Available online at www.sciencedirect.com
SCIENCE DIRECT

Pattern Recognition 38 (2005) 2174–2184

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Multi-stimuli multi-channel data and decision fusion strategies for dyslexia prediction using neonatal ERPs

Hyunseok Kook[a], Lalit Gupta[a,*], Dennis Molfese[b], K.C. Fadem[c]

[a]Department of Electrical & Computer Engineering, Southern Illinois University, Carbondale, IL 62901, USA
[b]Department of Psychological & Brain Sciences, University of Louisville, Louisville, KY 40292, USA
[c]Neuronetrix, Inc., Louisville, KY 40208, USA

Received 22 November 2004; received in revised form 28 March 2005; accepted 28 March 2005

Abstract

Data fusion and decision fusion classification strategies are introduced to predict dyslexia from multi-channel event related potentials (ERPs) recorded, at birth, in response to multiple stimuli. Two data and two decision fusion strategies are developed in conjunction with nearest-mean classification rank selection to classify multi-stimuli multi-channel (MSMC) ERPs. The fusion vector in the data fusion strategy is formed by directly combining the rank-ordered MSMC ERP vectors or the rank-ordered elements of the MSMC ERPS. The resulting fusion vector is classified using a vector nearest-mean classifier. The nearest-mean classification decisions of the rank-ordered MSMC ERP vectors or the rank-ordered MSMC ERP elements are combined into a fusion vector in the decision fusion strategy. The resulting decision fusion vector is classified using a discrete Bayes classifier. The MSMC fusion classification strategies are tested on the averaged ERPs recorded at birth of 48 children: 17 identified as dyslexic readers, 7 as poor readers, and 24 identified as normal readers at 8 years of age. The ERPs were recorded at 6 electrode sites in response to two speech sounds and two non-speech sounds. It is shown that through the MSMC ERP element decision fusion strategy, dyslexic readers and poor readers can be predicted with almost 100% accuracy. Consequently, future reading problems can be detected early using neonatal responses making it possible to introduce more effective interventions earlier to children with reading problems emerging later in their lives. Furthermore, it is noted that because of the generalized formulations, the fusion strategies introduced can be applied, in general, to problems involving the classification of multi-category multi-sensor signals.

Keywords: Bayes classification; Dyslexia; Event related potentials; Fusion; Nearest-mean classification

1. Introduction

The aim of this investigation is to show, through data and decision fusion classification strategies, that information from multi-stimuli event related potentials (ERPs) recorded at birth from multiple channels can predict dyslexia with very high accuracies. In a previous study on 48 children [1], it was shown that amplitude and latency measures of three auditory ERP components recorded at birth discriminated with 81.25% accuracy among three groups of children identified as normal, poor, or dyslexic readers based on reading and IQ scores obtained at 8 years of age. The significance of these results is that dyslexia can be detected early using neonatal ERPs. Consequently, early and more effective interventions can be provided to children before they enter school and thus improve their ability to learn. The ERPs were recorded from six scalp electrodes (FL, FR, T3,T4, PL,

* Corresponding author. Tel.: +1 6184537032; fax: +1 6184537972.
E-mail address: lgupta@siu.edu (L. Gupta).

0031-3203/$30.00
doi:10.1016/j.patcog.2005.03.003

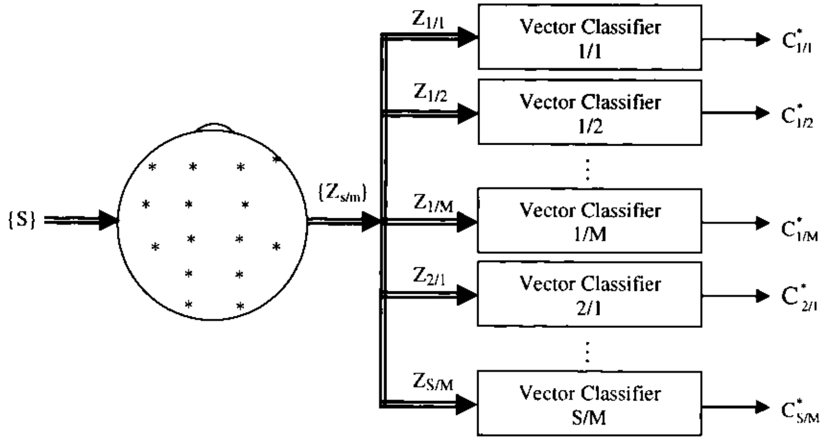

Fig. 1. Classification of each MSMC ERP type.

PR) to four stimuli consisting of two speech (/bi/ and /gi/) and two non-speech sounds (/nbi/ and /ngi/). Each child was represented by 24 (6 channels $\times$ 4 stimuli) averaged ERPs, that is, one averaged ERP for each stimulus–channel combination. Each ERP consisted of 200 data points, sampled at 5 ms intervals and collected sequentially over a 1 s period beginning at stimulus onset. The first 13 samples and the last 47 samples were dropped and the remaining 140 point signal was down-sampled by a factor of 2 to facilitate discriminant analysis. Each ERP, therefore, consisted of 70 samples. The six ERP features selected for discriminant analysis classification consisted of: (a) the first large negative peak (N1) latency to the speech syllable /gi/ recorded at FL; (b) the first large negative peak (N1) latency to the speech syllable /gi/ recorded at PL; (c) the first large negative peak (N1) latency to the speech syllable /gi/ recorded at T4; (d) the second large negative peak (N2) amplitude that differed between the two groups in response to the /gi/ speech syllable at FR; (e) the (N1) amplitude change recorded at T4, in response to /nbi/; and (f) the second large positive peak amplitude (P2) elicited in response to /bi/ at PR. Details regarding the subjects, stimuli, ERP procedures, and results of the study can be found in Ref. [2].

Given the complexity of the problem, the classification results reported in the previous study are quite impressive. However, even higher classification accuracies with lower false positives are desired in practice. The specific goal in this paper, therefore, is to improve the classification performance using exactly the same averaged data used in the previous study. In this investigation, the 48 children were first divided into two categories consisting of 24 normal readers and 24 poor readers (dyslexics and poor readers combined) and next, into three categories (24 normal readers, 17 dyslexic readers, and 7 poor readers). Given the small number of ERPs with respect to the dimension of the ERP vector, it is clearly not possible to design parametric classifiers that depend on second-order statistics. This limits the choice of parametric classifiers to those that depend only on the first-order statistics. Given this limitation, the nearest-mean parametric classifier is selected because the mean representing each category can be estimated from the training sets using either the "leave-one-out method" or the "random equi-partition training and testing method [1]." Two multi-stimuli multi-channel (MSMC) data fusion and two decision fusion strategies are introduced in conjunction with nearest-mean classifiers. In the both fusion strategies, the MSMC ERP vectors or the MSMC ERP elements are first ranked according to their individual Euclidean nearest-mean classification accuracies. In the data fusion strategies, the vectors or elements are selected according to their rank and fused into a data fusion vector. The resulting data fusion vector is classified using a vector Euclidean nearest-mean classifier. In the decision fusion strategies, the vectors or elements in the data fusion vector are classified independently using the Euclidean nearest-mean classifier and the decisions are fused into a single decision fusion vector. The decision fusion vector is classified using a discrete Bayes classifier. The formulations of the strategies are quite general and are not limited by the number of ERP categories, channels, or stimuli. The strategies are tested on the same data used in Ref. [2] and the results are presented in terms of the classification decision probabilities and the corresponding posterior probabilities. The performance is compared with the results reported in Ref. [2] and also with the nearest mean classification results of the single best stimulus and channel combination. In studies involving multiple channels and multiple stimuli, it is important to determine which channels, which stimuli, and which channel–stimuli give the best discriminatory information, therefore, a ranking strategy derived from the nearest-mean classification accuracies of the MSMC ERPs is first introduced. Ranking is also used to rank and select MSMC ERP vectors and MSMC ERP elements in the four fusion strategies.

2. Channel–stimulus, stimulus, and channel ranking

In the generalized multi-category formulations to follow, $Z_{s/m}$ represents the K-dimensional averaged ERP elicited by external stimulus s, $s = 1, 2, \ldots, S$, at channel m, $m = 1, 2, \ldots, M$, where S and M are the number of stimuli and channels, respectively. An averaged ERP elicited by external stimulus s at channel m given the category is c, $c = 1, 2, \ldots, C$, is represented by $Z_{s/m/c}$ where C is the total number of categories. The number of different types of MSMC ERPs in each category is, therefore $(S \times M)$.

The nearest-mean classifier is developed for each ERP type as shown in Fig. 1. The notation $\{\}$ in the figure is used to represent the entire collection. Therefore, $\{s\}$ is the entire set of S stimuli and $\{Z_{s/m}\}$ is the entire set of $(S \times M)$ MSMC ERPs. The Euclidean norm nearest-mean vector discriminant function for category c, is given by

$$g_{c;s/m}(Z_{s/m}) = (Z^{\mathrm{T}}_{s/m})(\bar{Z}_{s/m/c}) - (1/2)(\bar{Z}^{\mathrm{T}}_{s/m/c})(\bar{Z}_{s/m/c}), \quad (1)$$

where $\bar{Z}_{s/m/c}$ is the mean of the c-category ERPs of channel m in response to stimulus s. A test ERP $Z_{s/m}$ is assigned to the discriminant function that yields the highest value. That is, the test MSMC ERP $Z_{s/m}$ is assigned to the category $c^*_{s/m}$, $c^*_{s/m} \in \{c\}$, given by

$$c^*_{s/m} = \arg\max_{c}[g_{c;s/m}(Z_{s/m})]. \quad (2)$$

The classification accuracy of the classifier for each ERP type can be determined from the probability of classification error which is given by

$$P_{s/m}(\varepsilon) = \sum_{c=1}^{C} P_{s/m}(\varepsilon/c)P_c, \quad (3)$$

where $P_{s/m}(\varepsilon/c)$ is the probability of misclassifying $Z_{s/m}$ when the true category of $Z_{s/m}$ is c and P_c is the a priori probability of category c. The classification accuracy of the classifier for $Z_{s/m}$, expressed as a percentage, is given by

$$\alpha_{s/m} = [1 - P_{s/m}(\varepsilon)] \times 100\%. \quad (4)$$

The $(S \times M)$ MSMC ERP classifiers can be ranked according to their respective classification accuracies [3]. Furthermore, the stimuli and channels can be ranked individually using the ranks of the marginal rank-sums. The marginal rank-sums $\tilde{R}_s$ and $\tilde{Q}_m$ of the stimuli and of the channels are given, respectively, by

$$\tilde{R}_s = \sum_{m=1}^{M} R_{s/m}, \quad (5)$$

$$\tilde{Q}_m = \sum_{s=1}^{S} R_{s/m}, \quad (6)$$

where $R_{s/m}$ is the rank (1 = highest classification accuracy, $S \times M$ = lowest classification accuracy) of the classifier for ERP $Z_{s/m}$. The ranks R_s and R_m of the stimuli and channel are given by the ranks of the rank-sums, $\tilde{R}_s$ and $\tilde{Q}_m$, respectively.

3. MSMC vector data fusion

It is of interest to determine the classification accuracies that can be expected through different channel–stimuli combinations. Therefore, the goal in this section is to determine the classification performance as a function of the number of MSMC ERPs. The $(S \times M)$ ERPs can clearly be combined in numerous ways. To simplify the selection of the combinations, the MSMC ERPs can be systematically combined using the rankings established in the previous section. That is, let $R_{s/m}$ be the rank of $Z_{s/m}$ and let

$$Z_j = Z_{s/m} \quad \text{if } R_{s/m} = j. \quad (7)$$

As a result of the ranking, Z_j is ordered such that Z_1 is the MSMC ERP vector yielding the highest classification accuracy and $Z_{(M \times S)}$ is the MSMC ERP yielding the lowest classification accuracy. Let U_L be the data fusion vector formed according to

$$U_L = \mathop{\nabla}_{j=1}^{L} Z_j, \quad (8)$$

where ∇ represents the concatenation operation. That is, the fusion vector $U_L = (Z_1, Z_2, \ldots, Z_L)^{\mathrm{T}}$ is formed by concatenating the ERP vectors with the first L ranks to form a vector of dimension $(L \times K)$. Using U_L, the Euclidean-norm nearest-mean vector discriminant function for category c is given by

$$g_{c;L}(U_L) = (U^{\mathrm{T}}_L)(\bar{U}_{L/c}) - (1/2)(\bar{U}^{\mathrm{T}}_{L/c})(\bar{U}_{L/c}), \quad (9)$$

where $\bar{U}_{L/c}$ is the mean of the fusion vectors of category c. A test fusion vector U_L is assigned to the category c^*_L, $c^*_L \in \{c\}$, given by

$$c^*_L = \arg\max_{c}[g_{c;L}(U_L)]. \quad (10)$$

The MSMC vector data fusion strategy is summarized in Fig. 2.

4. MSMC vector decision fusion

An alternative to data fusion is decision fusion in which the decisions of individual classifiers are combined to decide the category of a test pattern (see Fig. 3). As in the previous section, let

$$Z_j = Z_{s/m} \quad \text{if } R_{s/m} = j \quad (11)$$

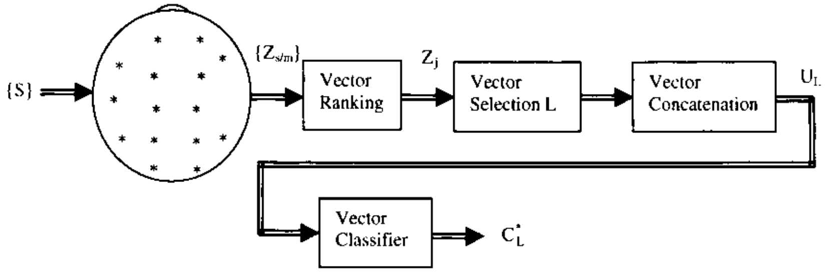

Fig. 2. The MSMC vector data fusion strategy.

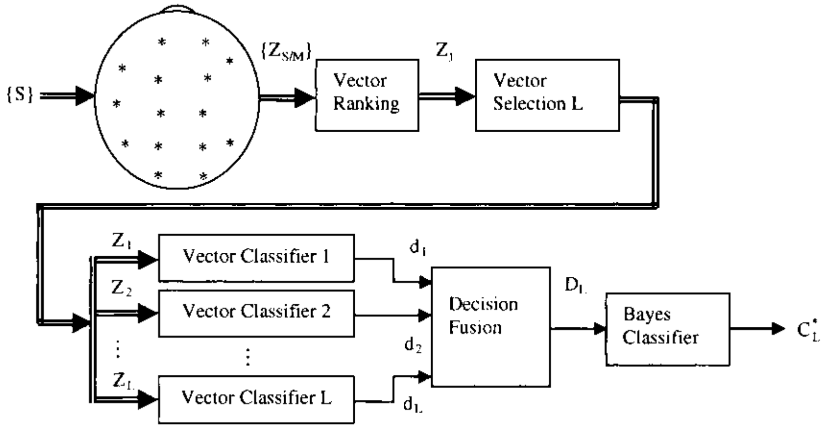

Fig. 3. The MSMC vector decision fusion strategy.

and let

$$c_j^* = \arg\max_c [g_{c;j}(Z_j)], \tag{12}$$

where

$$g_{c;j}(Z_j) = (Z_L^{\mathrm{T}})(\bar{Z}_j) - (1/2)(\bar{Z}_j^{\mathrm{T}})(\bar{Z}_j) \tag{13}$$

is the nearest-mean vector discriminant function for category c and $\bar{Z}_j$ is the mean of Z_j belonging to category c. In order to fuse the decisions, let

$$d_j = c_j^* \tag{14}$$

and let

$$D_L = \mathop{\nabla}_{j=1}^{L} d_j. \tag{15}$$

That is, $D_L = (d_1, d_2, \ldots, d_L)^{\mathrm{T}}$ is the fusion vector formed by concatenating the decisions of the MSMC ERP vector classifiers with the first L ranks. 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 category c be $P(D_L/c)$ and P_c be the a priori probability of category c, then, the Bayes decision function for category c is

$$g_c(D_L) = P_c P(D_L/c). \tag{16}$$

which can also be written as

$$g_c(D_L) = \ln P_c + \ln P(D_L/c). \tag{17}$$

The final decision c_L^*, $c_L^* \in \{c\}$, resulting from decision fusion is given by

$$c_L^* = \arg\max_c [g_c(D_L)]. \tag{18}$$

The discriminant function for the C-category case can be derived explicitly by letting

$$p_{j,a/c} = P(d_j = a/c), \quad j = 1, 2, \ldots, L. \tag{19}$$

That is, $p_{j,a/c}$ is the probability that $d_j = a$, $a \in \{c\}$, when the true category is c. The probability density function (PDF) of d_j under category c can be written as

$$P(d_j/c) = (p_{j,1/c})^{\delta(d_j-1)}(p_{j,2/c})^{\delta(d_j-2)} \ldots \times (p_{j,C/c})^{\delta(dj-C)}. \tag{20}$$

Because the classifiers are developed independently for each MSMC ERP type, we assume that the decisions of the MSMC ERP classifiers are independent, therefore, the PDF of D_L under the category c can be written as

$$P(D_L/c) = \prod_{j=1}^{L} (p_{j,1/c})^{\delta(d_j-1)} (p_{j,2/c})^{\delta(d_j-2)} \dots \times (p_{j,C/c})^{\delta(dj-C)}, \tag{21}$$

where

$$\delta(x-a) = \begin{cases} 1 & \text{if } x = a, \\ 0 & \text{if } x \neq a. \end{cases} \tag{22}$$

By substituting the PDFs into Eq. (17), it can be shown that the discriminant function for category c can be written as

$$g_c(D_L) = \sum_{j=1}^{L} [\delta(d_j - 1)\ln(p_{j,1/c}) + \delta(d_j - 2)\ln(p_{j,2/c}) + \cdots + \delta(d_j - C)\ln(p_{j,C/c})] + \ln P_c. \tag{23}$$

5. MSMC element data fusion

In this data fusion strategy, the fusion vector is formed by combining selected elements of MSMC ERPs. The fusion vector Z is formed by first concatenating all MSMC ERPs into a vector of dimension $(S \times M \times K)$. That is

$$Z = \mathop{\nabla}_{s=1}^{S} \mathop{\nabla}_{m=1}^{M} Z_{s/m}. \tag{24}$$

Let $z(j)$, $j = 1, 2, \dots, (S \times M \times K)$, be the jth element of Z and $d_j = c_j^*$, $c_j^* \in \{c\}$, where,

$$c_j^* = \arg\max_c [g_c(z(j))] \tag{25}$$

and

$$g_c(z(j)) = z(j)\bar{z}_c(j) - (1/2)[\bar{z}_c(j)]^2 \tag{26}$$

is the nearest-mean discriminant function for category c and $\bar{z}_c(j)$ is the mean of $z(j)$ under category c. Let α_j be the classification accuracy of the nearest-mean scalar classifier for element $z(j)$ and let R_j be the rank of $z(j)$ according to the classification accuracies. Let

$$u(j) = z(k) \quad \text{if } R_k = j. \tag{27}$$

As a result of the ranking, $u(1)$ is the element yielding the highest classification accuracy and $u(M \times S \times K)$ is the element yielding the lowest classification accuracy. Let U_L be the data fusion vector formed according to

$$U_L = \mathop{\nabla}_{j=1}^{L} u(j). \tag{28}$$

That is, the fusion vector $U_L = [u(1), u(2), \dots, u(L)]^{\mathrm{T}}$ is formed by concatenating the elements with the first L ranks to form a vector of dimension L. Note that the elements of U_L are samples from various MSMC ERPs. This corresponds to selecting different time-instants from different MSMC ERPs. Using U_L, the Euclidean-norm nearest-mean discriminant function for category c is given by

$$g_{c:L}(U_L) = (U_L^{\mathrm{T}})(\bar{U}_{L/c}) - (1/2)(\bar{U}_{L/c}^{\mathrm{T}})(\bar{U}_{L/c}), \tag{29}$$

where $\bar{U}_{L/c}$ is the mean of the fusion vector of category c. A test fusion vector U_L is assigned to the category c_L^*, $c_L^* \in \{c\}$, given by

$$c_L^* = \arg\max_c [g_{c:L}(U_L)]. \tag{30}$$

The MSMC element data fusion strategy is summarized in Fig. 4.

6. MSMC element decision fusion

The fusion vector for this case is formed by combining the independent decisions of selected MSMC ERP element scalar classifiers. Let $d_j = c_j^*$ where c_j^*, $c_j^* \in \{c\}$, is given by

$$c_j^* = \arg\max_c [g_c u(j)], \tag{31}$$

where $u(j)$ is defined in Eq. (27) and

$$g_c(u(j)) = u(j)\bar{u}_c(j) - (1/2)[\bar{u}_c(j)]^2 \tag{32}$$

is the nearest-mean discriminant function for category c and $\bar{u}_c(j)$ is the mean of $u(j)$ belonging to category c. The decision fusion vector for this case is given by

$$D_L = \mathop{\nabla}_{j=1}^{L} d_j. \tag{33}$$

That is, $D_L = (d_1, d_2, \dots, d_L)^{\mathrm{T}}$ is the fusion vector formed by concatenating the decisions of the nearest-mean element scalar classifiers with the first L ranks. Given D_L, the sequence of steps to derive the decision rule is identical to those described in Section 4. The discriminant function is given by Eq. (23). The probability $p_{j,a/c}$ in the discriminant function for this case is the probability that the scalar classifier for $u(j)$ decides category a when the true

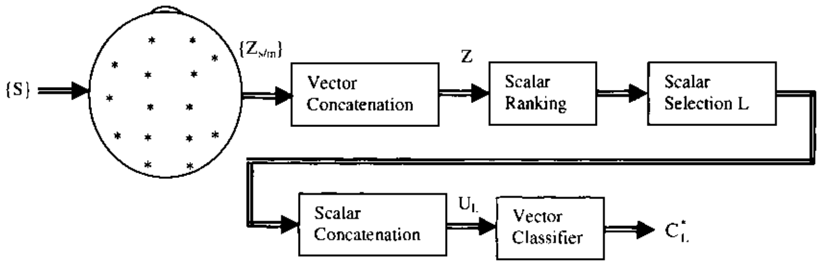

Fig. 4. The MSMC element data fusion strategy.

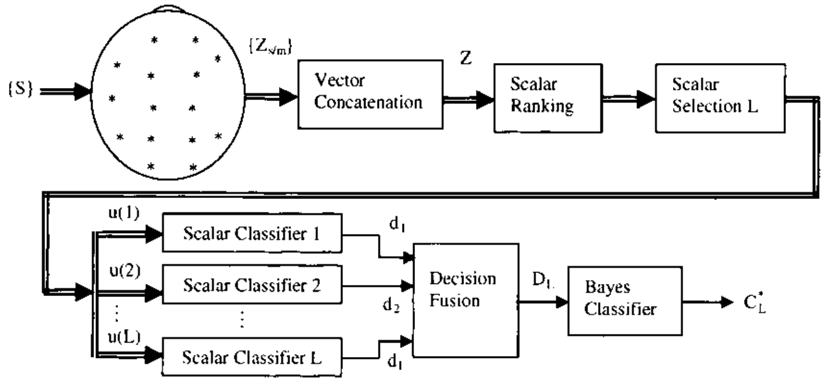

Fig. 5. The MSMC element decision fusion strategy.

category is c. The MSMC element decision fusion strategy is summarized in Fig. 5.

7. Performance measures

The performance of the four fusion classification strategies can be specified in terms of the classification decision probability or the classification accuracy which are computed from the classification rates. The classification rate for category c is given by

$$\alpha_{c.L} = [1 - P_{c.L}(\varepsilon)] \times 100\%, \tag{34}$$

where $P_{c.L}(\varepsilon)$ is the probability of misclassifying the data fusion vector U_L of Section 3 when the true category of U_L is c. $P_{c.L}(\varepsilon)$ is the probability of misclassifying the decision fusion vector D_L of Section 4 when the true category of D_L is c. $P_{c.L}(\varepsilon)$ is the probability of misclassifying the data fusion vector U_L of Section 5 when the true category of U_L is c. $P_{c.L}(\varepsilon)$ is the probability of misclassifying the decision fusion vector D_L of Section 6 when the true category of D_L is c.

The corresponding classification decision probability can be determined as

$$\alpha_L = [1 - P_L(\varepsilon)], \tag{35}$$

where

$$P_L(\varepsilon) = \sum_{c=1}^{C} P_{c.L}(\varepsilon) P_c. \tag{36}$$

The classification accuracy, expressed as a percentage, is given by $(\alpha_L \times 100)\%$.

8. Experiments and results

The data used to design and evaluate the classifiers were exactly the same as that used in the study described in Ref. [2] except that the signals were not down-sampled. Each ERP, therefore, consisted of 140 samples (140 elements in the ERP vector). Each child was represented by $(6 \times 4) = 24$ MSMC averaged ERPs. The means and the probabilities of each discriminant function were estimated from the respective training sets of each category.

8.1. Control/dyslexia (2-category classification)

The 48 children were grouped into two categories: control ($C = 1$) and dyslexia ($C = 2$). The dyslexic and poor readers were grouped into the dyslexic group. The number of children in each category was 24. Because dyslexia occurs in approximately 10% of the population, the prior

Table 1
2-category ranking

(a) Classification accuracies of each MSMC ERP type

s	m = 1	2	3	4	5	6
1	48.14	48.32	45.43	46.18	53.99	39.39
2	48.77	52.98	51.46	49.93	40.03	52.38
3	45.34	50.58	53.52	55.4	53.67	51.47
4	52	51.2	41.85	48.16	57.06	43.68

(b) The channel–stimulus, stimulus, and channel rankings

s	m = 1	2	3	4	5	6	R_s
1	17	15	19	18	3	24	4
2	14	6	10	13	23	7	2
3	20	12	5	2	4	9	1
4	8	11	22	16	1	21	3
R_m	5	2	4	3	1	6	

probabilities P_1 (control) and P_2 (dyslexia) were selected, to be 0.9 and 0.1, respectively. The MSMC ERPs of each category were partitioned randomly into two mutually exclusive equi-sized sets to form the training set and the test set. The channels $m = 1$, 2, 3, 4, 5, and 6 are FL, FR, T3, T4, PL, and PR and the stimuli $s = 1$, 2, 3, and 4 are /bi/, /gi/, /nbi/, and /ngi/, respectively. Table 1 shows the classification accuracies of each MSMC ERP type and the rankings of the channel–stimuli, stimuli, and channels obtained from the rankings of the classification accuracies. The entries in Table 1(a) show the classification accuracies $\alpha_{s/m}$ of each ERP type, where, as described in Section 2, an ERP type is the ERP elicited at a given channel in response to a given stimulus. The number of ERP types in this study is 24 (6 channels × 4 stimuli). The first entry in Table 1(a) is, therefore, the classification accuracy of the ERPs elicited at channel 1 (FL) in response to stimulus 1 (/bi/). The first entry in Table 1(b) shows the rank $R_{s/m} = R_{1/1}$ of the ERP at channel 1 (FL) in response to the stimulus $s = 1$ (/bi/). $R_{4/5}$ has rank 1 because the ERPs of channel 5 in response to stimulus 4, give the highest classification accuracy and $R_{1/6}$ has rank 24 because the ERPs of channel 6 in response to stimulus 1 yield the lowest classification accuracies. The last row gives the individual rank of each channel computed from the marginal rank sum of Eq. (6) and the last column gives the rank of each stimulus computed from the marginal rank sum of Eq. (5). For example, $R_m = R_5 = 1$ because the sum of the $R_{s/m}$ values in column 5 has the smallest value. Similarly, $R_s = R_3 = 1$ because the sum of the $R_{s/m}$ values in row 3 has the smallest value.

Table 2
MSMC vector data fusion results

(a) Classification decision probabilities for the maximum classification accuracy ($\alpha_2 = 58.29\%$)

True class	Classifier decision $Y = 1$	$Y = 2$
$C = 1$	0.58583	0.41417
$C = 2$	0.44333	0.55667

(b) Posterior probabilities for the maximum classification accuracy ($\alpha_2 = 58.29\%$)

Classifier decision	True class $C = 1$	$C = 2$
$Y = 1$	0.92244	0.077562
$Y = 2$	0.87006	0.12994

Table 3
MSMC vector decision fusion results

(a) Classification decision probabilities for the maximum classification accuracy ($\alpha_{24} = 90.29\%$)

True class	Classifier decision $Y = 1$	$Y = 2$
$C = 1$	0.99917	0.00083
$C = 2$	0.96333	0.036667

(b) Posterior probabilities for the maximum classification accuracy ($\alpha_{24} = 90.29\%$)

Classifier decision	True class $C = 1$	$C = 2$
$Y = 1$	0.90324	0.09676
$Y = 2$	0.16925	0.83075

Tables 2–5 show, for the four fusion strategies, the decision probabilities and the posterior probabilities for L giving the maximum classification accuracy. The best value of L was selected by evaluating the performances for all possible values of L (1–24 for vector data and vector decision fusion; 1–3360 (6 × 4 × 140 elements) for element data and element decision fusion). All classification results were estimated by averaging the decision probabilities over 100 × 12 = 1200 tests for each category. The classifier decision is represented by Y, where, $Y = 1$ when the classifier decision is the control category and $Y = 2$ if the classifiers decision is the dyslexic category. In order to interpret the results in Tables 2–5, consider Tables 2(a) and (b) which show the classifier's decision probabilities and the corresponding posterior probabilities, respectively. The first entry in Table 2(a) is the conditional probability $P(Y = 1/C = 1)$, that is, the probability that the

Table 4
MSMC element data fusion results

(a) Classification decision probabilities for the maximum classification accuracy ($\alpha_{336} = 87.97\%$)

True class	Classifier decision	
	$Y = 1$	$Y = 2$
$C = 1$	0.88667	0.15417
$C = 2$	0.18333	0.88583

(b) Posterior probabilities for the maximum classification accuracy ($\alpha_{336} = 87.97\%$)

Classifier decision	True class	
	$C = 1$	$C = 2$
$Y = 1$	0.97754	0.022458
$Y = 2$	0.55535	0.44465

Table 5
MSMC element decision fusion results

(a) Classification decision probabilities for the maximum classification accuracy ($\alpha_{2822} = 99.88\%$)

True class	Classifier decision	
	$Y = 1$	$Y = 2$
$C = 1$	1	0
$C = 2$	0.01167	0.98833

(b) Posterior probabilities for the maximum classification accuracy ($\alpha_{2822} = 99.88\%$)

Classifier decision	True class	
	$C = 1$	$C = 2$
$Y = 1$	0.99871	0.001295
$Y = 2$	0	1

classifier correctly classifies control subjects. The first entry in Table 2(b) is the conditional probability $P(C = 1/Y = 1)$ which gives the probability that the test subject is from the control group given that the classifier has decided that the test subject is from the control group (true-negative probability). The posterior probabilities are computed using Bayes rule. Table 6 shows the interpretation of the posterior probabilities in terms of the probabilities of true negatives, false negatives, false positives, and true positives.

For comparison, the results of discriminant analysis from the previous study, for the two-category case, using the prior probabilities of 0.9 and 0.1 are shown in Table 7. In studies such as the one reported in this paper, the posterior probabilities are quite important because the performance in terms of the classification accuracy by itself can be quite misleading. Consider, for example, the decision probability of 0.8125 which was reported in Ref. [2]. Although the classifier's decision probability is quite high, the probabilities of a false positive and a true positive are quite high (0.6735) and quite low (0.3264), respectively, as seen in Table 7(b).

Table 6
Positions of true negatives, false negatives, false positives, and true positives in Tables 2(b)–5(b)

Classifier decision	True class	
	$C = 1$	$C = 2$
$Y = 1$	True negative	False negative
$Y = 2$	False positive	True positive

Table 7
2-category discriminant analysis results from previous study

(a) Classification decision probabilities

True class	Classifier decision	
	$Y = 1$	$Y = 2$
$C = 1$	0.79	0.21
$C = 2$	0.083	0.916

(b) Posterior probabilities

Classifier decision	True class	
	$C = 1$	$C = 2$
$Y = 1$	0.9884	0.01153
$Y = 2$	0.6735	0.3264

8.2. Control/dyslexia/poor readers (3-category classification)

The 48 children were grouped into three categories: control ($C = 1$), dyslexia ($C = 2$), and poor readers ($C = 3$). The corresponding classifier decisions are represented by $Y = 1$, $Y = 2$, and $Y = 3$. The number of children in the categories were 24, 17, and 7, respectively. The prior probabilities P_1 (control), P_2 (dyslexia), and P_3 (poor readers) were selected, to be 0.9, $(17/24 \times 0.1) = 0.07$, and $(7/24 \times 0.1) = 0.03$, respectively. Because of the small number of MSMC ERPs in the poor reader category, the MSMC ERPs of each category were partitioned using the "leave-one-out" method to form the training set and the test set. Table 8 (a) shows the estimated classification accuracies $\alpha_{s/m}$ of each MSMC ERP type as described in Section 2. The channel–stimulus, stimulus, and channel rankings are presented in Table 8(b). All results were estimated by averaging the classification accuracies using the "leave-one-out method." Tables 9–12 show, for the four fusion strategies, the results for L giving the maximum decision probabilities and the corresponding posterior probabilities. For comparison, the results of

Table 8
3-category rankings

(a) Classification accuracies of each MSMC ERP type

s	m = 1	2	3	4	5	6
1	28.12	31.84	28.42	34.60	44.50	12.59
2	23.81	50.95	45.58	38.79	22.70	43.78
3	25.10	41.96	40.42	45.74	44.43	46.37
4	33.53	27.42	20.43	27.39	48.54	22.85

(b) The channel–stimulus, stimulus, and channel rankings

s	m = 1	2	3	4	5	6	R_s
1	16	14	15	12	6	24	3
2	20	1	5	11	22	8	2
3	19	9	10	4	7	3	1
4	13	17	23	18	2	21	4
R_m	6	2	4	3	1	5	

Table 9
MSMC vector data fusion results (3-category)

(a) Classification decision probabilities for the maximum classification accuracy ($\alpha_{11} = 60.78\%$)

True class	Classifier decision $Y = 1$	$Y = 2$	$Y = 3$
$C = 1$	0.64671	0.32913	0.02416
$C = 2$	0.56127	0.30462	0.1341
$C = 3$	0.18382	0.67332	0.14286

(b) Posterior probabilities for the maximum classification accuracy ($\alpha_{11} = 60.78\%$)

Classifier decision	True class $C = 1$	$C = 2$	$C = 3$
$Y = 1$	0.92806	0.063392	0.0085489
$Y = 2$	0.87786	0.063945	0.058199
$Y = 3$	0.61407	0.26826	0.11767

Table 10
MSMC vector decision fusion results (3-category)

(a) Classification decision probabilities for the maximum classification accuracy ($\alpha_{24} = 95.65\%$)

True class	Classifier decision $Y = 1$	$Y = 2$	$Y = 3$
$C = 1$	0.99125	0.007703	0.001050
$C = 2$	0.382	0.59139	0.026611
$C = 3$	0.15371	0.07458	0.77171

(b) Posterior probabilities for the maximum classification accuracy ($\alpha_{24} = 95.65\%$)

Classifier decision	True class $C = 1$	$C = 2$	$C = 3$
$Y = 1$	0.96585	0.02930	0.004854
$Y = 2$	0.13594	0.8214	0.042654
$Y = 3$	0.03731	0.07439	0.8883

Table 11
MSMC element data fusion results (3-category)

(a) Classification decision probabilities for the maximum classification accuracy ($\alpha_{260} = 77.14\%$)

True class	Classifier decision $Y = 1$	$Y = 2$	$Y = 3$
$C = 1$	0.83333	0.15406	0.012601
$C = 2$	0.39881	0.30182	0.29937
$C = 3$	0	1	0

(b) Posterior probabilities for the maximum classification accuracy ($\alpha_{260} = 77.14\%$)

Classifier decision	True class $C = 1$	$C = 2$	$C = 3$
$Y = 1$	0.9637	0.036298	0
$Y = 2$	0.73285	0.113	0.15416
$Y = 3$	0.34853	0.65147	0

discriminant analysis from the previous study, for the three-category case, using the prior probabilities of 0.9, 0.07, and 0.03 are shown in Table 13.

9. Conclusions

The goal in this paper was to obtain very high accuracies for predicting dyslexia in children from their MSMC ERPs recorded at birth. A ranking strategy was developed to rank the channel–stimuli combinations, channels, stimuli, and ERP elements in terms of their effectiveness in classifying dyslexia. The prediction problem was formulated as a classification problem and the rankings were used to develop two MSMC data-fusion and two decision-fusion classification strategies. The data-fusion methods systematically combined the rank-ordered MSMC ERP vectors or the rank-ordered MSMC ERP elements. For each case, the resulting data fusion vector was classified using a single vector nearest-mean classifier. In the decision fusion strategy, the independent classification decisions of rank-ordered MSMC ERP vectors or the independent classification decisions of rank-ordered MSMC ERP elements

Table 12
MSMC element decision fusion results (3-category)

(a) Classification decision probabilities for the maximum classification accuracy ($\alpha_{497} = 100\%$)

True class	Classifier decision		
	$Y = 1$	$Y = 2$	$Y = 3$
$C = 1$	1	0	0
$C = 2$	0	1	0
$C = 3$	0	0	1

(b) Posterior probabilities for the maximum classification accuracy ($\alpha_{497} = 100\%$)

Classifier decision	True class		
	$C = 1$	$C = 2$	$C = 3$
$Y = 1$	1	0	0
$Y = 2$	0	1	0
$Y = 3$	0	0	1

Table 13
3-category discriminant analysis results from previous study

(a) Classification decision probabilities

True class	Classifier decision		
	$Y = 1$	$Y = 2$	$Y = 3$
$C = 1$	0.79	0.08	0.13
$C = 2$	0.12	0.76	0.12
$C = 3$	0	0	1

(b) Posterior probabilities

Classifier decision	True class		
	$C = 1$	$C = 2$	$C = 3$
$Y = 1$	0.9882	0.0118	0
$Y = 2$	0.5752	0.4278	0
$Y = 3$	0.7665	0.0550	0.1886

were combined into a decision fusion vector. The resulting decision fusion vector for each case was classified using a Bayes discrete vector classifier.

The fusion classification strategies were tested using exactly the same data in the study reported in Ref. [2]. The results presented facilitate performance comparisons not only between the four fusion classification strategies but also between ERP vector fusion versus ERP element fusion as well as data fusion versus decision fusion. The best results were obtained using MSMC element decision fusion. It was shown that a classification accuracy of 99.88% with zero false positives are possible for the control/dyslexia prediction problem using $L = 2822$ ERP elements out of the total of 3360 elements. For the control/dyslexia/poor reader prediction problem, a classification accuracy of 100% was possible using "leave-one-out" evaluations using $L = 497$ ERP elements. Thus it can be concluded that future reading difficulties can be predicted with almost 100% accuracy. The significance of the decision fusion results can be appreciated by noting that, individually, the best ($L = 1$) MSMC ERP vector and MSMC ERP element give classification accuracies of only 57.06% and 74.43% for the two-category case, respectively, and 50.95% and 69.95% for the three-category case, respectively. Additionally, the significant reduction in the false positive probabilities must be noted by comparing the results with those reported in Ref. [2] (Table 7(b)). In terms of complexity, the data fusion strategies require only a single vector classifier whereas the decision fusion classifiers require multiple element (scalar) classifiers and an additional Bayes discrete vector classifier. The scalar classifiers are, however, computationally relatively simple.

The results presented in this paper further support the findings reported in Ref. [2]. That is, auditory ERPs recorded within 36 h of birth can successfully predict reading performance in children 8 years later. Therefore, potential problems in language or cognitive development can be identified at birth and, consequently, planned interventions can be introduced earlier to the child and be more successful in the remediation of the child's later emerging language problems. This will have a tremendously positive impact by avoiding the psychological damage resulting from being labeled "slow", enabling the child to take full advantage of his/her schooling and thus make it possible to reach his/her full intellectual potential.

In summary, it is shown that through the MSMC element decision fusion classification strategy, dyslexia and poor reading can be predicted with almost 100% accuracy from the ERPs of infants. Furthermore, because the formulations of the MSMC fusion classification strategies are quite general, they can also be applied to other multi-category prediction/classification problems involving ERPs and, in general, to multi-category multi-sensor signal classification problems.

Acknowledgements

We would like to thank the reviewer for the helpful comments and suggestions to improve the paper.

References

[1] L. Gupta, J. Phegley, D.L. Molfese, Parametric classification of multichannel evoked potentials, IEEE Trans. Biomed. Eng. 49 (8) (2002) 905–911 49(9) (2002) 1070.
[2] D.L. Molfese, Predicting dyslexia at 8 years of age using neonatal brain responses, Brain Language 72 (2000) 238–245.
[3] J. Phegley, K. Perkins, L. Gupta, L. Hughes, Multi-category prediction of multifactorial diseases through risk factor fusion and rank sum selection, IEEE Trans. Syst. Man Cyber. - A, in press.

About the Author—H. KOOK received the B.E. degree in electronics engineering from Wonkwang University, Korea (1995), the M.S. degree in electronics engineering from Wonkwang University, Korea (1997), and the M.S. degree in electrical engineering from Southern Illinois University at Carbondale (2000). He is currently finishing his Ph.D. in the Department of Electrical and Computer Engineering at Southern Illinois University, Carbondale, IL. His research interests are in pattern recognition, signal processing, neuroinformatics, and image processing.

About the Author—L. 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, TX (1986). He is currently a Professor of Electrical and Computer Engineering at Southern Illinois University, Carbondale, IL. His research interests are in pattern recognition, signal processing, neuroinformatics, and neural networks. He has served as a consultant for the Army Research office (ARO) on smart munitions design and for Seagate Technology on image compression. He is currently a signal processing and pattern recognition consultant for Think-A-Move, Neuronetrix Inc., and Cleveland Medical Devices on NIH funded projects. Dr. Gupta has numerous publications in the fields of pattern recognition, neural networks, and evoked potentials. He is an Associate Editor of the Pattern Recognition Journal, is a Senior Member of the IEEE, and his accomplishments are listed in the prestigious Marquis Who's Who in America (2002).

About the Author—D.L. MOLFESE Dr. Molfese received his Ph.D. in Psychology from the Pennsylvania State University in 1972. He is currently a Distinguished University Scholar and serves as Chair of the Department of Psychological and Brain Sciences at the University of Louisville. He has served as the Chair of a number of national panels on Learning Disabilities. He is co-director of one of 15 national laboratories that make up the National Institutes of Health Reading and Learning Disabilities Research Network. Dr. Molfese is the recipient of a number of honors for outstanding research contributions from societies such as Sigma Xi and Phi Kappa Phi and is also the recipient an Outstanding Scholar Award from Southern Illinois University. He is a Fellow of the American Psychological Association and the American Psychological Society and serves as the Editor-in-Chief for the scientific journal, Developmental Neuropsychology. In addition to his collaborations and consultations with a number of laboratories in the USA, Finland, and The Netherlands, Dr. Molfese also serves as a consultant on infant and child development issues to the country of Belize and is working with the State of South Carolina to set up clinical trials to test a new form of early intervention to eliminate learning disabilities in the next generation of Americans. Dr. Molfese has organized over 30 national and international conferences on brain processes and language, has given some 80 invited lectures, and has over 110 national and international conference presentations on this topic. He has published over 120 books, journal articles, and book chapters on the relationship between brain functions and cognitive processes. Dr. Molfese has obtained research funding for his work from the National Institutes of Health, the National Science Foundation, Department of Education, The National Foundation/March of Dimes, the MacArthur Foundation, the Kellogg Foundation, NATO, and NASA.

About the Author—K.C. FADEM graduated from Georgia Institute of Technology's School of Aerospace Engineering in 1981. He spent 11 years working on software systems for solids modeling, finite element modeling and analysis, computational fluid dynamics, materials characterization, and mechanisms analysis. In 1993, he founded Val-U-Med, a development and marketing company for non-invasive surgical devices. After selling the company to Dexterity Surgical, Mr. Fadem developed novel instrumentation for hand-assisted laparoscopy, and minimally invasive pelvic surgery. Currently, he is a founder and the Chief Operating Officer for Neuronetrix, Inc. where he is developing a high-throughput screening device which will identify infants who may be at risk for neurological disorders.