Response Error Correction - A Demonstration of Improved Human-Machine Performance Using Real-Time EEG Monitoring

Lucas C. Parra, Clay D. Spence, Adam D. Gerson and Paul Sajda

Abstract— We describe a brain computer interface (BCI) system, which uses a set of adaptive linear preprocessing and classification algorithms for single-trial detection of Error Related Negativity (ERN). We use the detected ERN as an estimate of a subject's perceived error during an alternative forced choice visual discrimination task. The detected ERN is used to correct subject errors. Our initial results show average improvement in subject performance of 21% when errors are automatically corrected via the BCI. We are currently investigating the generalization of the overall approach to other tasks and stimulus paradigms.

Keywords— Brain Computer Interface, Error Related Negativity, single-trial detection, electroencephalography, eye blink removal

I. ADAPTIVE HUMAN-COMPUTER INTERFACE

The performance of a human subject executing a task while interacting with a computer can be highly variable, depending upon such individual factors as level of alertness, reaction speed, working memory capacity, and capacity to perform parallel tasks. Most current human-computer interfaces (HCI) do not adapt to the physiological or psychological state of the user. The goal of an *adaptive* interface is to estimate variables correlated with human performance and adapt the HCI accordingly (e.g. adjust speed of display, provide appropriate cues, automatically correct errors, etc.) Several behavioral and physiological measures, such as reaction time, eye motion, and pupil dilation, have been proposed as variables having utility for adapting an HCI [1], [2]. More recently, research in neuroimaging has identified electroencephalography (EEG) signals that are correlated with attention [3], memory encoding [4], motor imagery [5], perceived error and/or conflict [6], perception/recognition [7] and which therefore might be useful for such adaptation.

In this paper we describe a brain computer interface (B-CI) capable of monitoring a subject's cognitive state associated with specific observable events. We argue that this information can be used to adapt the HCI, and ultimately maximize performance. As an example we demonstrate our initial results using a high throughput, alternative forced choice visual discrimination task. In this task a subject discriminates between two visual stimuli by pressing one of two buttons. When subjects attempt to minimize their response time they often commit errors that are

Lucas Parra and Clay Spence are with Sarnoff Corporation, Princeton, NJ, USA. E-mail: lparra@sarnoff.com, cspence@sarnoff.com. Paul Sajda and Adam Gerson are with The Department of Biomedical Engineering, Columbia University, New York, NY, USA. E-mail: ps629@columbia.edu, adg71@columbia.edu

perceived shortly after the button-push response.[1] Interestingly, such perceived errors are accompanied by a negative fronto-central deflection in the EEG signal. This signal is known as the error related negativity (ERN) [9]. Single-trial detection of the ERN has been proposed as a means of correcting communication errors in a BCI system [10]. For off-line processing we have recently reported, using a simple linear classifier, up to 79% correct detection of the ERN within 100 ms of the erroneous response [8].

In this paper we describe a BCI system, which uses the detected ERN to correct erroneous responses of subjects so as to minimize the overall human-machine error rate. In particular, we focus on a set of adaptive, linear algorithms for data preprocessing and ERN detection. The algorithms are designed for single-trial, on-line processing and run in real-time on data that is streaming from a 64 channel EEG system through Ethernet. The latency of the total system is approximately 230 ms which is primarily constrained by the 200 ms integration time required for ERN detection.

BCI systems are currently being developed primarily as a communications means for severely motor impaired subjects. We are currently exploring other applications of the proposed real-time monitoring. For instance, we argue that one may be able to increase the speed of visual search for image analysts by detecting the activity associated with fast visual recognition. Single-trial detection of this activity permits bypassing slow motor response thereby increasing human performance as we have demonstrated in a rapid visual serial presentation (RSVP) paradigm [11].

II. ADAPTIVE LINEAR ON-LINE REAL-TIME PROCESSING

We describe in the following two sections novel algorithms for EEG eye blink removal and denoising. These algorithms preserve the ERN signal and in fact improve detection accuracy as presented in section II-C. In all cases we first remove baseline drifts, most likely due to slow skin conductivity changes, by filtering each channel with a recursive linear high-pass filter (2nd order Butterworth with 1 Hz cutoff frequency).

A. Robust eye blink removal with principal component analysis

The muscle activity of eye blinks generates strong electrical signals that are linearly superimposed with smaller magnitude signals of interest (e.g. the ERN). We present here a novel algorithm for eye blink removal. Conventional

---

[1]For more detailed description of the experimental paradigm see [8]

algorithms detect eye blinks to simply discard the corresponding segment of data. This is not feasible in practice given the frequent eye motion in most real world situations. A better approach is to subtract the artifacts using linear regression algorithms (e.g. [12]). Existing methods typically use electro-oculogram (EOG) electrodes as a reference. Unfortunately in addition to eye motion, EOG signals also contain frontal cortical activity which should not be subtracted. We construct a better reference signal using a linear combination of all electrodes thereby increasing the power of the eye blink activity in the reference.

Let $\mathbf{x}(t)$ be the observed EEG sensor reading, $y(t)$ the eye blink signal, and $\mathbf{s}(t)$ the remaining signal of interest,

$$\mathbf{x}(t) = \mathbf{a}y(t) + \mathbf{s}(t). \qquad (1)$$

The linear coupling of the eye blink source $y(t)$ with the EEG sensors is denoted as $\mathbf{a}$. The arbitrary scale in the factorization, $\mathbf{a}y(t)$, can be constrained by setting, $\mathbf{a}^T\mathbf{a} = 1$. Given $\mathbf{a}$, we can generate a new signal $\tilde{\mathbf{x}}(t)$, which only contains the signal of interest when the orientation of the eye blink is removed,

$$\tilde{\mathbf{x}}(t) = (\mathbf{I} - \mathbf{a}\mathbf{a}^T)\mathbf{x}(t) = (\mathbf{I} - \mathbf{a}\mathbf{a}^T)\mathbf{s}(t). \qquad (2)$$

Only source signals that are exactly collinear with $\mathbf{a}$ are removed. All other source signals will have at least a residual contribution to $\tilde{\mathbf{x}}(t)$ and may be detected with appropriate algorithms.

The key challenge lies in estimating the coupling $\mathbf{a}$. Recent work suggests to use independent component analysis (ICA) in order to identify such linear projection associated with EEG artifacts [13]. In our experience, however, ICA methods have not proven robust when applied on-line. Fortunately, during eye blinks the activity $y(t)$ is large compared to the signals of interest, $\mathbf{s}(t)$. This is in particular true for frontal electrodes and electro-oculogram (EOG) signals commonly acquired with EEG. The projection $\mathbf{a}$ can therefore be identified using principal component analysis (PCA) for which a number of robust on-line algorithms are available.

Our method partitions the 64 recorded channels into two sets: (1) the EOG and frontal electrodes containing strong eye blink signals and (2) the remaining parietal, temporal, and occipital electrodes with weaker eye blink signal contributions. Let us denote this partitioning with, $\mathbf{x}(t) = [\mathbf{x}_1(t), \mathbf{x}_2(t)]$, and correspondingly $\mathbf{a} = [\mathbf{a}_1, \mathbf{a}_2]$. We propose to identify $\mathbf{a}_1$ as the first principal component of $\mathbf{x}_1(t)$ estimated *during eye blinks* as demonstrated in Figure 1. Given the coupling $\mathbf{a}_1$ we obtain an estimate for the eye blink signal as, $\hat{y} = \mathbf{a}_1^T\mathbf{x}_1(t)$, since $\hat{y}(t) \approx y(t)$ for the case that $y(t) \ll \mathbf{a}^T\mathbf{s}$. The partitioning further weights the eye muscle activity in the estimate $\hat{y}$. The remaining coupling factors $\mathbf{a}_2$ can be determined with conventional regression, i.e. as the linear predictors of $\mathbf{x}_2$ given $\hat{y}(t)$. Interestingly the update equations for the principal components $\mathbf{a}_1$, and the regression coefficients $\mathbf{a}_2$ have the same analytic form and can be combined into a single on-line update equation. After normalizing with the output power to
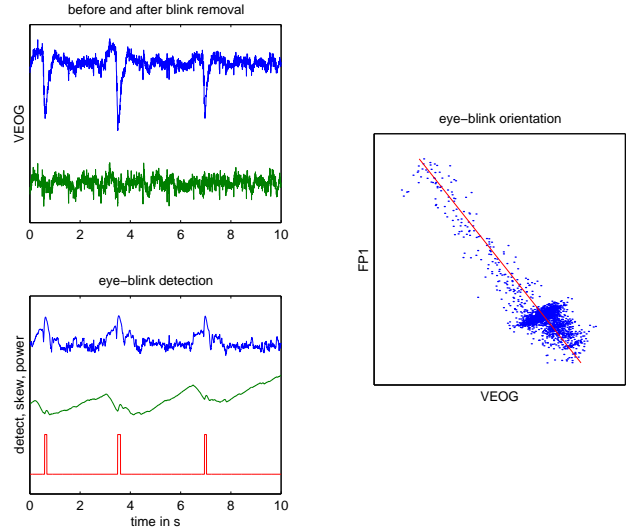

Fig. 1. Eye-blink detection and subtraction using first principal component. Top left: EOG channel before and after subtraction of eye blinks. Right: Scatter plot of the activity in EOG electrode and a frontal electrode during the same 10s time interval. Solid line shows the orientation of the first principal component estimated during eye blinks. Bottom left: Power magnitude (upper trace) and time averaged skew (center trace) of the power in frontal and EOG electrodes, as well as eye blink detector output (lower trace).

accelerate convergence with a constant learning rate μ, the update becomes,

$$\Delta\mathbf{a} = -\mu\left(\mathbf{a} - \frac{\mathbf{R}_{\mathbf{x}\mathbf{x}_1}(t)\mathbf{a}}{\mathbf{a}^T\mathbf{R}_{\mathbf{x}_1\mathbf{x}_1}(t)\mathbf{a}}\right), \qquad (3)$$

where $\mathbf{R}_{\mathbf{x}\mathbf{x}_1}(t)$ and $\mathbf{R}_{\mathbf{x}_1\mathbf{x}_1}(t)$ are the corresponding portions of a running estimate of the covariance of $\mathbf{x}$ during eye blinks, e.g. $\mathbf{R}_{\mathbf{x}\mathbf{x}}(t) = \gamma\mathbf{R}_{\mathbf{x}\mathbf{x}}(t-1) + (1-\gamma)\mathbf{x}(t)\mathbf{x}_1^T(t)$ with a forgetting factor γ. For the required eye blink detection we use the instantaneous magnitude, and a time averaged skew of the power in $\mathbf{x}_1(t)$. A blink is detected when these quantities cross a predetermined threshold.

Figure 1 shows a scatter plot of the signals of two electrodes in the set $\mathbf{x}_1$ and the corresponding orientations in $\mathbf{a}_1$ as they have been identified a few seconds after observing only a single blink. Eye blinks last for approximately 100 ms, which at 250 Hz sampling rate corresponds to about 25 samples to estimate $\mathbf{R}_{\mathbf{x}\mathbf{x}}(t)$. Only a few hundred iterations of (3) are required for convergence, i.e. 2-3 seconds after that eye blink. Subsequent eye blinks are accurately subtracted as seen in Figure 1. We find that $\mathbf{a}_1$ does not change much in the course of an experiment, and only a few eye blinks are required for an accurate estimation of the covariance matrixes. We can therefore set the detector thresholds and learning constant to very conservative values.

B. Denoising the ERN using temporal Hidden-Markov-Trees

One challenge in detecting the cognitive state of a subject via single-trial EEG is the inherently low signal to noise ratio (SNR). In developing denoising algorithms for

improving the SNR, it is critical to obtain accurate noise estimates as well as exploit properties of the signal which preserve its structure. For the purpose of detection noise is defined here as temporal activity in the EEG that is unrelated to the ERN. We therefore develop a statistical noise model using the activity during some unrelated baseline period. This model is then used to remove noise during the time period of the evoked response. We expect the wavelet transform of evoked potentials, as with many other types of natural signals [14], to exhibit the persistence of large or small coefficients across scale and the clustering of coefficients within scale. In addition, observations on the distribution of wavelet coefficients of EEG evoked potentials reveal that the coefficients have near zero mean and long tails–i.e. they are super-Gaussian (see Figure 2A). Here we propose to estimate noise statistics by modeling such signal properties with a hierarchical probability model.

Super-Gaussian distributions may be approximated using a two state, zero mean, Gaussian mixture model in which the large number of small coefficients are modeled with a low variance Gaussian, and the small number of large coefficients are modeled with a high variance Gaussian (Figure 2B). To efficiently describe the statistics of wavelet coefficients, each coefficient is associated with a hidden state variable that describes whether the coefficient is in either a high or low variance state (Figure 2C).

In order to model the conditional relationships described by persistence properties of wavelet coefficients, hidden state variables are linked across scales. Coefficients within scale are assumed to have the same probability density function, an approximation referred to as "tying" within scale. Figure 2D shows the hidden Markov Tree (HMT), which was first developed by Crouse et al. [14] and represents a graphical model of a set of dependencies between wavelet coefficients. The HMT, and similar models, have been used for a wide variety of signal and image processing applications including classification, segmentation, compression, synthesis and denoising [15], [16], [17], [18].

The parameters of the two state zero mean HMT consist of, (1) the probability mass function p_i^S describing the high/low variance state of the wavelet coefficients of the coarsest scale, (2) the state transition matrices describing persistence between adjacent scales and (3) the variances of the Gaussian mixture model, σ_L^2 and σ_H^2.

These parameters can be estimated using a modified Baum-Welch/upward-downward expectation maximization algorithm [19]. For the data shown in Figure 3 we find that after training the transition matrices across scale are not uniformly distributed between states. This confirms the modeling assumption of the persistence of large or small coefficients across scale.

We follow the denoising methods outlined in [14]. Assume that the relationship between the wavelet coefficients of the observed EEG signal w_i^k, clean signal y_i^k, and noise n_i^k may be described as,

$$w_i^k = y_i^k + n_i^k \,, \qquad (4)$$

where the subscript i indicates the ith wavelet coefficien-

t and the superscript k indicates it comes from the kth wavelet tree. Assuming independent Gaussian white noise, we can estimate the state dependent clean signal variance $\sigma_{i,m}^2$ from the noisy signal variance $\gamma_{i,m}^2$,

$$\sigma_{i,m}^2 = (\gamma_{i,m}^2 - \sigma_n^2)_+ \,, \qquad (5)$$

where m indicates the state and $(x)_+$ indicates a rectification allowing for only zero and positive values of its argument.[2] In the results we present, an estimate of noise variance is derived from the variance of the finest scale wavelet coefficients. However this estimate can be adjusted to use only a specific subspace of the wavelet basis, depending upon prior knowledge of the noise source or interfering signal.

An estimate of the clean signal wavelet coefficients y_i^k, given the noisy wavelet coefficients and state variables is given by,

$$E[Y_i^k | W_i^k = w_i^k, S_i^k = m] = \frac{\sigma_{i,m}^2}{\sigma_n^2 + \sigma_{i,m}^2} w_i^k \,. \qquad (6)$$

Marginalizing over the state variable S_i^k results in

$$E[y_i^k | \mathbf{w}^k] = \sum_m p(S_i^k = m | \mathbf{w}^k) \times \frac{\sigma_{i,m}^2}{\sigma_n^2 + \sigma_{i,m}^2} w_i^k \,, \qquad (7)$$

where $\mathbf{w}^k$ is a vector representing all coefficients in the kth tree. Given these estimates for the clean signal wavelet coefficients, the inverse wavelet transform is applied to reconstruct the signal.

Examples of original and HMT denoised EEG signals, after eye blink removal, are demonstrated in Figure 3 for a 10s segment of data. For the single-trial analysis the HMT denoising is applied to a short segment of data of about 500 ms around the event of interest (button push). The HMT parameters are estimated for each channel from a single event at the beginning of the experiment.

C. ERN detection with linear discriminant analysis

We have previously reported [8] off-line single trial detection of ERN, in a time window 0-100ms following the button-push response, with an accuracy of $A_z = 0.79 \pm 0.05$ using a linear classifier on 64 electrodes.

In addition to the preprocessing methods just described, we have looked to improve detection accuracy by adding a second detection interval of 100ms following the conventional time window of 0-100 ms after the response. The input to the linear classifier is a combination of two time windows of 64 leads (128 inputs). To improve classification and reduce the effect of high frequency noise, we bootstrap the data by selecting multiple samples within a single trial as training data. At 250Hz sampling rate the 200ms following the response results in 25 training samples, each of dimension 128.

Our on-line detector is a Gaussian classifier and can be adapted by simply updating the mean and covariance estimates for the two classes using the most recent sample.

---

[2]This rectification can cause deviation from linearity. However, in practice it is not a factor since the difference is almost always positive.
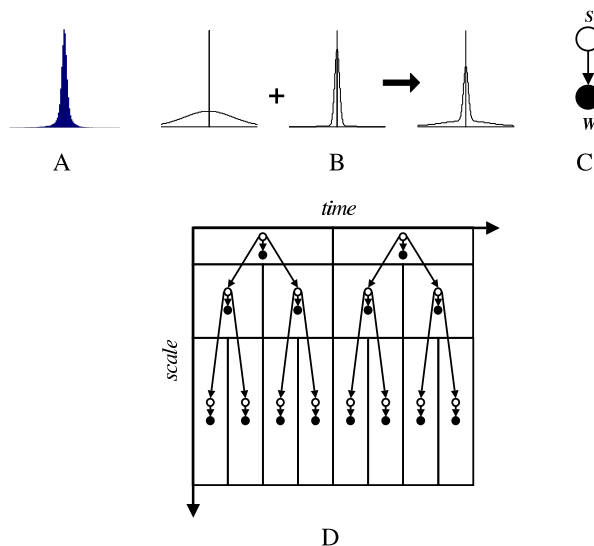
A   B   C

D

Fig. 2. Overview of the HMT. (A) Marginal distribution of evoked response EEG wavelet coefficients showing large peak and long tails (kurtosis=17.5). (B) Two state Gaussian mixture model for the marginal distribution of evoked response EEG. (C) Graphical representation of state dependencies on observed wavelet coefficients (white nodes are hidden variables, black nodes are observed). (D) hidden Markov tree indicating the dependencies between scales.
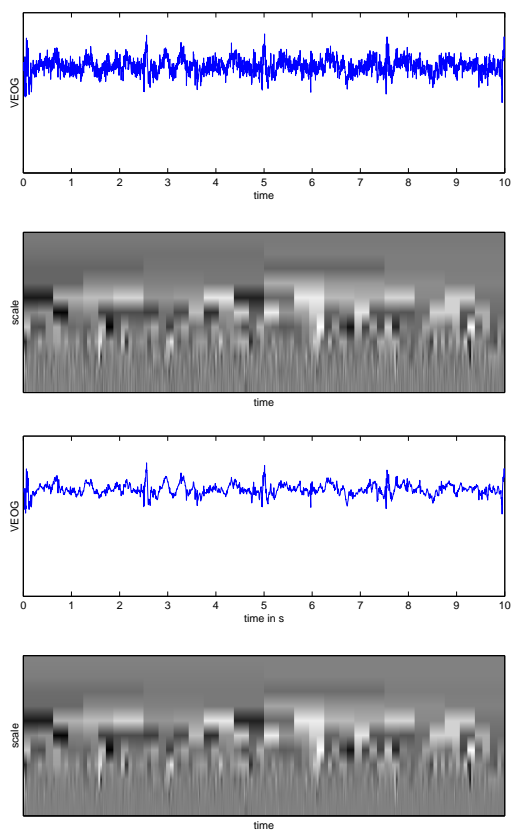

Fig. 3. Results of single-trial denoising using the HMT. (Top) original (noisy) signal after eye blink removal. (Bottom) clean signal denoised using HMT. Shown are the signals and wavelet decompositions.
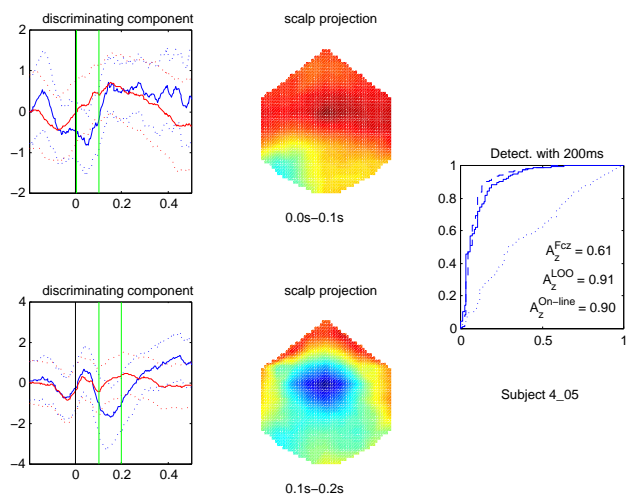

Fig. 4. ERN detection using linear discrimination combining the time intervals 0-100ms and 100-200ms after response. (left) Discriminating component and (center) scalp projection graphs results were obtained with off-line linear classification. Similar results are obtained with on-line adaptation.(right) Single-trial ROC results compare A_z for off-line (LOO), on-line and using only F_{cz} electrode. Note that on-line and off-line are comparable.

With equal covariance for the two classes (errors/corrects) this results in linear classification. The advantage of linear classification is that we can compute the coupling of the discriminating source activity with the sensors. This provides a spatial map of the origin of the discriminating activity. The results obtained for a typical subject are shown in Figure 4. The previously described fronto-central negativity is observed during the 100ms following the response. In addition, a more prolonged bilateral posterior positivity is observed for correct trials, which further improves discrimination. The single-trial discrimination performance for the 7 subjects analyzed in [8] increases to $A_z = 0.90 \pm 0.04$ when using off-line linear preprocessing. Note the processing sequence: eye blink removal, followed by HMT denoising, followed by linear classification. The largest jump in performance was due to the eye blink removal (A_z increases from 0.79 to 0.89). HMT denoising increased A_z from 0.89 to 0.90.The performance for on-line processing[3] is $A_z = 0.91 \pm 0.03$, where performance is measured by predicting the classification of the next sample using only past data.

III. CORRECTING ERRORS USING THE DETECTED ERN

Using the detected ERN to correct human response errors requires choosing a threshold on the output of the linear classifier so as to minimize the number of errors, E_{HM}, made by the combined human-machine system. This is the number of errors, E_H, made by the subject alone, minus the number of those errors detected by our BCI system, *plus* the number of correct responses by the subject that were incorrectly classified as errors by our BCI system. Denote by f_{TE} the fraction of subject errors detected by the BCI system (True Errors), and by f_{FE} the frac-

---

[3]The on-line version does not yet include the HMT denoising.

tion of correct subject responses incorrectly classified as errors by our BCI system (False Errors). We can write $E_{HM} = E_H(1 - f_{TE}) + C_H f_{FE}$. However, $1 - f_{TE} = f_{FC}$, the fraction of subject errors that were incorrectly classified as correct responses (False Correct). The number of human-machine errors is therefore $E_H f_{FC} + C_H f_{FE}$, which is the total number of classifier errors.

If we construct a classifier that estimates the class probability, we minimize the number of errors by assigning new trials to the class with the highest probability under the classifier. In terms of the log-odds-ratio, we minimize the error rate by choosing a threshold of zero. Since we employ Gaussian models with equal covariance matrices for the class distributions, the log-odds ratio is a linear function of the input. However, if the actual class distributions do not fit our model distributions, our linear discriminator is only an approximation to the true log-odds ratio, and the optimal threshold may differ from zero. This appears to be the case for this set of experiments, as we see in Figure 5.

In our on-line implementation, we choose an optimal threshold by keeping a record of the outputs of the linear discriminator for the last one hundred trials of both correct and erroneous responses. We then search for the threshold that minimizes the error rate on these sets, i.e., $E_H f_{FC} + C_H f_{FE}$. We estimate f_{FC} and f_{FE} from the outputs we have saved, and keep count of E_H and C_H from the trials we have seen thus far. The threshold is applied to a new sample that is not part of this training set. Table I summarizes the results for each subject. An experiment consists of about 600 button push responses of which 5-20% are incorrect. The middle column show the error rate before correction. The average relative reduction in error rate for 7 subjects under on-line processing is $21.4\% \pm 21.7$. Note the large variance indicates very large improvements for some subjects and no improvement or slight degradations in performance for other subjects.

In a practical applications the error correction system will not have access to the correct response. The parameters for the linear classifier and the detection threshold must be derived from an initial training sequence and are kept constant during operation. The performance numbers reported here reflect this scenario in that we report classification performance on the current trial using a classifier that is trained only on information from previous trials.

IV. CONCLUSION

We have described a set of linear preprocessing and classification algorithms for providing an accurate single-trial estimate of the ERN, an EEG signal which has been found to correlate with perceived error. The goal of measuring the ERN is to monitor a subject's task specific error rate and adapt an HCI to maximize overall performance. We have shown initial results for both off-line and on-line correction of subject errors for an alternative forced choice visual discrimination task. Future work will investigate generalization of the approach to other tasks, particularly those which require more complex adaption of the HCI.
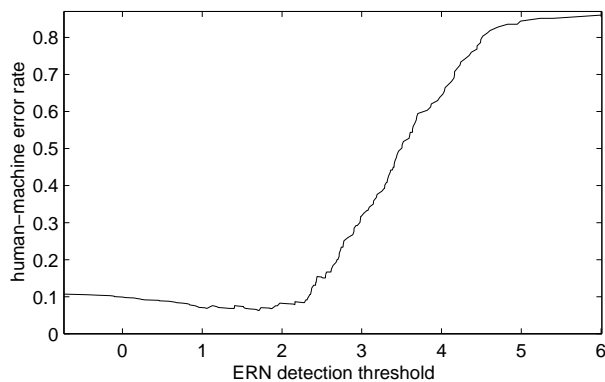

Fig. 5. Total human-machine performance after correcting human response errors based on ERN detection as a function of classifier threshold. Note that the minimum is not at zero, indicating that the underlying distributions deviate from Gaussian.

TABLE I
SUMMARY OF ON-LINE ERROR CORRECTION FOR EACH OF 7 SUBJECTS

Subject ID	original error rate in %	% error reduction[a]
3_27	6	23
3_30	10	-6
4_03	15	-1
4_04	15	49
4_05	13	27
4_09	14	47
4_10	18	12

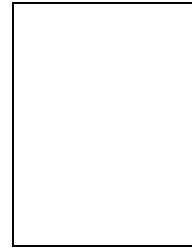[a]Negative values indicate degradation in performance.

V. ACKNOWLEDGMENTS

REFERENCES

[1] A.F. Kramer and J. Spinks, "Capacity views of information processing," in *Psychophysiology of human information processing: An integration of central and autonomic nervous system approaches*, R. Jennings and M. Coles, Eds. 1991, pp. 179–250, New York: Wiley.

[2] J. Beatty, "Task-evoked pupillary responses, processing load, and the structure of processing resources," *Psychological Bulletin*, pp. 276–292, 1982.

[3] A. Gevins, M.E. Smith, L.K. McEvoy, H. Leong, and J. Le, "Electroencephalographic imaging of higher brain function," *Philosophical Transactions of the Royal Society, London, Series B*, vol. 354, pp. 1125–1134, 1999.

[4] C. Tallon-Baudry, O. Bertrand, F. Peronnet, and J. Pernier, "Induced gamma-band activity during the delay of a visual short-term memory task in humans," *J. Neuroscience*, , no. 11, pp. 4244–4254, 1998.

[5] F. Babiloni e.t al., "Linear classification of low-resolution EEG patterns produced by imagined hand movements," *IEEE Trans. Rehab. Eng.*, pp. 186–188, 2000.

[6] W.H.R. Miltner, C.H. Braun, and M.G.H Coles, "Event-related brain potentials following incorrect feedback in a time estimation task: evidence for a generic neural system for error detection," *J. Cog. Neurosci*, pp. 788–798, 1997.

[7] Thorpe S., D. Fize, and C. Marlot, "Speed of processing in the human viusal system," *Nature*, pp. 520–522, 1996.

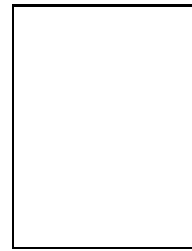[8] L. Parra, C. Alvino, A.C. Tang, B.A. Pearlmutter, N. Yeung,

A. Osman, and P. Sajda, "Linear spatial integration for single trial detection in encephalography," *NeuroImage*, 2002, To appear.

[9] M. Falkenstein, J. Hoorman, S. Christ, and J. Hohnsbein, "ERP components on reaction errors and their functional significance: A tutorial," *Biological Psychology*, vol. 52, pp. 87–107, 2000.

[10] G. Schalk, J. Wolpaw, D. McFarland, and G. Pfurtscheller, "EEG-based communication: presence of an error potential," *Clinical Neurophysiology*, vol. 111, pp. 2138– 2144, 2000.

[11] P. Sajda, A. Gerson, and L. Parra, "High-throughput image search via single-trial event detection in a rapid serial visual presentation task," in *1st International IEEE EMBS Conference on Neural Engineering*, Capri, Italy, March 2003.

[12] J.C. Woestenburg, M.N Verbaten, and J.L. Slangen, "The removal of the eye-movement artefact from the EEG by regression analysis in the frequency domain," *Biological Psychology*, vol. 16, no. 1-2, pp. 127–147, 1983.

[13] T.-P. Jung, S. Makeig, C. Humphries, T.-W. Lee, M.J. Mckeown, V. Iragui, and Sejnowski T.J., "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, pp. 163–178, 2000.

[14] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 886–902, 1998.

[15] H. Coi and R. G. Baraniuk, "Multiscale imaage segmentation using wavelet-domain hidden markov models," *IEEE Transactions on Image Processing*, vol. 10, no. 9, pp. 1309–1321, 2001.

[16] H. Cheng and C. A. Bouman, "Multiscale bayesian segmentation using a trainable context model," *IEEE Transactions on Image Processing*, vol. 10, no. 4, pp. 511–525, Apr. 2001.

[17] J. K. Romberg, H. Coi, and R. G. Baraniuk, "Bayesian tree-stuctured image modeling using wavelet domain hidden markov models," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1056–1068, 2001.

[18] C.D. Spence, L. Parra, and P. Sajda, "Detection, synthesis and compression in mammographic image analysis using a hierarchical image probability model," in *Mathematical Methods in Biomedical Image Analysis*, M. Staib, Ed. 2001, pp. 3–10, IEEE.

[19] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–285, 1989.

Clay D. Spence (A'93) received the B.S. (cum laude), M.S. and Ph.D. degrees in theoretical physics from the University of California, Irvine, in 1979, 1981, and 1987, respectively. He subsequently joined the Technical Staff at the Sarnoff Corporation, Princeton, N-J, where he has modeled the auditory nervous system of the barn owl, developed algorithms for noise suppression in video, studied the effects of noise and noise suppression on video compression, and developed algorithms for object recognition, medical image analysis, and blind source separation.

Adam D. Gerson received the B.S.E. degree in bioengineering in 1997 from the University of Pennsylvania and the M.S. degree in biomedical engineering in 2002 from Columbia University where he is working toward the Ph.D. degree. His current research interests include machine learning, computational neuroscience, computer vision, and pattern recognition.

Paul Sajda received his B.S. in Electrical Engineering from MIT (1989) and his M.S. and Ph.D. in Bioengineering from the University of Pennsylvania (1992, 1994). In 1994 he joined the David Sarnoff Research Center where he went on to become the Head of the Adaptive Image and Signal Processing Group. He is currently an Associate Professor of Biomedical Engineering and Radiology at Columbia University where he is Director of the Laboratory for Intelligent Imaging and NeuroComputing (LIINC). His research focuses on neural engineering, neuroimaging, computational neural modeling and machine learning applied to image analysis. He has received several awards for his research including an NSF CAREER Award (2002), Sarnoff Technical Achievement Award (1996), The Pollack Award for outstanding dissertation research in Bioengineering (UPenn, 1994), The Flexner Award for outstanding thesis research in the Neurosciences (UPenn, 1993) and the Adler Award for outstanding undergraduate research in Electrical Engineering (MIT, 1989). He holds 5 US patents.
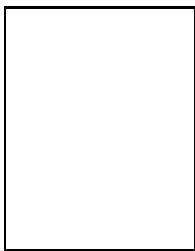
Lucas C. Parra received his Diploma in Physics in 1992, and Doctorate in Physics in 1996 from the Maximilian-Ludwig-Universiaet, Munich, Germany. From 1992 to 1995 he worked at the Neural Networks Group at Siemens Central Research in Munich, Germany and at the Machine Learning Department at Siemens Corporate Research (SCR) in Princeton, NJ. During 1995-1997 he was member of the Imaging Department at SCR and worked on medical image processing and novel reconstruction algorithms for nuclear medicine. He joined Sarnoff Corporation in Princeton, NJ in 1997 where he is Technology Leader for the Adaptive Image and Signal Processing group. Currently he is also Adjunct Professor of Biomedical Engineering at Columbia University. His areas of expertise include machine learning, statistical pattern recognition, acoustic array processing, medical imaging, biomedical signal processing and encephalography. His research concentrates on probabilistic models of natural signals.