estimate the Bayes classifier parameters $p_{j,a/c}$ from the classification accuracies of the Gaussian classifier. The steps involved in classifying an unknown r-averaged ERP array $Z^{r}(m,k)$ are summarized in Figure 4.

Parameter Estimation

From the model assumptions, each element of $\tilde{z}_c^{r}(m,k)$ of $\tilde{Z}_c^{r}(m,k)$ can be written as

$$\tilde{z}_c^{r}(m,k) = s_c(m,k) + \tilde{n}_c^{r}(m,k) \quad (15)$$

where $\tilde{n}_c^{r}(m,k)$ is the noise component averaged over r trials. Because $n_c(m,k)$ is assumed Gaussian, $\tilde{n}_c^{r}(m,k)$ is also Gaussian. If $[\tilde{\sigma}_c^{r}(m,k)]^2$ is the variance of $\tilde{n}_c^{r}(m,k)$, then, $\tilde{z}_c^{r}(m,k)$ is Gaussian with mean $s_c(m,k)$ and variance $[\tilde{\sigma}_c^{r}(m,k)]^2$. Note that $\tilde{\mu}_c^{r}(m,k) = \tilde{\mu}_c(m,k)$ and the estimates of the variances for any r can be determined from the single-trial variances using the following relationship:

$$[\tilde{\sigma}_c^{r}(m,k)]^2 = (1/r)[\tilde{\sigma}_c(m,k)]^2. \quad (16)$$

The set $\{Z_c^{A}(m,k)\}$ in Figure 3 is used to determine the dynamic selection masks using a significance level α for the K-S and Lilliefors tests. Increasing α decreases the acceptance region of the null hypotheses, therefore, the number of elements selected can be increased by increasing the significance levels of the tests. The second set $\{Z_c^{B}(m,k)\}$ is used to estimate Bayes discriminant function parameters $p_{j,a/b}$, $j = 1,2,...,L$, which can be estimated as the ratio of the number of r-EPs of class b classified as class a to the total number r-EPs belonging to class b in $\{Z_c^{B}(m,k)\}$

Note that although the entire formulation of the classification strategy is presented for ERPs averaged over r trials, the strategy is also applicable to classifying single-trial ERPs simply by setting r=1.

V. NORMALIZATION

As noted in Section II, the trial-to-trial variations in single-trial ERPs result in large intra-class variations which lead to a

5

decrease in the accuracy of classifying ERPs. A close examination of single-trial ERPs shows that single-trial ERPs typically experience both amplitude and slope variations from trial-to-trial. Therefore, we can expect an improvement in the classification accuracy if the ERPs are amplitude and slope normalized, at the single-trial level. The amplitudes of a single-trial ERP can be normalized by dividing each ERP sample by the standard deviation of all samples of the ERP. By subtracting the least-squares fit line, the single-trial ERPs are normalized to have zero slope. These two normalization operations not only decrease the intra-class scatter but also increase the inter-class separation which in general, tends to improve the classification accuracy. The spatio-temporal array formed by the normalized single-trial ERPs is represented by $\bar{Z}(m,k)$ and the normalized and averaged ERPs are represented by $\bar{Z}^r(m,k)$.

VI. ERP EXPERIMENT DATA

We emphasize that the focus of this paper is on ERP classifier development and not on a particular ERP study. The ERP data that were used to demonstrate the design and evaluation of the dynamic channel selection-based classification strategy were from an experiment in which children were engaged in a Stroop color-naming test [39]. ERP-based Stroop testing has become a standard measure in neuropsychological assessment [40]-[44]. The test measures cognitive processing and provides valuable diagnostic information on brain dysfunction, cognition, and psychopathology. In this particular set of attention related experiments, the subjects were presented names of color. The color of the actual ink was varied so that it was either congruent or incongruent with the word. For example, if the word RED was presented and it was in red ink, it was the congruent condition. If the word RED were in blue ink, it was the incongruent condition. Additionally, neutral stimuli consisting of a string of X's were presented in varying colors. The children were instructed to press one button for the congruent and neutral conditions and another for the incongruent condition. The responses were recorded at the onset of each stimulus presentation. The standard Stroop interference test measures the reaction times (RTs) for the different conditions. In general, when compared with the RT for the neutral condition, the RT decreases for the congruent condition and increases for the incongruent condition. ERPs provide a means to study the neural basis of the Stroop interference tasks. Our goal is to show that the ERPs elicited by the 3 conditions can be classified accurately for each subject.

ERP data were collected from a high-density array of 256 Ag/AgCl electrodes arranged into a net (Geodesic Sensor Net, EGI Inc.) from eight 8-year old children, 5 males and 3 females. The subjects are represented by $S_1, S_2,\ldots,S_8$. Brain wave data recording was controlled by the Net Station 2.0 software package (EGI, Inc.). During recording, all electrodes were referenced to Cz and then later re-referenced to an average reference for data analysis. All impedances remained at or under 40 kΩ as indicated by impedance measures made immediately before and after the test session. E-prime v.1.0

(SP2; PST, Inc) software packages controlled stimulus presentation. Single-trial ERPs were digitized over a 1.8 sec interval using a 250Hz sampling rate. Each 450 sample ERP epoch included a 100 ms pre-stimulus baseline and 700 ms post-stimulus interval. After removing artifacts using the Net Station artifact rejection routines, for each subject, a total of forty eight 256-channel single-trial ERPs were obtained for each of the 3 conditions (congruent, incongruent, and neutral). The 25 pre-stimulus and 175 post-stimulus samples were removed so that the single-trial ERPs initially consisted of 250 samples. As a first step towards dimensionality reduction, the 250-sample ERPs were down-sampled to 125 by dropping every other sample. Our initial exploratory experiments showed that there was no loss in classification accuracy when the ERPs were down-sampled to 125, however, any further reduction in the dimension resulted in a drop in the classification accuracy.

VII. CLASSIFICATION EXPERIMENTS

The first set (SET 1) involved classifying the 32,000 elements of the original spatio-temporal array $Z^r(m,k)$ directly. That is, the 32,000 spatio-temporal elements were assumed to be Gaussian and the 32,000 univariate Gaussian classifier decisions were fused into a decision fusion vector which was classified using a discrete Bayes classifier for the final decision. The Gaussian assumption was also tested by applying the Lilliefors test for normality to each spatio-temporal element across the $C=3$ classes. For each subject, the test showed that on the average, more than 93% of the elements fit a Gaussian model. The results from this set of experiments will serve to show the improvement of our dynamic channel selection strategy over simply selecting all the elements even though the Gaussian assumption is valid for a large majority of the elements. The second set (SET 2) was exactly the same as SET1 except that the spatio-temporal arrays were the normalized arrays $\bar{Z}^r(m,k)$. In the third set (SET 3) of experiments, only the dynamically selected elements of the normalized spatio-temporal arrays $\tilde{Z}^r(m,k)$ were classified and the resulting univariate Gaussian decisions were fused and classified using a Bayes classifier. Experiments were designed to classify the ERPs of the 3 stimuli conditions (**C/N/I**) as well as combinations of the 2 stimuli conditions (**C/N, C/I, N/I**), where, **C**, **N**, and **I** represent congruent, neutral, and incongruent, respectively. Furthermore, for each set of experiments, classifiers were designed for each subject independently and for the subjects pooled into a single group. It is important to note that for the group case, the classifiers were trained using the pooled training sets, however, the testing was conducted using the test set of each subject.

Performance Evaluation

Unlike estimating multivariate classifier parameters such as the covariance matrices, the univariate parameters (means and variances) required to design the univariate Gaussian classifiers can be estimated from a relatively small data set without being faced with the dimensionality problem. For

example, 48 single-trial spatio-temporal arrays were collected from each subject for each of the 3 stimulus conditions. Therefore, the number of samples of each spatio-temporal element was 48 for each subject. A fairly good estimate of the mean and variance of each spatio-temporal element could be obtained even from one-half of the 48 samples for designing the univariate classifiers. The remaining half could be used to estimate the performance. We would, however, like to get a more robust estimate of the performance over a much larger set. Instead of collecting a large ERP set which we know is quite difficult, a re-sampling approach is used to generate a large number of averaged ERP arrays for training and testing [21].

The resampling technique is similar to bootstrapping which is often used to evaluate the performance of classifiers from a finite set of patterns [20]. For example, given a total of 48 signals, a training set consisting of 24 signals can be formed by randomly selecting 24 signals out of the 48 signals. The performance can be estimated by averaging the results of the 24 signals in the remaining set which is used for testing. The random sampling process to generate a training set and the estimation of the performance from the remaining test set, can be repeated a large number of times. A final estimate of the performance is then obtained by averaging the test results across all the repetitions. The only difference in our approach is that we use resampling to form a large number of averaged ERPs. In a previous study [27], we investigated the use of resampling for generating averaged ERPs and showed that the total number of distinct ERPs averaged across r single-trials that can be formed from a total of N single-trial ERPs is given by $\{N!/[r!(N-r)!]\}$. For example, for ERPs averaged over $r=8$ trials, the total number of distinct 8-averaged ERP arrays that could be formed from $N=24$ ERP arrays is 735,471 in the training and the test sets of each stimulus condition.

In the experiments to follow, 200 training and 200 test set arrays were generated for each classification experiment. In order to be sure that the 200 arrays were distinct in each set, the sum-of-the-squared distances between the elements of each newly generated averaged ERP array and elements of every previously generated averaged ERP array were determined. If the sum-of-the-squared distance between the newly generated array and any other array was zero, the new array was dropped and the process was repeated until we had 200 averaged ERP arrays that were distinct from each other. For each stimulus condition, the 256-channel single-trial ERP arrays were randomly partitioned into 2 equi-sized sets to form the training set and the test set. Consequently, each subject had 24 single-trial spatio-temporal arrays in the training set and 24 single-trial spatio-temporal arrays in the test set for each stimulus condition. The dimension of each array was 256×125. In each set, the single-trial arrays were randomly selected and averaged to form the r-ERP training and test set arrays.

3-Category Classification Results
The classification accuracies, expressed as a percentage, of the test congruent, neutral and incongruent conditions using the *original* spatio-temporal arrays (SET 1), the *normalized* spatio-temporal arrays (SET 2), and the *dynamically selected*

spatio-temporal elements (SET 3) for each *individual* subject are presented in Table I. The ERPs were averaged over $r=8$ single-trials. The dynamic channel selection results are presented for significance levels $\alpha = 0.05$, 0.1, 0.15, and 0.2. The number enclosed in parentheses is the parameter L which is the number of elements, out of the total of 32,000 elements, that were dynamically selected for classification. For the combined case, S_{1-8}, the congruent, incongruent, and neutral training sets of the 8 subjects were pooled (mixed) into single congruent, incongruent, and neutral training sets. The results of testing the ERPs of each individual subject are shown in Table II. The best result for each subject (each row) is shown in boldface. The classification accuracies presented in the two tables are quite extensive because each single result is averaged over 200 trials.

Also included in the tables are the corresponding results for Gaussian multivariate classifiers using features derived from principal components analysis (PCA). This is a classic example of making the Gaussian assumption in multivariate classifier design. The principal components were computed by pooling the zero-mean congruent, incongruent, and neutral normalized training sets. For the single subject cases, the training set used to compute the principal components, therefore, consisted of 24x3=72 single-trial arrays. The rows of each array were concatenated to form a vector with dimension 32000. The resulting covariance matrix would have a dimension of 32000×32000 and the number of eigenvectors would be 32000. Given that the training set had 72 vectors, only 72 of the 32000 eigenvectors can be determined. Instead of trying to determine the eigenvectors of the very large covariance matrix, the eigenface method [45] was used to determine the 72 non-trivial eigenvectors. The resulting multivariate classifier had a dimension equal to 72. Similarly, by combining the grouped training ERP arrays of all 8 subjects, 72x8=576-dimensional Gaussian multivariate PCA classifiers were developed for the grouped subject cases.

Comparisons of Results
It is clear from the results in the tables that there is a significant improvement in performance by simply amplitude and slope normalizing the ERPs of the spatio-temporal arrays in order to decrease the intra-class scatter. Furthermore, it is also clear that the best results are obtained using a combination of normalization and dynamic channel selection. It is also interesting to note that the number of elements selected dynamically (L) is a small fraction of the total number of array elements. The improvement in performance can be attributed to the selection of the spatio-temporal data that fit the classification model rather than simply assuming the entire spatio-temporal data fit the model. The results also show the classification accuracies obtained using the dynamic channel selection strategy are significantly superior to those obtained using Gaussian multivariate PCA classifiers.

As demonstrated in the experiments, the dynamic channel selection classification strategy is applicable to classifying the ERPs of individual subjects as well as the ERPs of groups of subjects. The next comparison, therefore, is between the results for individual subjects and grouped

subjects because it was commented in the Introduction that combining ERPs of a large group of subjects will increase the risk of over-generalization which can have an adverse effect on the performance for each individual. Observe that the classification accuracies in Table II are lower than those in Table I. Furthermore, for each α, the average classification accuracies of the combined subjects in Table II is less than the worst of the 8 individual classification accuracies in Table I. The drop in classification accuracy is in spite of the fact that the size of the training set was essentially eight-times larger when the subjects were grouped. These results confirm the risk of pooling the training sets of different individuals into single groups. The drop could be even more significant if the ERPs of a larger number of subjects are pooled due to the larger increase in the intra-class scatter in the grouped ERPs. Another point worth noting is that although the group results are not as high as the results for individual subjects, the average classification accuracies across the eight subjects for each α in Table II are greater than twice that can be expected through random classification. This is because the expected classification accuracy for a three-class problem through random classification is only 33.3%.

2-Category Classification Results
In order to demonstrate another application, experiments were also designed to classify the 2-category cases: congruent and incongruent, incongruent and neutral, and congruent and neutral. The results are presented in Table III. Because dynamic channel selection gave the best results, only the results for the dynamically selected channels of the normalized ERP arrays are presented (similar to SET 3). As expected, much higher classifications are obtained for the 2-category cases when compared with the 3-category cases.

Illustration of Dynamic Channel Selection
As noted in Section IV, the selection of elements from the spatio-temporal arrays is dynamic in the sense that the samples of different channels are selected at different time-instants according to the non-zero elements in mask $\tilde{F}(m,k)$. Figure 5 shows an example of the ensemble averaged spatio-temporal arrays belonging to the 3 different conditions obtained from the normalized grouped ERPs of subject S_1 using $\alpha = 0.2$. A close examination of the arrays of the 3 classes shows that they are similar across a large majority of the elements. Conversely, only a small number of elements are different across the classes. Figure 6 shows the selection mask $\tilde{F}(m,k)$ which resulted from an application of the pairwise K-S tests and the Lilliefors tests to the the ERP data used in Figure 5. Observe the sparseness of the selection mask. Only

463 elements were selected out of the 32000 elements. Figure 7 shows the electrode layout for the 256-channel EGI geodesic sensor net used in the experiments. The channels that were selected according to the mask in Figure 6 are shaded in yellow. The number of channels selected was 188 out of the total of 256. Furthermore, the number of time-instants that was selected was 81 out of the total of 125.

These results validate our claims that the most useful discriminatory information is spread across the dense array channels and these channels do not carry useful discriminatory information at all times. The channels selected give us an insight as to which electrode sites are the best for separating the brain categories. The weights ($p_{j,a/b}$) of the selected spatio-temporal elements derived from the univariate classifiers give further insights into which channel-time-instant combinations carry the most useful discriminatory information. The weights are also an indication of relative importance of each selected spatio-temporal element in the determination of the final decision using the Bayes classifier.

VIII. CONCLUSIONS

In conclusion, the dynamic channel selection-based strategy developed in this paper provides a new method for accurately classifying dense array ERPs by exploiting the fact that different channels carry useful discriminatory information of the brain activity at different times. The most notable feature is that dimensionality problem is solved by selecting the spatio-temporal elements whose data fit the assumed classification model. The main advantage of the strategy is that it is not constrained by the dimensionality of the ERP vector and can, therefore, be used to classify ERPs without having to collect a prohibitively large number of ERPs. We demonstrated that classifiers can be designed and evaluated for individual subjects even though the number of ERPs in the training set was substantially smaller than the dimension of the dense array signal space. The significance of this contribution is that clinical assessments can be made more accurately for conditions that are unique to an individual using a practical number of his/her own ERPs. Furthermore, the strategy is quite simple in its formulation and implementation. The strategy also provides a generalized means for analyzing which dense array channels yield the most discriminatory information at different time-instants. Also noteworthy is that the strategy is quite general in its formulation. Therefore, it can be applied to select and classify the most useful discriminatory information, in a time-window, from different sensors in various multi-sensor signal classification problems. Although the sensors of all channels are identical for EEGs and ERPs, the sensors can be heterogeneous which an additional advantage of the strategy.
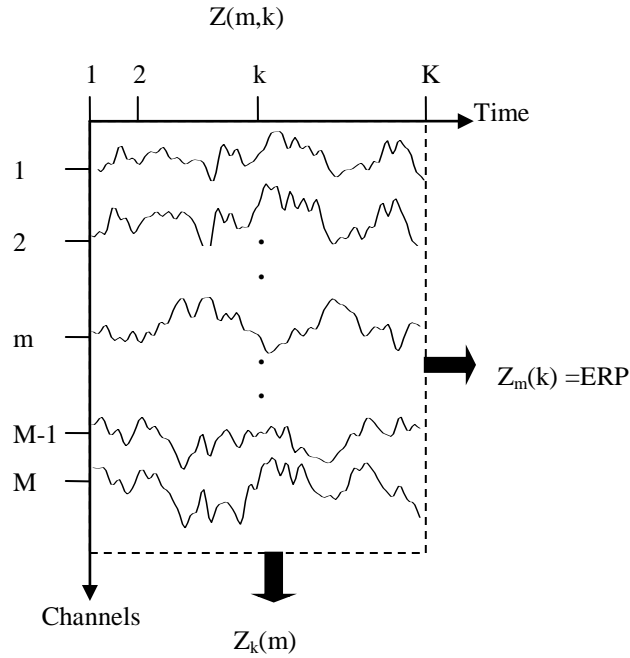
$Z(m,k)$

1   2   k   K

Time

1

2

$Z_m(k) =$ ERP

m

M-1

M

Channels

$Z_k(m)$

Fig. 1. Spatio-temporal array representing a
single-trial of dense array ERPs.

M channels
$E_1, E_2, \cdots, E_M$

E_{14}
E_2 E_{11}
E_7 E_3 E_5 E_8 E_4

1   2 t_j t_{j+1} K
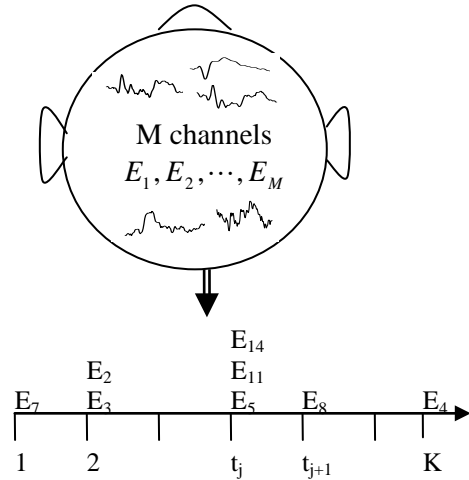
Fig. 2. Dynamic selection of
dense array channels.

$\{Z_c^A(m,k)\}$ → $\hat{F}(m,k)$ → $\tilde{F}(m,k)$ → Gaussian Parameter Estimation → $\tilde{\mu}_c(m,k), [\tilde{\sigma}_c(m,k)]^2$

$\{Z_c(m,k)\}$

$\{Z_c^B(m,k)\}$ → $\tilde{F}(m,k)$ → Averager → Gaussian Classifier → Bayes Parameter Estimation → $p_{j,a/c}$
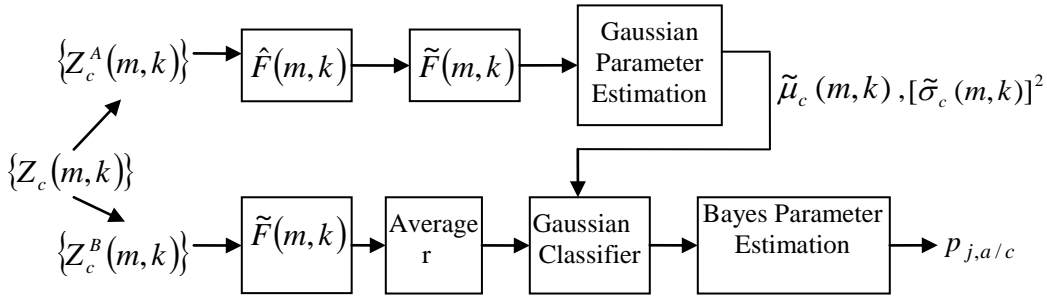
Fig. 3. Dynamic channel selection and parameter estimation during training

$Z^r(m,k)$ → $\tilde{F}(m,k)$ → $g_c[\tilde{z}^r(m,k)]$ → Decision Fusion → $g_c[D_L]$ → c_L^*
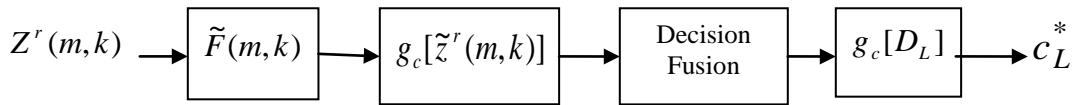
Fig. 4. Testing an averaged spatio-temporal ERP array

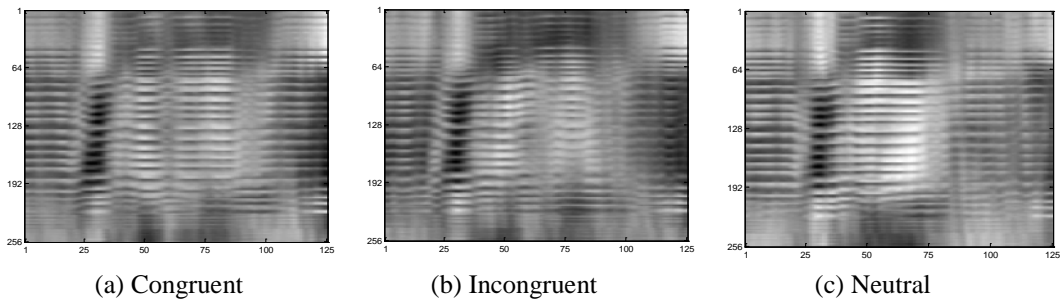(a) Congruent   (b) Incongruent   (c) Neutral
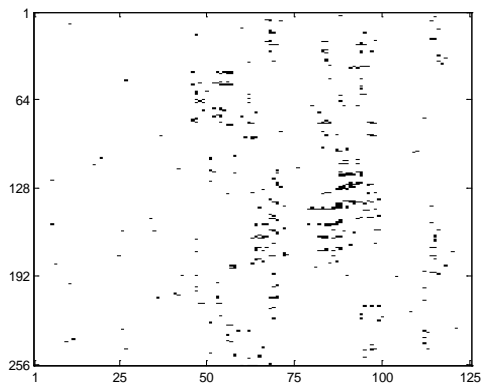
Fig. 5. Ensemble averaged spatio-temporal arrays

Fig. 6. The spatio-temporal element
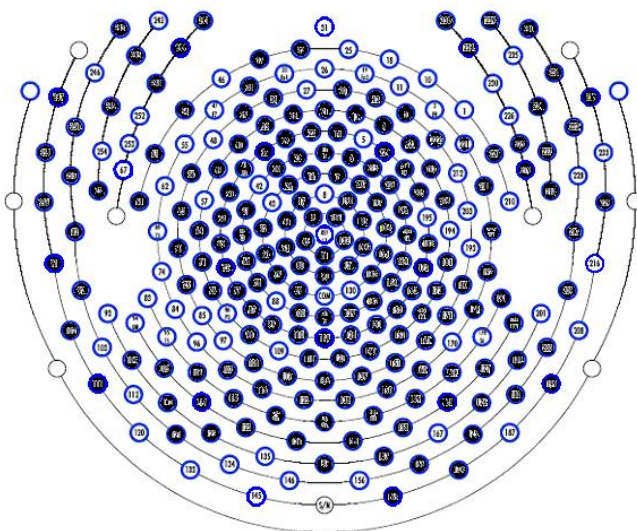
selection mask $\tilde{F}(m,k)$

Fig. 7. The electrode layout for the 256-channel EGI geodesic sensor. The channels that were selected according to the mask in Figure 6 are shaded in yellow.

Table I
Results of classifying the *original* spatio-temporal arrays, *normalized* spatio-temporal arrays, *dynamically selected* spatio-temporal elements of each individual subject. The PCA results are shown in the last column.

Subject	Original data (Set 1)	Normalized data (Set 2)	Dynamic channel Selection (Set 3)				PCA (Set2)
			α = 0.05	α = 0.10	α = 0.15	α = 0.20	
S₁	61.32	87.29	93.09 (82)	**94.52 (215)**	94.04 (357)	92.96 (463)	75.91
S₂	69.55	81.11	85.96 (51)	89.33 (147)	89.78 (261)	**90.67 (273)**	81.66
S₃	64.70	75.37	83.44 (41)	86.28 (127)	**89.36 (207)**	88.18 (307)	65.92
S₄	59.61	73.33	86.52 (64)	93.14 (150)	**93.36 (252)**	91.92 (268)	54.66
S₅	63.41	72.43	91.91 (87)	**94.58 (183)**	94.23 (325)	92.81 (446)	67.84
S₆	65.89	68.86	87.61 (58)	85.62 (105)	84.56 (237)	**88.77 (375)**	55.35
S₇	63.92	69.53	81.18 (73)	**86.34 (198)**	85.14 (370)	85.54 (561)	75.57
S₈	63.44	73.19	83.70 (55)	85.25 (152)	86.87 (244)	**87.25 (383)**	77.35
Average	63.98	75.14	86.68	89.38	89.67	**89.76**	69.28

10

Table II
Results of classifying the *original* spatio-temporal arrays, *normalized* spatio-temporal arrays, *dynamically selected* spatio-temporal elements of each individual subject using *pooled* training sets. The PCA results are shown in the last column.

Subject	Dynamic channel Selection (Set 3)				PCA (Set2)
	α = 0.05	α = 0.10	α = 0.15	α = 0.20	
S₁	75.58 (208)	82.47 (368)	**84.97 (516)**	82.53 (595)	61.33
S₂	63.58 (208)	67.96 (368)	69.58 (516)	**72.74 (595)**	64.49
S₃	63.86 (208)	**67.42 (368)**	60.85 (516)	59.37 (595)	60.39
S₄	72.49 (208)	76.43 (368)	75.60 (516)	**77.32 (595)**	70.42
S₅	79.02 (208)	**79.68 (368)**	79.08 (516)	80.58 (595)	63.81
S₆	72.19 (208)	73.10 (368)	73.09 (516)	**75.91 (595)**	66.91
S₇	52.34 (208)	**55.20 (368)**	52.80 (516)	54.01 (595)	41.69
S₈	**74.26 (208)**	73.63 (368)	64.52 (516)	61.96 (595)	43.22
Average	69.17	**71.99**	70.06	70.55	59.03

11

Table III
Summary of 2-category classification accuracies

Subject		Dynamic channel Selection			
		α = 0.05	α = 0.10	α = 0.15	α = 0.20
S₁	(C,I)	98.63 (2530)	98.51 (3403)	98.99 (3870)	**99.27 (4446**)
	(I,N)	**99.98 (5525)**	99.72 (6177)	99.96 (6317)	99.86 (6332)
	(N,C)	99.59 (6033)	**99.61 (5989)**	98.42 (6674)	99.16 (5807)
S₂	(C,I)	**96.70 (2359)**	95.13 (2884)	95.12 (3686)	95.28 (3084)
	(I,N)	99.85 (5990)	**100.0 (4566)**	**100.0 (6536)**	99.90 (5320)
	(N,C)	99.96 (4768)	**100.0 (5189)**	**100.0 (2718)**	99.37 (5707)
S₃	(C,I)	92.07 (1461)	96.17 (2461)	95.63 (2963)	**96.95 (3809)**
	(I,N)	99.07 (5507)	**99.65 (6180)**	99.30 (7503)	99.26 (6074)
	(N,C)	97.93 (3642)	98.67 (4765)	**99.20 (5291)**	98.17 (6109)
S₄	(C,I)	98.20 (1190)	**99.17 (2114)**	91.67 (3604)	96.17 (3276)
	(I,N)	99.83 (4049)	97.16 (5803)	99.62 (7208)	**100.0 (8116)**
	(N,C)	99.95 (3539)	**100.0 (5195)**	**100.0 (6401)**	93.33 (72.73)
S₅	(C,I)	99.89 (1261)	92.00 (2080)	**100.0 (2810)**	98.33 (3266)
	(I,N)	**100.0 (4616)**	**100.0 (5609)**	**100.0 (6251)**	**100.0 (6741)**
	(N,C)	**100.0 (5405)**	**100.0 (6295)**	100.0 (6889)	99.19 (6908)
S₆	(C,I)	96.24 (1520)	95.83 (2633)	**98.67 (3473)**	97.67 (3558)
	(I,N)	**100.0 (4098)**	**100.0 (4736)**	**100.0 (5720)**	**100 (6627)**
	(N,C)	**100.0 (4660)**	**100.0 (5599)**	**100.0 (6101)**	**100.0 (5741)**
S₇	(C,I)	95.50 (1472)	97.17 (2661)	**99.33 (3074)**	97.83 (3616)
	(I,N)	99.63 (4645)	98.10 (5976)	**100.0 (6864)**	**100.0 (7181)**
	(N,C)	97.68 (4705)	**99.67 (5758)**	98.06 (6446)	97.79 (6559)
S₈	(C,I)	99.03 (2198)	99.17 (3548)	97.67 (4051)	**99.35 (4872)**
	(I,N)	93.28 (5469)	**94.82 (7242)**	92.23 (8308)	94.14 (8828)
	(N,C)	**100.0 (6197)**	**100.0 (7436)**	**100.0 (8119)**	98.87 (8384)
Average	(C,I)	97.03	96.64	97.14	**97.61**
	(I,N)	98.76	98.68	98.89	**99.15**
	(N,C)	99.39	**99.74**	99.46	98.24

REFERENCES

[1] D.M. Tucker, "Spatial sampling of head electric fields: the geodesic sensor net," *Electroencephalography and Clinical Neurophysiology*, 87(3), pp. 154-163, 1993.

[2] H. Jasper, "The ten-twenty electrode system of the International Federation of Societies for Electroencephalography: Appendix to report of the committee on methods of clinical examination in electroencephalography," *Electroencephalography and Clinical Neurophysiology*, 10, pp. 371-375, 1958.

[3] T. Curran, "The electrophysiology of incidental and intentional retrieval: erp old/new efftets in lexical decision and recognition memory", *Neuropsychologia,* vol. 37, no. 7, pp. 771-785, 1999.

[4] M. Junghofer, M.M. Bradley, T.R. Elbert, & P.J. Lang, "Fleeting images: A new look at early emotion discrimination," *Psychophysiology*, vol. 38, issue. 2, pp. 175-178, 2001.

[5] A. Keil, M.M. Bradley, O. Hauk, B. Rockstroh, T. Elbert, & P.J. Lang, "Large-scale neural correlates of affective picture processing," *Psychophysiology*, vol. 39, issue. 5, pp. 641-649, 2002.

[6] G.M. Doniger, J.J. Foxe, C.E. Schroeder, M.M. Murray, B.A. Higgins, & D.C. Javitt, "Visual perceptual learning in human object recognition areas: a repetition priming study using high-density electrical mapping", *NeuroImage,* vol. 13, no. 2, pp. 305-313, 2001.

[7] J. Russeler, E. Henninghausen, T. F. Muntee, & F. Rosler, "Differences in incidental and intentional learning of sensorimotor sequences as revealed by event-related potentials," *Cognitive Brain Research*, vol. 15, issue. 2, pp. 116-126, 2003.

[8] A.A. Benasich, N. Choudhary, J.T. Friedman, T. Realpe-Bonilla, C. Chojnowska, & Z. Gou, "The infant as a prelinguistic model for language learning impairments: predicting from event-related potentials to behavior", *Neuropsychologia,* vol. 44, no. 3, pp. 396-411, 2006.

[9] G.A. Frishkoff, R.M. Frank, J. Rong, D. Dou, J. Dien, & L.K. Halderman, "A framework to support automated classification and labeling of brain electromagnetic patterns", *Computational Intelligence and Neuroscience,* vol. 2007, Article ID 14567, 13 pages.

[10] S. C. Sereno & K. Rayner, "Measuring word recognition in reading: eye movements and event-related potentials," Trends in Cognitive Sciences, vol. 7, no. 11, pp. 489-493, 2003.

[11] A.P.F. Key, M. Ferguson, D.L. Molfese, K. Peach, C. Lehman, & V.J. Molfese, "Smoking during pregnancy affects speech-processing ability," *Environmental Health Perspectives*, vol. 115, no. 4, pp. 623-629, 2007.

[12] Y. Liu, & C.A. Perfetti, "The time course of brain activity in reading english and chinese: an ERP study of chinese bilinguals", *Human Brain Mapping*, vol. 18, no. 3, pp. 167-175, 2003.

[13] C. Poulsen, P. Luu, C. Davey, D. Tucker, "Dynamics of task sets: Evidence from dense-array event-related potentials," *Cognitive Brain Research*, vol. 24, issue. 1, pp. 133-154, 2005.

[14] C.D. Ladouceur, R.E. Dahl, B. Birmaher, D.A. Axelson, & N.D. Ryan, "Increased error-related negativity (ERN) in childhood anxiety disorders: ERP and source localization," *Journal of Child Psychology and Psychiatry* 47:10, pp. 1073-1082, 2006.

[15] L.J. Trejo, & M.J. Shensa, "Feature extraction of event-related potentials

using wavelets: An application to human performance monitoring", *Brain and Language,* vol. 66, no. 1, pp. 89-107, 1999.

[16] K.M. Spencer, J. Dien, & E. Donchin, "Spatiotemporal analysis of the late ERP responses to deviant stimuli", *Psychophysiology*, vol. 38, issue. 2, pp. 343-358, 2001.

[17] J. Kayser, & C.E. Tenke, "Principal components analysis of laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks", *Clinical Neurophysiology*, vol. 177, no. 2, pp. 348-368, 2006.

[18] M. Mørup, L.K. Hanser, C.S. Herrmann, J. Parnas, & S.M. Arnfred, "Parallel Factor Analysis as an exploratory tool for wavelet transformed event-related EEG", *NeuroImage,* vol. 29, no. 3, pp. 938-947, 2006.

[19] K-A. Do & K. Kirk, "Discriminant analysis of event-related potential curves using smoothed principal components," *Biometrics*, vol. 55, issue. 1, pp. 174-181, 1999.

[20] R.O. Duda, P.E. Hart, & D.G. Stork, Pattern Classification, John Wiley & Sons, 2001.

[21] L. Gupta, B. Chung, M.D. Srinath, D.L. Molfese, & H. Kook "Multi-channel fusion models for the parametric classification of differential brain activity," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 11, pp. 1869-1881, 2005.

[22] A-K. Solbak, I. Reinvang, C. Nielsen, & K. Sundet, "ERP indicators of disturbed attention in mild closed head injury: A frontal lobe syndrome?," *Psychophysiology*, vol. 36, issue. 6, 802-817, 1999.

[23] I. Reinvang, H. Nordby, & C.S. Neilsen, "Information processing deficits in head injury assessed with ERPs reflecting early and late processing states," *Neuropsychologia*, vol. 38, no. 7, pp. 995-1005, 2000.

[24] H.L. Lew, J.H. Poole, A. Castaneda, R. Marie, & M. Gray, "Prognostic value of evoked and event-related potentials in moderate to severe brain injury," *The Journal of Head Trauma Rehabilitation*, vol. 21, issue. 4, pp. 350-360, 2006.

[25] R. Doi, K. Morita, M. Shigemori, T. Tokutomi, & H. Maeda, "Characteristics of cognitive function in patients after traumatic brain injury assessed by visual and auditory event-related potentials," *American Journal of Physical Medicine & Rehabilitation*, vol. 86, issue. 8, pp. 641-649, 2007.

[26] H. Kook, L. Gupta, S. Kota, D. Molfese , & H. Lyytinen, "An offline/real-time artifact rejection strategy to improve the classification of multi-channel evoked potentials," *Pattern Recognition* (in press).

[27] L. Gupta, J. Phegley, & D.L. Molfese, "Parametric classification of multichannel evoked potentials," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 8, pp. 905-911, Aug. 2002 (vol. 49, no. 9, 1070, September 2002).

[28] C.E. Vasios, G.K. Matsopoulos, K.S. Nikita, & N. Uzunoglu, "Classification of event-related potentials using multivariate autoregressive modeling combined with simulated annealing," *Journal of Automatic Control*, vol. 13(1), pp.7-11, 2003.

[29] D.P. Burke, S.P. Kelly, P. Chazal, R.B. Reilly, & C. Finucane, "A parametric feature extraction and classification strategy for brain-computer interfacing," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 1, 12-17, 2005.

[30] D.L. Molfese, "Predicting dyslexia at 8 years of age using neonatal brain responses," *Brain and Language*, vol. 72, issue. 3, pp. 238-245, 2000.

[31] A. Kiesel, J. Miller, P. Jolccoeur, & B. Brisson, "Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods," *Psychophysiology*, 44, 2007.

[32] C.W. Anderson, E.A. Stolz, and S. Shamsunder, "Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 3, pp. 277-286,1998.

[33] H. Heinrich, H. Dickhaus, A. Rothenberger, V. Heinrich, & G.H. Moll, "Single-sweep analysis of event-related potentials by wavelet networks-methodological basis and clinical application", *IEEE Transactions on Biomedical Engineering,* vol. 46, no. 7, pp. 867-879, 1999.

[34] D. Garrett, D.A. Peterson, C.W. Anderson, & M.H. Thaut, "Comparison of Linear, Nonlinear, and feature selection methods for EEG signal classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, No. 2, pp. 141-144, 2003.

[35] R. Davey, P. McCullagh, G. Lightbody, G. McAllister, "Auditory brainstem response classification: A hybrid model using time and frequency features", *Artificial Intelligence in Medicine*, vol. 40, issue 1, pp. 1-14, 2007.

[36] P.K. Varshney, "Multisensor data fusion," *Electronics & Communication Engineering Journal*, vol. 9, issue. 6, pp. 245-253, December 1997.

[37] W.J. Conover, "Practical Nonparametric Statistics, Third Edition, John Wiley & Sons, New York, 1999.

[38] S. Makeig, M. Westerfield, T-P Jung, S. Enghoff,, J. Townsend, E. Courchesne, T.J. Sejnowski, "Dynamic brain sources of visual evoked responses," Science, 295, 690-64, 2002.

[39] J.R. Stroop, "Studies of interference in serial verbal reactions," *Journal of Experimental Psychology*, vol. 18(6), pp. 643-662, 1935.

[40] R. West & C. Alain, "Age-related decline in inhibitory control contributes to the increased Stroop effect observed in older adults," *Psychophysiolgy*, vol. 37, issue. 2, pp. 179-189, 2000.

[41] T. Sitnikova, D. F. Salisbury, G. Kuperberg, & P. J. Holcomb, "Electrophysiological insights into language processing in schizophrenia," *Psychophysiolgy*, vol. 39, issue. 6, pp. 851-860, 2002.

[42] J. Kray, B. Eppinger, & A. Mecklinger, "Age differences in attention control: An even-related potential approach," *Psychophysiolgy*, vol. 42, issue. 4, pp. 407-416, 2005.

[43] K. Ito, N. Tomiya, M. Iimori, J. Yoshihama, & H. Matsuda, "Visual cognitive function in dementia of Alzheimer's type: a comparison of visual event-related potentials and Stroop test," *Psychogeriatrics*, vol. 6, issue. 2, pp. 49-54, 2006.

[44] M-S. Kim, S.S. Kang, K.S. Shin, S.Y. Yoo, Y.Y. Kim, & J.S. Kwon, "Neuropsychological correlates of error negativity and positivity in schizophrenia patients," *Psychiatry and Clinical Neurosciences*, vol. 60, issue. 3, pp. 303-311, 2006.

[45] M. Turk and A. Pentland, "Eigenfaces for recognition," Journal of Congnitive Neuroscience, vol. 3, No. 1, pp. 71-86, 1991.

Srinivas Kota (S'05) received the B.E. degree in electronics and communications engineering from Andhra University, Visakhapatnam, India, in 2000 and the M.S. degree in electrical engineering from Southern Illinois University, Carbondale, in 2004. He is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, Southern Illinois University.

His research interests are in signal processing, image processing, pattern recognition and neuroinformatics.

Lalit Gupta received the B.E. (Hons) degree in electrical engineering from the Birla Institute of Technology and Science, Pilani, India (1976), the M.S. degree in digital systems from Brunel University, Middlesex, England (1981), and the Ph.D degree in electrical engineering from Southern Methodist University, Dallas, Texas (1986).

He is currently a Professor of Electrical and Computer Engineering at Southern Illinois University, Carbondale, Illinois. His research interests are in pattern recognition, neuroinformatics, neural networks, and signal processing. He has been awarded contracts from the Army Research office to conduct research in the development of smart munitions, from Seagate Technology on image compression research, from Cleveland Medical Devices on NIH funded projects related to

13

brain wave form analysis and classification, from Think-A-Move Inc. on a NIH funded project related to Human-Machine Interfacing, from Neuronetrix on a NIH funded project on detecting neurological disorders from evoked potentials, and from the Naval Postgraduate School on research related to Human-Machine Interfacing.

Dr. Gupta has numerous publications in the areas of neural networks, evoked potential analyses and classification, and multi-channel/sensor information fusion strategies. Dr. Gupta serves as an Associate Editor of the Pattern Recognition Journal and is a Senior Member of the IEEE.

D. L. Molfese. B.A., Psychology, The Oklahoma City University, Oklahoma City , OK, 1969, M.S., Psychology, The Pennsylvania State University, State College, PA, 1970, Ph.D., Psychology, The Pennsylvania State University, State College, PA, 1972. Major field of study: developmental neuroscience.

He is currently Distinguished University Scholar and Professor in the Birth Defects Center at the University of Louisville, Louisville, KY. Previously, he served as Chair of the Department of Psychological and Brian Sciences at the University of Louisville (1999-2006), Chair and Professor of Behavioral & Social Sciences in the School of Medicine at Southern Illinois University (1992-1999), and Professor of Psychology at Southern Illinois University (1972-1999). He is the author of 150+ journal articles and books. His current research interests include predicting cognitive and linguistic skills from infancy; developmental changes in brain, language and cognitive processes from infancy; impact of space travel on cognition, brain function and sleep; development of intervention strategies for learning disabilities in infancy, man - machine interactions using neuroelectrophysiological techniques. Sample publications include: Molfese, D., Molfese, V., Barnes, M., Warren, C., & Molfese, P. (2008). Familial predictors of dyslexia: Evidence from preschool children with and without familial dyslexia risk. In V. Berninger, A. Fawcett, G. Reid & L Siegel (Eds). *Handbook of Dyslexia*. New York Sage Publications, 99-120; Tan, A. & Molfese, D. L. (2008). ERP Correlates of Noun and Verb Processing in Preschool-Age Children. *Biological Psychology*.

Dr. Molfese is a Fellow of the American Psychological Association and the American Psychological Society, as well as an elected member of The Belgian Society of Electromyography and Clinical Neurophysiology, Phi Kappa Phi, Psychonomic Society, Sigma Xi. He is also a member of the American Academy of Sleep Medicine, the Society for Neuroscience, the Acoustical Society of America, the International Neuropsychological Society, and the Society for Research in Child Development. He has been recognized as Psychologist of the Year in 2006 by the Kentucky Psychological Association, Collaborative Investigator to the National Institute of Child Health and Human Development in 2005, received the Dean's Award for 2004-2005 Outstanding Department Chair in the College of Arts & Sciences at University of Louisville, the Phi Kappa Phi Outstanding Scholar Award in 1997, and the University Outstanding Scholar Award in 1990 from Southern Illinois University. He also received the Sigma Xi Kaplan Research Award for Significant Contributions in Child Psychology in April, 1987.

Ravi Vaidyanathan received the M.S. and Ph.D. degrees in mechanical engineering from Case Western Reserve University, Cleveland, OH, USA in 1996 and 2001 respectively.

He is presently a Lecturer in Biodynamics at the Bristol Robotics Laboratory in the University of Bristol, Bristol, UK with a co-appointment as an Assistant Professor in Systems Engineering at the US Naval Postgraduate School, Monterey, CA, USA. His research interests include biologically inspired robotics, human-machine interface, and complex adaptive systems. He has served as Principal Investigator in more than 20 research separate programmes in the US, Singapore, and UK, has authored over 50 refereed publications, and two patents.

Dr. Vaidyanathan has been the recipient of several awards from the IEEE Robotics and Automation Society, the Robotics Society of Japan (RSJ), and the American Institute of Aeronautics and Astronautics (AIAA), including Best Paper in Conference (2004) at the IEEE International Conference on Intelligent Robots and Systems (IROS) and being a finalist for the New Technology Foundation (NTF) Award on Entertainment and Robotic Systems celebrating top innovations in robotics from 1987-2007.

14