Early Detection of Numerical Typing Errors Using Data Mining Techniques

Shouyi Wang, *Student Member, IEEE*, Cheng-Jhe Lin, Changxu Wu, and Wanpracha Art Chaovalitwongse, *Senior Member, IEEE*

***Abstract*—This paper studies the applications of data mining techniques in early detection of numerical typing errors by human operators through a quantitative analysis of multichannel electroencephalogram (EEG) recordings. Three feature extraction techniques were developed to capture temporal, morphological, and time–frequency (wavelet) characteristics of EEG data. Two most commonly used data mining techniques, namely, linear discriminant analysis (LDA) and support vector machine (SVM), were employed to classify EEG samples associated with correct and erroneous keystrokes. The leave-one-error-pattern-out and leave-one-subject-out cross-validation methods were designed to evaluate the in- and cross-subject classification performances, respectively. For the in-subject classification, the best testing performance had a sensitivity of 62.20% and a specificity of 51.68%, which were achieved by SVM using morphological features. For the cross-subject classification, the best testing performance was achieved by LDA using temporal features, based on which it had a sensitivity of 68.72% and a specificity of 49.45%. In addition, the receiver operating characteristic (ROC) analysis revealed that the averaged values of the area under ROC curves of LDA and SVM for the in- and cross-subject classifications were both greater than 0.60 using the EEG 300 ms prior to the keystrokes. The classification results of this study indicated that the EEG patterns of erroneous keystrokes might be different from those of the correct ones. As a result, it may be possible to predict erroneous keystrokes prior to error occurrence. The classification problem addressed in this study is extremely challenging due to the very limited number of erroneous keystrokes made by each subject and the complex spatiotemporal characteristics of the EEG data. However, the outcome of this study is quite encouraging, and it is promising to develop a prospective early detection system for erroneous keystrokes based on brain-wave signals.**

***Index Terms*—Early detection, electroencephalography (EEG) classification, mental state monitoring, typing errors.**

Manuscript received March 29, 2010; revised October 11, 2010; accepted December 11, 2010. Date of publication April 21, 2011; date of current version October 19, 2011. This paper was recommended by Associate Editor J. M. Carmena.

S. Wang is with the Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ 08854 USA (e-mail: shouyi@rci.rutgers.edu).

C.-J. Lin and C. Wu are with Department of Industrial and System Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260 USA (e-mail: cl224@buffalo.edu; changxu@buffalo.edu).

W. A. Chaovalitwongse is with the Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ 08854 USA. He is also with the Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544 USA, and also with the Department of Industrial and Systems Engineering, University of Washington, Seattle, WA 98195-2650 USA (e-mail: wchaoval@rci.rutgers.edu; wchaoval@princeton.edu; artchao@uw.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCA.2011.2116006

I. Introduction

NUMEROUS types of electronic devices with alphabetical or numerical keyboards have become very important tools in modern times. An erroneous keystroke can be easily caused by many reasons, such as the operators' inexperience, fatigue, and carelessness. At the present time, many typing error correction systems have been developed for computer users. For example, current word processing software such as Microsoft Word provides automatic spelling checks as well as automated corrective actions. Some other methods have also been developed to detect and remove errors due to overlapped keystrokes [47]. It is noted that most of the automatic typing error detection systems are designed for text typing; very few studies have focused on detecting numerical typing errors. In fact, numerical typing is as a common task as text typing in practice [41], [42]. In particular, in some crucial tasks, numerical typing errors may result in serious consequences or accidents. For instance, numerical typing errors in medical records may result in inaccurate diagnoses and/or drug administrations. In financial transactions, numerical errors may cause significant losses at the stock exchanges. In aviation control, incorrect numerical inputs may lead to serious air traffic accidents [51].

Human typing involves intricate interactions of concurrent perceptual and cognitive processes [40]. Numerous studies of typing behaviors have been conducted to explore their underlying cognitive mechanisms in the past decade. However, most studies in the literature were in the field of transcription (text) typing. Error correction of numerical data is much more difficult and challenging because there is no pattern database to look up. The operator cannot visualize and identify if there are errors in the data because there is no contextual information for verification. This is very common in hear-and-type tasks such as telling a phone number to a phone representative, bank account number to a teller, and tracking number of a parcel to a customer service agent. The operators are very susceptible to making errors when they receive auditory inputs while typing.

Although double data entry (DDE) and read-aloud (RA) [22] methods are commonly used to assure the data quality, these methods are very tedious and inefficient. Moreover, in reality, to avoid typing errors, each data entry may accompany a confirmative action, such as pressing an ENTER key, before which input data can be checked and corrected. However, such a mechanism does not always exist. For example, a selection menu may be coded by numbers, and pressing a number already commands the execution. In this kind of situations, afterward, detection mechanism is too late to reverse the outcome. Particularly in a crucial task, such numerical errors may result in serious or even life-threatening consequences. As a result, when afterward checking/confirmation is limited or even impossible, predicting and avoiding numerical typing errors are becoming very critical to assure data quality. Unfortunately, to the best of our knowledge, there are no effective tools available to assist human operators in this task. If numerical typing errors can be detected in advance, the detection can be integrated in an error prevention system for many crucial typing works.

Erroneous keystrokes are possibly caused by an operator's psychophysiological state such as a lack of attention, external distractions, and fatigue. In our previous study, we have successfully built a computational model, called the queuing-network-model human processor, to establish mathematical representations of cognitive functions of typing behaviors. The results of our study on brain modeling and human factor analysis of erroneous keystrokes suggested that the brain activity before erroneous keystrokes might be different from that of the correct ones [24], [49]. The goal of this paper is to develop an early detection system of erroneous keystrokes. We employ feature extraction and data mining techniques to perform quantitative analysis of electroencephalogram (EEG) recordings prior to keystrokes. Although there have been numerous EEG studies in various fields, very few studies in the literature have been conducted to investigate the early detection (or prediction) of typing behaviors based on EEG data. The characterization of the underlying EEG patterns before someone is about to make an error is still in a great need of further investigation. The development of an effective method to classify erroneous keystrokes based on their (generating) mechanisms remains a difficult but worth-pursuing task.

The rest of this paper is organized as follows. In Section II, the research background and previous related work are discussed, including the background of data entry correction methods, error-related EEG potentials, and the data mining techniques for quantitative EEG analysis. In Section III, the proposed EEG feature extraction techniques and the employed classification techniques, namely, linear discriminant analysis (LDA) and support vector machine (SVM), are described. Section IV presents the design of human experiments and computational data analysis. The computational results of the classification systems are provided in Section V. Finally, the concluding remarks and future work are given in Section VI.

II. Background

A. Data Entry Correction Methods

A number of studies have been performed to develop correction methods for numerical typing errors. Scholtus [41], [42] developed an algorithm for automatic correction of typing errors in numerical data. However, this algorithm can be only applied to some systematic typing errors, such as checking the inconsistencies when there are mathematical relations between the data digits. Kawado *et al.* [22] compared the efficiencies of the two commonly used data verification methods: DDE and RA. In the DDE method, the DDE was performed by either an identical or a different operator. In the RA method, one operator read the typed data on a printed sheet or computer screen aloud, and another operator compared the data (that were) heard with the data (that were) recorded to confirm whether they are the same. The error detection rates were 59.5% and 69.0% for the RA and DDE methods, respectively. Their results surprisingly showed that there might still be a large portion of undetected errors even after the two commonly used data verification methods were applied. Their study also indicated that it is very hard to achieve full accuracy for large amount of numerical data input even when data verification methods are used to the data management. As mentioned in Arndt *et al.* [5], the databases of large projects may contain a great absolute number of mistakes in data collection and thus have data quality problems. They investigated the types and frequencies of data errors in 688 forms from seven sites in a multicenter field trial. It was found that 2.4% of the received data had errors even though conscientious efforts had been made in checking and correcting the data.

B. ErrPs

Typing behaviors involve complex interactions of concurrent perceptual and cognitive processes [40]. The brain-wave activity measured by EEG is often an essential and natural way to study the brain activity during typing. The event-related potentials (ERPs) in response to a perceptual, cognitive, or motor event have been extensively studied in neuroscience and brain–computer interfaces (BCIs) [19]. Since the early 1990s, many studies have found that a subject's recognition of response errors is often associated with some specific error-related EEG potentials [14], [30]. More recently, the work of Ferrez and Millán [36] has shown that the error-related potential (ErrP) of a BCI can be reliably recognized. The pioneering work in ErrP detection provided a prelude for us to explore the underlying mechanisms of erroneous keystrokes. It should be noted that current ErrP studies mostly focused on the EEGs after response errors, and the EEGs prior to errors were much less studied. However, EEGs prior to errors are of great importance to prevent errors from occurring, particularly in some crucial typing tasks mentioned in the introduction part. Therefore, this study particularly focused on early error detection using EEGs prior to keystrokes.

C. Data Mining in EEG: Feature Extraction

Over the past decade, numerous studies have been performed to apply quantitative signal processing methods and time series techniques to analyze the characteristics of EEG data. The simplest feature extraction can be obtained by downsampling an EEG signal from its usually high sampling rate (such as 1000 Hz) into a low-frequency range of particular interest (such as 0–30 Hz). The resulting features are supposed to be a representative of the temporal characteristics of EEG data in this low-frequency band [7]. Another common univariate feature extraction method uses the morphological characteristics of EEG data, such as curve length [32], zero crossings [38], number of peaks [48], nonlinear energy [2], etc. In addition, grounded in signal processing techniques, some more complex EEG feature extraction techniques have also been developed. Traditional linear methods include frequency and power spectrum analysis and the parametric modeling of EEG time series (e.g., autoregressive, moving average, and autoregressive moving average models). Although widely used in EEG analysis, these methods actually treat EEG as statistically stationary signals. To deal with nonstationarity in EEG, various methods based on time–frequency analysis have been developed. The most well-known time–frequency technique is called wavelet transform (WT), which is capable of providing a representation of nonstationary EEG signals in both time and frequency domains accurately.

D. Data Mining in EEG: Classification

Over the past decade, there were increasing interests in using classification techniques to discriminate different brain activities based on EEG recordings. Numerous data mining techniques have been proposed to EEG classification. Those methods include decision trees [35], neural networks [44], association rule induction [13], K-nearest-neighbor method [10], and genetic algorithms [31]. There have been many studies suggesting that EEG signals at different mental states or in different mental tasks may be classifiable [3], [4], [6], [29], [43].

The error detection task in this paper is, in principle, a binary classification problem of correct and erroneous EEG samples. LDA and SVM are two popular classification techniques for binary classification tasks. Both of them construct a hyperplane to separate data into two subsets based on optimization theories. Parra *et al.* [18] used LDA to detect response errors for seven subjects in a forced choice visual discrimination task. Using 64 EEG electrodes and two time windows of 100 ms, they were able to reach an accuracy of 79% on average. Blankertz *et al.* [7] adopted the sparse Fisher discriminant to differentiate an index finger movement from a small finger movement in a self-paced key typing task. Using EEG data 120 ms prior to keystrokes, they achieved overall classification accuracies of 96.7% and 93.6% for filtered and nonfiltered EEG data, respectively. SVMs have also been widely applied to a large number of EEG classification problems [17], [50]. Our group successfully applied data mining techniques to classify normal and abnormal (epileptic) brain activities based on EEG recordings of a number of epileptic patients [8]–[10]. Garrett *et al.* [18] applied both LDA and SVM to classify EEG signals in five mental tasks. The averaged classification accuracies of LDA and SVM were 66% and 72%, respectively.

III. Methods

A. Feature Extraction

We employed three feature extraction techniques to capture the characteristics of EEG signals. They were temporal, morphological, and wavelet features. For an EEG epoch with n channels, we first extracted features from each channel and then concatenated the features of all the n channels to construct the feature vector of this multichannel EEG epoch. Let $X = \{x_1, x_2, \ldots, x_m\}$ denote a single-channel EEG with m sampling points, and the extractions of the temporal, morphological, and wavelet features of X are described as follows.

Temporal Features: The temporal features can be obtained by downsampling of EEG signals. The downsampling of EEG data reduces the amount of data that needs to be analyzed while it is still capable of capturing patterns for slow brain activity [7]. Since the most common EEG patterns (e.g., alpha, beta, delta, and theta wave patterns) contain frequency elements that are mainly below 30 Hz, we downsampled the EEG data from 1000 to 30 Hz in this study. In particular, the downsampling was accomplished by calculating the means of consecutive nonoverlapping intervals of every 33 points. For example, a 100-ms EEG epoch of 1000 Hz has 100 points in each channel, then three temporal features are extracted for each channel of EEG through the downsampling process.

Morphological Features: Seven morphological features were extracted from each channel of EEG. These features were based on the features previously described in Wong *et al.* [48]. A brief description of the morphological features is given in the following.

1) Curve length: This feature is also known as "line length," which was first proposed by Olsen *et al.* in [32]. Curve length is the sum of the distances between successive points, given by

$$\sum_{i=1}^{m-1} |x_{i+1} - x_i|. \tag{1}$$

Since the curve length increases as the signal magnitude or frequency increases, it can be used to measure the amplitude-frequency variations of the EEG signals. It has been used in many EEG studies, such as epileptic seizure detection [11] and stimulation responses of the brain [12].

2) Standard deviation: It is among the most widely used measures of signal variability. It indicates how all the points of the signal are clustered around the mean. The standard deviation can be obtained by

$$\sqrt{\frac{\sum_{i=1}^{m}(x_i - \bar{X})^2}{m-1}} \tag{2}$$

where $\bar{X}$ is the mean of the single-channel EEG X.

3) Number of peaks: The number of peaks per second is a commonly used characteristic to measure the overall frequency of EEG signals. The number of peaks in the single-channel EEG X can be calculated by

$$\frac{1}{2}\sum_{i=1}^{m-2} \max\left\{0, \operatorname{sgn}(x_{i+2} - x_{i+1}) - \operatorname{sgn}(x_{i+1} - x_i)\right\}. \tag{3}$$

4) Root-mean-square (rms) amplitude: RMS is one of the most commonly used methods to determine the power changes of a signal [28], particularly for complex waveforms, such as EEG signals. The rms amplitude of the single-channel EEG X is defined as

$$\sqrt{\frac{\sum_{1}^{m} x_i^2}{m}}. \tag{4}$$

5) Average nonlinear energy: Nonlinear energy was first proposed by Kaiser [20]. It is a measure of signal energy that is proportional to both signal amplitude and frequency. It has been found that the nonlinear energy is sensitive to spectral changes. Thus, it is also useful to capture the spectral information of an EEG signal [2]. The average nonlinear energy of the single-channel EEG X is computed as

$$\frac{1}{m-2}\sum_{i=2}^{m-1} x_i^2 - x_{i-1}x_{i+1}. \tag{5}$$

6) Zero crossings: The frequency information of EEG signals can also be estimated by the number of times that its value crosses the zero axis. Zero-crossing feature extraction has been applied in many signal processing and pattern recognition tasks, including EEG signal analysis [38]. The zero crossings of the single-channel EEG X can be mathematically defined as

$$\frac{1}{2}\sum_{i=1}^{m-1}\left|\operatorname{sgn}(x_{i+1})-\operatorname{sgn}(x_i)\right|. \tag{6}$$

7) Variance-to-range ratio: This feature calculates the ratio of the variance to the magnitude range of the EEG signal. It takes into account both variation and range of EEG magnitudes. The ratio of the single-channel EEG X is given by

$$\frac{\sum_{i=1}^{m}(x_i-\bar{X})^2}{(m-1)(X_{\max}-X_{\min})} \tag{7}$$

where $X_{\max}$ and $X_{\min}$ are the maximum and minimum values of X, respectively.

Time–Frequency Features: WT was employed to analyze the time–frequency characteristics of the EEG signals. The basic idea of wavelet analysis is to express a signal as a linear combination of a particular set of functions obtained by shifting and dilating one single function called mother wavelet. The WT of the signal $X(t)$ is defined as

$$C(a,b)=\int_R X(t)\frac{1}{\sqrt{a}}\Psi\left(\frac{t-b}{a}\right)dt \tag{8}$$

where Ψ is the mother wavelet, $C(a,b)$ denotes the WT coefficients of the signal $X(t)$, a is the scale parameter, and b is the shifting parameter. Continuous WT (CWT) has $a \in R^+$ and $b \in R$; and discrete WT (DWT) has $a = 2^j$ and $b = k2^j$ for all $(j,k) \in Z$ given the decomposition level of j. Analyzing the signal by CWT at every possible scale a and shifting b requires substantially more computations than the DWT. As a result, the DWT with dyadic scaling and shifting is often employed in many studies to decompose EEG signals into different frequency subbands [39]. The coefficients of DWT decomposition provide a nonredundant and highly efficient representation of a signal in both time and frequency domains. At each level of decomposition, DWT works as filters to divide the signal into two bands called approximations and detail signals. The approximations (A) are the low-frequency components of the signal, and the details (D) are the high-frequency components. For more detailed mathematical formulations of WT, refer to [1].

Among different wavelet families, Daubechies wavelets are well known for their orthogonality property and efficient filter implementation, and the db4 is frequently used in EEG analysis [46]. In this paper, we applied the typical db4 to decompose EEG signals into eight levels. Table I shows the frequency bands of different levels of DWT decomposition. Since the frequency band of EEG signals is often considered to be less than 30 Hz, we employed the coefficients of levels A7, D7, D6, and D5, which roughly correspond to the commonly recognized delta, theta, alpha, and beta brain waves, respectively. The other four levels of signals were considered as the high-frequency background noises and thus were eliminated in the wavelet feature vector. Moreover, to further decrease the feature dimensionality for classification, the statistics of the DWT coefficients were extracted. They are the mean, standard deviation, maximum, and minimum of the wavelet coefficients of the four used levels. By doing so, each channel of EEG can be represented by a $4 \times 4 = 16$ dimensional feature vector, and an EEG epoch of n channels can be represented by a $16n$-dimensional feature vector.

TABLE I
FREQUENCY RANGES AND THE CORRESPONDING BRAIN-WAVE BANDS OF THE EIGHT LEVELS OF SIGNALS BY DISCRETE WAVELET DECOMPOSITION

Decomposed Signal	Frequency Range (Hz)	Approximate Band
D1	250-500	-
D2	125-250	-
D3	62.5-125	-
D4	31.3-62.5	-
D5	15.7-31.3	Beta
D6	7.9-15.7	Alpha
D7	4.0-7.9	Theta
A7	0-4.0	Delta

B. Classification Methods

Let Y denote the $n \times k$ dimensional feature vector for a multichannel EEG epoch, where n is the number of channels and k is the number of features of each single channel of EEG. In this paper, $n = 36$ and $k = 3, 7,$ and 16 for temporal, morphological, and wavelet features, respectively. Let l denote the class label of the EEG epoch, for which $l = 1$ denotes a correct EEG sample and $l = -1$ means an erroneous EEG sample. Given $p+q$ training samples (Y_i, l_i), $i = 1,\ldots,p+q$, the data set of p correct EEG epochs is denoted by $D_1 = \{(Y_1,l_1),(Y_2,l_2),\ldots,(Y_p,l_p)\}$, and the data set of q erroneous epochs is denoted by $D_2 = \{(Y_{p+1},l_{p+1}),(Y_{p+2},l_{p+2}),\ldots,(Y_{p+q},l_{p+q})\}$. The difference between them is that the optimal decision boundary is determined based on different optimization theories, which will be briefly discussed in the following.

Fisher's LDA: Fisher's LDA aims to find an optimal projection by minimizing the intraclass variance and maximizing the distance between the two classes simultaneously [16]. Mathematically, LDA tries to find an optimal direction $\omega^* \in R^{n\times k}$ as a solution of the following optimization problem:

$$\omega^* = \arg\max_{\omega}\frac{\omega^{\mathrm{T}}S_b\omega}{\omega^{\mathrm{T}}S_\omega\omega} \tag{9}$$

where ω is the direction of the hyperplane that is used to separate the two data sets. S_b and S_ω are the interclass and intraclass covariance matrices, respectively. They are defined as follows:

$$S_b = (m_1-m_2)^{\mathrm{T}}(m_1-m_2) \tag{10}$$

$$S_\omega = \sum_{i\in 1,2}\sum_{i\in D_i}(Y_i-m_i)^{\mathrm{T}}(Y_i-m_i) \tag{11}$$

where m_1 and m_2 are the means of the feature vectors Y in the two data sets D_1 and D_2, respectively. They can be calculated by

$$m_1 = \frac{1}{p} \sum_{Y \in D_1} Y = \frac{1}{p} \sum_{i=1}^{p} Y_i \tag{12}$$

$$m_2 = \frac{1}{q} \sum_{Y \in D_2} Y = \frac{1}{p} \sum_{i=p+q}^{p+1} Y_i. \tag{13}$$

When S_ω is not singular, the aforementioned optimization problem can be solved by applying the eigen-decomposition to the matrix $S_\omega^{-1} S_b$. The eigenvector corresponding to the largest eigenvalue forms the optimal direction w^* by

$$\omega^* = S_\omega^{-1}(m_1 - m_2). \tag{14}$$

When S_ω is singular, an identity matrix with a small scalar multiple can be used to tackle this problem [27]. The optimal w^* then becomes

$$\omega^* = (S_\omega + \lambda I)^{-1}(m_1 - m_2). \tag{15}$$

Once ω^* is obtained, the optimal decision boundary of LDA can be represented by

$$\omega^{*\mathrm{T}} Y + b = 0 \tag{16}$$

where b is the bias term. There is no general rule to determine the bias term; the most commonly used bias term is $b = -\omega^{*\mathrm{T}}(m_1 + m_2)/2$. The class of an EEG epoch Y depends on which side of the hyperplane that its feature vector is on. In particular, for a new EEG epoch represented by a feature vector Y_{new}, then the prediction rule is as follows:

$$\begin{cases} \omega^{*\mathrm{T}} Y_{\mathrm{new}} + b > 0, & l_{\mathrm{new}} = 1 \text{ (an erroneous keystroke)} \\ \omega^{*\mathrm{T}} Y_{\mathrm{new}} + b < 0, & l_{\mathrm{new}} = -1 \text{ (a correct keystroke).} \end{cases}$$

SVM: SVMs are another group of binary classification tools, which have been successfully applied in many EEG classification problems [7], [18], [21], [23], [37]. The fundamental problem of SVM is to build an optimal decision boundary to separate two categories of data. In the data sets of EEG epochs D_1 and D_2, each EEG epoch is represented by an $n \times k$ dimensional feature vector. One can actually find infinitely many hyperplanes in $R^{n \times k}$ to separate the two data groups. Based on the statistics learning theory, an SVM selects a hyperplane which maximizes its distance from the closest point from the samples. This distance is referred to as *margin*. The standard SVM formulation that maximizes the *margin* and minimizes the training error is as follows:

$$\min_{\omega,\xi,b} \left\{ \frac{1}{2}\|\omega\|^2 + C \sum_{i=1}^{p+q} \xi_i : D(Y^{\mathrm{T}}\omega + be) \geq e - \xi_i \right\} \tag{17}$$

where ω is the weight vector, e is a vector of ones, b is an offset parameter, and the slack variables ξ are introduced to measure the degree of misclassification during training. The parameter $b/\|\omega\|$ determines the offset of the hyperplane from the origin along the weight vector ω. The penalty cost C is used to control the tradeoff between a large *margin* and a small prediction error penalty. Each column of Y is an observation Y_i, and D is a diagonal matrix with class-label elements $D_i i$ equal to 1 if Y_i belongs to one class or -1 if otherwise. Vector e has all its elements equal to one. The first term of the objective function in (17) is due to maximizing the *margin* of separation $2/\|w\|$, and the second term measures how much emphasis is given to the minimization of the training error.

Since the standard SVM classifiers usually require a large amount of computation time for training, the proximal SVM (PSVM) algorithm was introduced by Mangasarian and Wild [26] as a fast alternative to the standard SVM formulation. The formulation for the linear PSVM is as follows:

$$\min_{\omega,\xi,b} \left\{ \frac{1}{2}\left(\|\omega\|^2 + b^2\right) + \frac{1}{2} C \xi_i{}^{\mathrm{T}} \xi_i : D(Y^{\mathrm{T}}\omega + be) = e - \xi_i \right\} \tag{18}$$

where the traditional SVM inequality constraint is replaced by an equality constraint. This modification changes the nature of the support hyperplanes ($\omega^{\mathrm{T}} Y + b = \pm 1$). Instead of bounding planes, the hyperplanes of PSVM can be thought of as "proximal" planes, around which the points of each class are clustered and which are pushed as far apart as possible by the term $(\|\omega\|^2 + b^2)$ in the aforementioned objective function. It has been shown that PSVM has comparable classification performance to that of standard SVM classifiers, but can be an order of magnitude faster [26]. Therefore, we employed PSVM in this study.

IV. Typing Experiment

A. Experimental Design

The experimental task was a typical hear-and-type task which emulated daily work done by bank tellers or representatives in customer services. A computer program read out 30 random numbers of nine digits in a trial, and the subjects were told to type out those numbers. The numbers were not linguistically grouped, i.e., every digit was read out separately without chunking two or three digits (e.g., read 123 as "one two three" instead of "one twenty three" or "one hundred twenty three"). In addition, there was a small pause (300 ms) in between every three digits. The numbers were read out this way because, based on an observation and interview by the author in a pilot study, this was the most natural way to read out numbers without any specific format known beforehand. The interval between two digits is 750 ms on average. A short pause of 2.5 s existed after each nine-digit number, during which the subjects would be reminded of pressing the enter key.

Nine subjects were recruited from the student body of University at Buffalo. All subjects were native speakers of English without any hearing disability. Before the experiment, each subject was pretested on his/her typing skill to assure his/her familiarity with typing. The subjects were allowed to adjust the volume, posture, and other settings of typing environment to his/her preference. Each subject was given two practice trials prior to formal experimental trials. If a subject did not show any inability in the hear-and-type task, she or he was then allowed to continue eight trials of hear-and-type tasks. During each trial, the EEG data of each subject were recorded. A 5-min break was given to subjects after four trials so that their EEG would not be influenced by long exposure to a relatively boring task.

TABLE II
TYPING PERFORMANCE OF EACH SUBJECT

subject	# Keystrokes	# Erroneous Keystrokes	Percentage of Erroneous Keystrokes
1	2122	36	1.70%
2	2113	51	2.41%
3	2419	69	2.85%
4	2401	10	0.42%
5	2422	60	2.48%
6	2117	76	3.59%
7	2420	45	1.88%
8	2405	63	2.62%
9	2134	54	2.53%

The descriptive statistics of the typing performance of the nine subjects are summarized in Table II. No significant difference was found in terms of age, accuracy, or typing speed between male and female subjects. Hence, male and female subjects can be regarded as a homogeneous group. The percentage of erroneous keystrokes was ranged from 0.42% on subject 4 to 3.59% on subject 7. The latency between auditory stimuli and keystrokes was 728 ms on average.

B. EEG Acquisition and Preprocessing

During the experiment, EEG data were collected with an EEG cap containing 40 Ag/AgCl electrodes according to the international 10–20 system. There are four electrodes that were used for measuring eye movements to remove muscular artifacts. The rest 36 electrodes were mounted on the scalp and thus used for analyses in this paper. The placement of the 36 scalp electrodes is shown in Fig. 1. The signals were amplified by NuAmps Express system (Neuroscan Inc., USA) and sampled at 1000 Hz. The typed number, as well as the timing of each keystroke, was recorded simultaneously by the system. After comparing with the reference number, each keystroke was labeled as either "correct" or "erroneous" by using "1" and "-1," respectively. The raw EEG data were first processed by a 0.1–30-Hz bandpass filter [34]. Then, the EEG epochs were extracted from the filtered database on the keystroke events recorded during typing. The length of each EEG epoch was set to 500 ms before a keystroke according to the minimal interval between two successive keystrokes. The flowchart of the EEG acquisition and the epoch sampling is shown in the upper part in Fig. 2.

C. Classification Procedure

As shown in Fig. 2, each 500-ms EEG epoch was divided into five nonoverlapping subepochs with equal length of 100 ms. The size of subepochs was chosen empirically with the goal of obtaining salient information of the brain activity prior to keystrokes. The temporal, morphological, and wavelet features of each 100-ms EEG epoch were extracted. The feature vector of a multichannel EEG epoch was constructed by concatenating the feature vectors of all the channels. For example, if we want to classify the 100-ms EEG epochs based on temporal features, then each epoch was represented by a $3 \times 36 = 108$ dimensional feature vector. Similarly, the morphological feature dimension for an EEG epoch is $7 \times 36 = 252$, and the wavelet feature vector has a dimension of $16 \times 36 = 576$.

D. Evaluation Metric of a Single Prediction

Sensitivity and specificity are commonly used performance measures of binary classification tests. For example, people are tested for a disease in a clinic study. Sensitivity is defined as the proportion of actual positives which are correctly identified as positive, and specificity is the proportion of negatives which are correctly identified as negative. In this paper, we labeled the erroneous EEG samples as positive and the correct EEG samples as negative. Then, we use sensitivity to measure the percentage of erroneous EEG samples that are correctly identified as positive and specificity to measure the percentage of correct EEG samples that are correctly identified as negative. For each testing EEG sample, the classification result can be always categorized into one of the following four subsets:

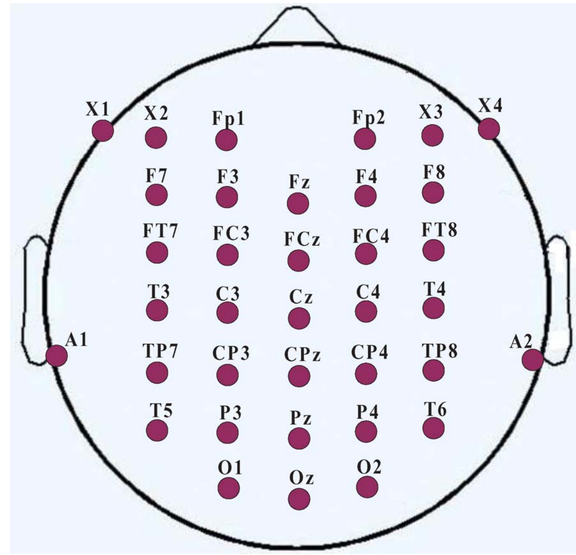

Fig. 1. Allocations of the 36 scalp electrodes.

1) true positive (TP): if an erroneous EEG epoch is classified as positive;
2) false positive (FP): if a correct EEG epoch is classified as positive;
3) true negative (TN): if a correct EEG epoch is classified as negative;
4) false negative (FN): if an erroneous EEG epoch is classified as negative.

Then, sensitivity and specificity can be calculated as follows:

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{19}$$

$$\text{specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}}. \tag{20}$$

E. Training and Evaluation

A standard classification problem generally follows a two-step procedure which consists of training and testing phases. During the training phase, a classifier is trained to achieve the optimal separation for the training data set. Then, in the testing phase, the trained classifier is used to discriminate new samples with unknown class information. The leave-one-out cross-validation is an attractive method of model evaluation, and it is capable of providing almost unbiased estimate of the generalization ability of a classifier [45]. In this paper, we trained and tested the classifiers under two frameworks, namely, in- and cross-subject error detections. Correspondingly, two leave-one-out cross-validation methods with

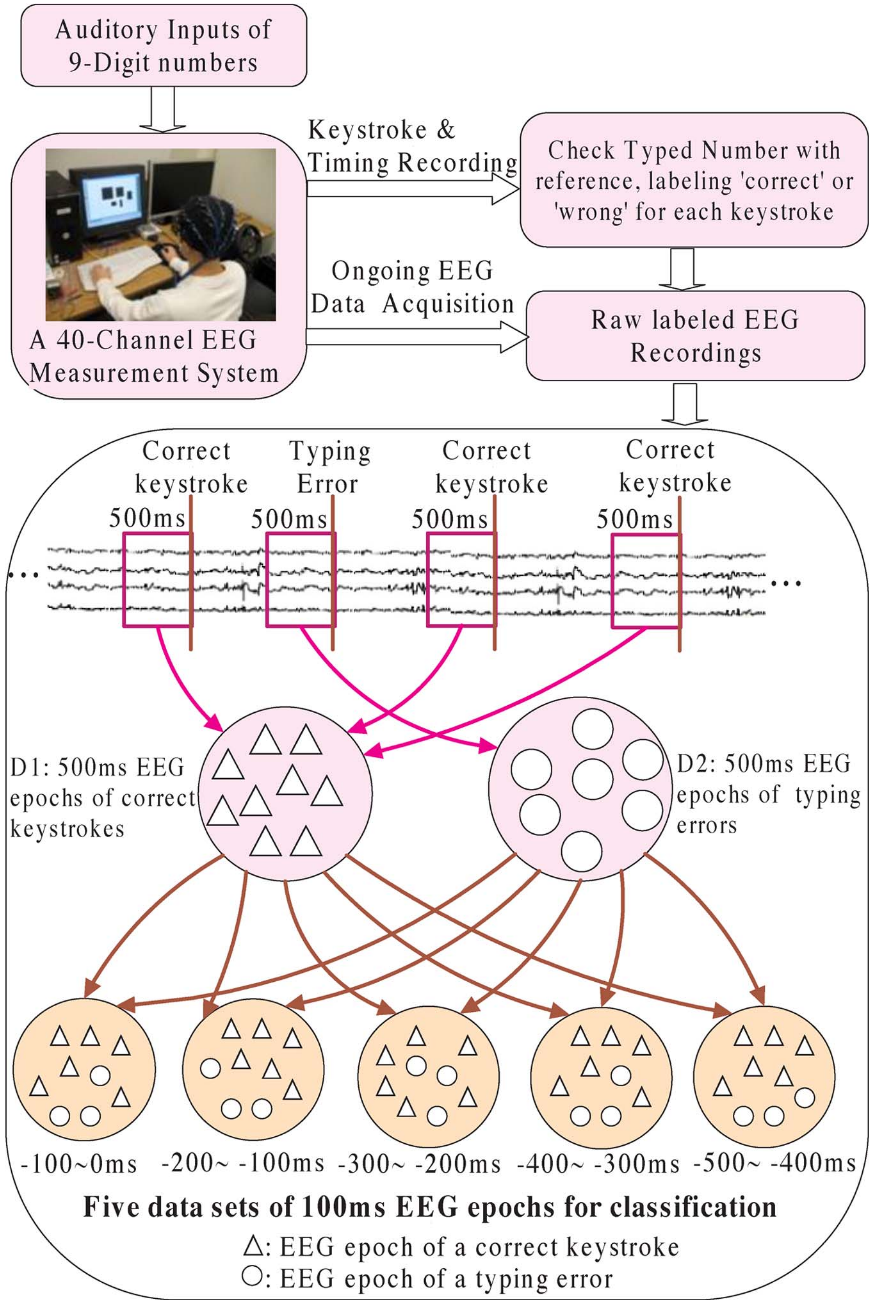

Fig. 2. Flowchart of the typing experiment, as well as the EEG acquisition and epoch sampling procedure. The 500-ms EEG epochs were first extracted from the raw EEG data, and then, they were divided into five 100-ms subepochs corresponding to the five nonoverlapping time intervals prior to keystrokes.

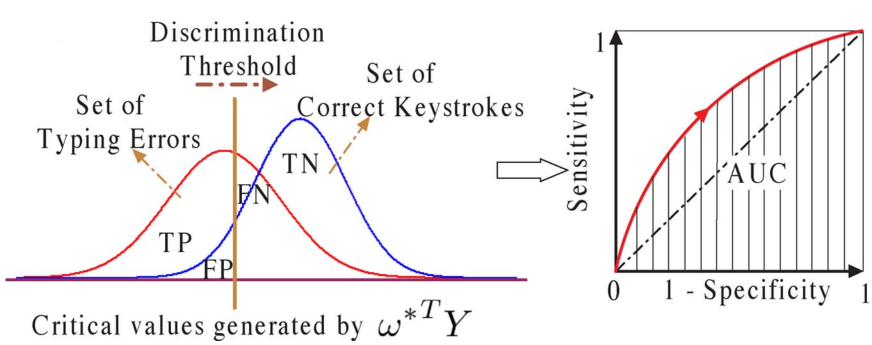

Fig. 3. Demonstration of ROC as the discrimination threshold of a classifier (LDA or PSVM) is varied through the whole range of its possible values. The value of AUC indicates the overall performance of a classifier. It may also indicate the classifiability of the two data sets without knowing the distributions of the two data sets based on the current classification framework.

perturbed duplications of erroneous samples were designed to achieve an unbiased estimate of the classification performance. The two methods are described in the following.

1) In-subject error detection: training and testing on each subject individually. We employed a leave-one-error-pattern-out cross-validation method for each subject. Let n_c and n_e denote the numbers of correct and erroneous keystrokes of a subject. Each time, we picked one erroneous EEG sample and $\lceil n_c/n_e \rceil$ correct samples out and trained the classifier by the rest of the samples. To eliminate the unbalanced problem during training, we employed an oversampling method with perturbed replications of erroneous samples. Let n_c^t and n_e^t denote the numbers of correct and erroneous samples in the training data set $(n_c^t \gg n_e^t)$. Then, the feature vector of each erroneous sample was replicated $\lceil n_c^t/n_e^t \rceil$ times. For each replication, a synthetic erroneous feature vector was generated by adding a random perturbation to the original erroneous feature vector. In particular, let Y_j be a feature vector of an erroneous EEG sample; then, a synthetic erroneous feature vector Y_j' can be created by

$$Y_j' = Y_j + \alpha \times \frac{\bar{Y}_e - Y_j}{\tau_e} \tag{21}$$

where $\bar{Y}_e$ and τ_e are the $1 \times 36k$ vectors, which contain the means and standard deviations of the $36k$ features for all the erroneous samples in the training data set and α is a random number that is uniformly generated in $[-1, 1]$. The trained classifiers were tested on the left-out samples. Repeat the procedure for all the erroneous samples of a subject. The averaged prediction result was used to indicate the classification effectiveness based on the current data set.

2) Cross-subject error detection: In this framework, we speculate that the erroneous EEG patterns of different subjects may share some common characteristics due to a high level of uncertainty or anxiety prior to making typing errors. There have been a number of recent BCI studies focusing on subject-independent ERP classification. The studies showed that the EEG potentials of different subjects may exhibit similar waveform characteristics in performing the same mental task [15], [25]. Stemmed from this consideration, we designed a leave-one-subject-out method to train and evaluate the classifiers. Each time, we picked one subject out and trained the classifiers by the EEG samples from the rest eight subjects. The oversampling method with perturbed replications of erroneous samples was also used to form a balanced training data set. The EEG samples of the left-out subject were considered as unknown samples to test the trained classifiers. Repeating this procedure for all the subjects, the averaged prediction accuracy can be used to indicate the effectiveness of the trained classification models.

F. ROC Analysis

Receiver operating characteristic (ROC) analysis is another popular method to evaluate the performance of a prediction model. An ROC curve is a plot of sensitivity versus false-alarm rate (1-specificity) as the discriminant threshold of a classifier varied throughout its possible ranges. The ROC curve for a perfect prediction model is the line connecting [0, 0] to [0, 1] and [0, 1] to [1, 1]. Moreover, the diagonal line connecting [0, 0] to [1, 1] is the ROC curve corresponding to a random model. Generally, an ROC curve lies between these two extreme lines. The *area under the ROC curve* (AUC) is often used as an important metric to evaluate a prediction model. The AUC is an overall summary of prediction accuracy across the spectrum of its decision-making values. AUC values are usually between 0.5 and 1. The AUC of a perfect predictor is 1 while a purely random chance model has an AUC of 0.5 on average. The higher the AUC value is to one, the better the prediction power a predictor has. A typical generation procedure of the ROC curve for a classifier is shown in Fig. 3. One may also find that the value of AUC may also be a classificability index of the two data sets without knowing their exact distributions based on the current classification framework.

V. Results

A. In-Subject Sensitivity and Specificity Analysis

Table III summarizes the in-subject training and testing sensitivity and specificity of LDA and PSVM based on the leave-one-error-pattern-out cross-validation methodology using the three choices of EEG features. The best training performance was achieved by the wavelet features for both LDA and PSVM. Using wavelet features, the training sensitivity and specificity were above 90% for both LDA and PSVM at all the five time intervals. It had an averaged training sensitivity of above 90% and an averaged specificity of above 80% when using morphological features. The temporal features had the worst training performance, which had an averaged training sensitivity of above 80% and an averaged specificity of above 70%. As for the testing performance, a noteworthy observation is that the best testing results of LDA and PSVM were both achieved at the time interval of $-100 \sim 0$ ms. In particular, the best testing performance of LDA was achieved at a sensitivity of 62.77% and a specificity of 51.03% when using morphological features, while the best testing performance of PSVM had a sensitivity of 62.20% and a specificity of 51.68% when using morphological features. In a contrast experiment, we also tested a randomized detection model with prior probability of error rate (RDPP). For a subject with an error rate of p, the RDPP classified each EEG sample as erroneous with a probability of p and as correct with a probability of $1 - p$. The testing results of the RDPP are shown in the last row in Table III. It was noted that only about 2% of the erroneous keystrokes can be detected on average by the RDPP, while both LDA and PSVM detected more than 60% of the erroneous keystrokes at a time interval of $-100 \sim 0$ ms. Our trained classification models considerably increased the error detection rate.

In addition, the averaged testing sensitivities of LDA and PSVM over the nine subjects and the three choices of features for the five time intervals are shown in Fig. 4. Interestingly, the error detection accuracies tended to increase as the time interval became closer to the timing of keystrokes, particularly at the last three 100-ms time intervals. This observation may indicate that, the closer the analyzed EEGs to the keystrokes, the more

TABLE III
IN-SUBJECT TRAINING AND TESTING RESULTS OF LDA, PSVM, AND A RANDOM MODEL BASED ON THE LEAVE-ONE-ERROR-PATTERN-OUT CROSS-VALIDATION METHODOLOGY (RESULTS WERE ALL AVERAGED OVER THE NINE SUBJECTS)

	Classifier	Feature	-500ms		-400ms		-300ms		-200ms		-100ms	
			sen.	spe.	sen.	spe.	sen.	spe.	sen.	spe.	sen.	spe.
Training Results	LDA	Temporal	87.11%	76.33%	87.64%	76.76%	88.26%	76.84%	85.07%	76.66%	89.63%	79.19%
		Morph.	95.08%	84.64%	95.62%	85.20%	95.28%	85.20%	96.41%	85.21%	95.71%	86.16%
		Wavelet	99.59%	93.67%	99.78%	94.43%	99.59%	93.28%	99.34%	93.64%	99.23%	93.71%
	PSVM	Temporal	87.16%	75.74%	87.33%	76.37%	87.03%	76.70%	86.90%	76.77%	87.40%	78.08%
		Morph.	93.73%	82.71%	95.57%	84.01%	93.22%	83.36%	96.08%	83.88%	95.89%	85.05%
		Wavelet	99.28%	92.42%	99.74%	92.64%	99.52%	92.60%	99.63%	93.45%	99.66%	93.68%
Testing Results	LDA	Temporal	56.03%	50.02%	57.47%	49.86%	54.88%	49.26%	61.28%	49.68%	61.74%	49.83%
		Morph.	54.48%	50.35%	55.93%	51.13%	58.81%	50.73%	55.17%	51.23%	**62.77**%	**51.03**%
		Wavelet	53.10%	49.72%	55.91%	50.41%	52.03%	50.85%	54.74%	51.97%	56.95%	49.70%
	PSVM	Temporal	55.62%	50.16%	57.83%	49.58%	55.89%	49.44%	59.88%	49.61%	63.15%	49.37%
		Morph.	52.87%	50.48%	55.82%	50.75%	60.13%	51.05%	58.92%	50.91%	**62.20**%	**51.68**%
		Wavelet	50.98%	50.49%	57.00%	50.28%	51.37%	50.63%	55.30%	51.64%	57.00%	50.52%
	RDPP	-	2.35%	97.73%	2.18%	97.74%	2.33%	97.72%	2.21%	97.72%	2.32%	97.73%

LDA: Linear Discriminant Analysis; PSVM: Proximal Support Vector Machine; RDPP: Randomized Detection Model Based on Prior Probability of Error Rate.

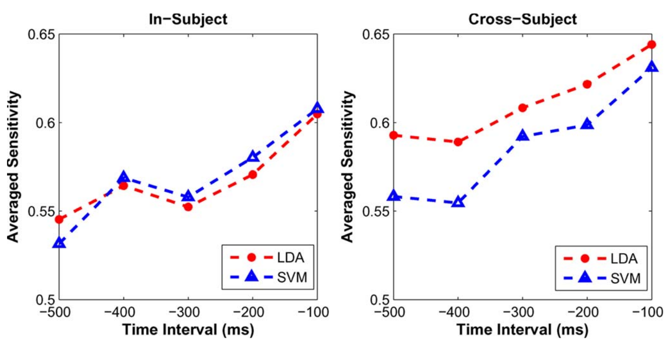

Fig. 4. Averaged testing sensitivity of LDA and PSVM over the nine subjects and the three choices of features for the five time intervals. In both in- and cross-subject experiments, there is an increasing trend of error detection accuracy as the time interval moves closer to the timing of keystrokes.

prominent brain-wave patterns can be captured to discriminate an upcoming erroneous keystroke from the correct ones. This result nicely matches with our physiological intuition and the previous study of Blankertz *et al.* in [7], which also reported increased classification accuracies in detecting upcoming finger movements (keystrokes) based on EEG recordings prior to the keystrokes. They also claimed that the most salient information of brain may be gained within 230 ms before the finger movements based on their experiments. However, this hypothesis still needs further investigation in future work.

B. Cross-Subject Sensitivity and Specificity Analysis

Table IV summarizes the cross-subject training and testing performance based on the leave-one-subject-out cross-validation methodology. It is noted that the cross-subject training performance was worse than the in-subject training performance. The best training performance of LDA has a sensitivity of 52.64% and a specificity of 86.00%, and that of PSVM was achieved at a sensitivity of 46.69% and a specificity of 93.06%. As for the testing performance, it is interesting to observe that the cross-subject testing performance was comparable to the in-subject testing performance. Moreover, it is worth mentioning that the best testing performance was achieved at a time interval of $-100 \sim 0$ ms for both LDA and PSVM. In particular, the best testing performance of LDA has a sensitivity of 68.72% and a specificity of 49.45%, and PSVM has a sensitivity of 66.63% and a specificity of 51.30% at best. These results indicate that the erroneous EEG at the time interval of $-100 \sim 0$ ms may exhibit more prominent patterns than the other four time intervals, which lead to increased classification accuracies. More importantly, the classifiability of erroneous and correct EEG samples across the subjects confirmed our hypothesis that different subjects may exhibit some similar EEG patterns prior to erroneous actions. Otherwise, the leave-one-subject-out method would produce an overall accuracy no better than a chance level. The subject-independent erroneous EEG potentials may be associated with a high level of uncertainty or anxiety prior to wrong response actions. Such uncertainty/anxiety-related EEG potentials may have much in common for human beings.

C. ROC Analysis

The ROC analysis is an important method to further investigate the classifiability of the erroneous and correct EEG samples. Tables V and VI present the in- and cross-subject AUC values of the nine subjects based their best choices of features. The corresponding in- and cross-subject ROC curves are shown in Figs. 5 and 6, respectively. From the ROC plots, one can observe that both in- and cross-subject ROC curves of the nine subjects are apparently deviated from the 45 $^\circ$ diagonal line which represents a random chance level, particularly the last three time intervals. These ROC curves suggested that the distribution of erroneous EEG patterns might be different from that of correct ones.

In addition, AUC is a convenient indicator of the discrimination between the two distributions of erroneous and correct EEG samples. As for the in-subject experiments, the best AUC value of LDA was 0.76, achieved at subject 4 using temporal features at $-200 \sim -100$ ms. The best AUC value of PSVM was 0.80, achieved also at subject 4 using temporal features at $-200 \sim -100$ ms. The best averaged AUC values were 0.63 and 0.64 for LDA and PSVM, respectively. They were both achieved at a time interval of $-100 \sim 0$ ms. In the cross-subject experiments, the best AUC values of LDA and PSVM were both 0.78, achieved at subject 4 using temporal features at a time interval of $-100 \sim 0$ ms. The best averaged cross-subject AUC

TABLE IV
CROSS-SUBJECT TRAINING AND TESTING RESULTS OF LDA, PSVM, AND A RANDOM MODEL BASED ON THE LEAVE-ONE-SUBJECT-OUT CROSS-VALIDATION METHODOLOGY (RESULTS WERE ALL AVERAGED OVER THE NINE SUBJECTS)

	Classifier	Feature	-500ms		-400ms		-300ms		-200ms		-100ms	
			sen.	spe.	sen.	spe.	sen.	spe.	sen.	spe.	sen.	spe.
Training Results	LDA	Temporal	51.67%	79.52%	52.57%	85.15%	52.64%	86.00%	54.89%	71.75%	55.04%	72.96%
		Morph.	68.09%	64.75%	65.59%	63.25%	69.06%	66.18%	68.04%	62.01%	70.00%	63.89%
		Wavelet	62.47%	69.66%	66.03%	68.38%	63.64%	64.73%	66.10%	68.34%	67.81%	71.06%
	PSVM	Temporal	46.69%	93.06%	47.48%	91.99%	48.83%	89.00%	52.09%	79.39%	56.06%	73.15%
		Morph.	66.46%	66.20%	66.40%	62.63%	67.44%	64.80%	66.04%	63.57%	69.25%	64.69%
		Wavelet	60.94%	74.19%	64.97%	72.66%	61.52%	69.18%	62.74%	69.87%	67.29%	71.98%
Testing Results	LDA	Temporal	63.39%	48.50%	63.98%	49.52%	59.75%	50.43%	64.17%	48.67%	**68.72%**	**49.45%**
		Morph.	55.74%	54.41%	55.53%	56.29%	60.39%	53.84%	60.84%	53.12%	60.88%	53.66%
		Wavelet	58.73%	49.97%	57.21%	50.88%	62.37%	48.83%	61.51%	50.21%	63.63%	51.84%
	PSVM	Temporal	54.85%	55.43%	54.65%	58.40%	55.08%	56.55%	61.21%	53.46%	**66.63%**	**51.30%**
		Morph.	54.40%	55.68%	54.25%	56.20%	60.35%	53.51%	56.87%	56.41%	60.61%	54.46%
		Wavelet	58.23%	50.97%	57.49%	52.79%	62.23%	48.93%	61.52%	51.38%	62.09%	52.86%
	RDPP	-	2.31%	97.74%	2.25%	97.74%	2.26%	97.74%	2.29%	97.74%	2.23%	97.74%

LDA: Linear Discriminant Analysis; PSVM: Support Vector Machine; RDPP: Randomized Detection Model Based on Prior Probability of Error Rate.

TABLE V
IN-SUBJECT AUC VALUES OF LDA AND PSVM BASED ON THE BEST CHOICE OF FEATURES

	sub.	-500ms		-400ms		-300ms		-200ms		-100ms	
		AUC	Feat.	AUC	Feat.	AUC	Feat.	AUC	Feat.	AUC	Feat.
LDA	1	0.56	Morp.	0.53	Temp.	0.62	Temp.	0.63	Wave.	0.68	Temp.
	2	0.59	Wave.	0.62	Morp.	0.55	Morp.	0.53	Morp.	0.58	Temp.
	3	0.59	Wave.	0.58	Temp.	0.59	Morp.	0.53	Morp.	0.64	Morp.
	4	0.6	Temp.	0.61	Morp.	0.75	Morp.	**0.76**	Temp.	0.75	Temp.
	5	0.61	Temp.	0.59	Wave.	0.58	Temp.	0.57	Temp.	0.61	Wave.
	6	0.68	Morp.	0.65	Temp.	0.66	Morp.	0.63	Morp.	0.64	Morp.
	7	0.61	Temp.	0.57	Morp.	0.62	Temp.	0.65	Morp.	0.61	Morp.
	8	0.57	Temp.	0.53	Temp.	0.6	Temp.	0.58	Wave.	0.54	Temp.
	9	0.59	Morp.	0.58	Morp.	0.62	Morp.	0.56	Morp.	0.62	Temp.
	ave.	**0.60**	-	**0.58**	-	**0.62**	-	**0.60**	-	**0.63**	-
PSVM	1	0.61	Morp.	0.53	Temp.	0.67	Temp.	0.65	Wave.	0.65	Temp.
	2	0.59	Wave.	0.67	Temp.	0.55	Temp.	0.55	Temp.	0.66	Morp.
	3	0.59	Wave.	0.58	Temp.	0.61	Morp.	0.59	Morp.	0.64	Morp.
	4	0.59	Temp.	0.59	Wave.	0.74	Morp.	**0.80**	Temp.	0.72	Temp.
	5	0.6	Temp.	0.57	Wave.	0.56	Temp.	0.54	Morp.	0.62	Temp.
	6	0.65	Morp.	0.63	Temp.	0.64	Morp.	0.67	Morp.	0.65	Morp.
	7	0.6	Temp.	0.54	Morp.	0.6	Temp.	0.63	Morp.	0.62	Wave.
	8	0.57	Temp.	0.54	Morp.	0.59	Temp.	0.61	Wave.	0.57	Wave.
	9	0.59	Morp.	0.6	Morp.	0.55	Temp.	0.57	Wave.	0.67	Morp.
	ave.	**0.60**	-	**0.58**	-	**0.61**	-	**0.62**	-	**0.64**	-

TABLE VI
CROSS-SUBJECT AUC VALUES OF LDA AND PSVM BASED ON THE BEST CHOICE OF FEATURES

	subject	-500ms		-400ms		-300ms		-200ms		-100ms	
		AUC	Feat.	AUC	Feat.	AUC	Feat.	AUC	Feat.	AUC	Feat.
LDA	1	0.56	Temp.	0.57	Temp.	0.57	Temp.	0.68	Wave.	0.55	Temp.
	2	0.53	Morp.	0.58	Temp.	0.56	Morp.	0.59	Morp.	0.57	Temp.
	3	0.6	Morp.	0.58	Wave.	0.58	Wave.	0.57	Temp.	0.59	Temp.
	4	0.67	Temp.	0.6	Temp.	0.73	Wave.	0.76	Temp.	**0.78**	Temp.
	5	0.6	Temp.	0.56	Morp.	0.57	Temp.	0.6	Temp.	0.64	Temp.
	6	0.64	Morp.	0.64	Morp.	0.64	Morp.	0.66	Morp.	0.63	Morp.
	7	0.59	Morp.	0.59	Temp.	0.62	Morp.	0.56	Wave.	0.61	Morp.
	8	0.55	Temp.	0.56	Temp.	0.57	Temp.	0.6	Temp.	0.58	Wave.
	9	0.57	Morp.	0.6	Morp.	0.61	Morp.	0.59	Morp.	0.57	Temp.
	ave.	**0.59**	-	**0.59**	-	**0.61**	-	**0.62**	-	**0.61**	-
PSVM	1	0.56	Wave.	0.57	Morp.	0.57	Temp.	0.68	Wave.	0.57	Wave.
	2	0.53	Morp.	0.59	Wave.	0.56	Morp.	0.59	Morp.	0.57	Temp.
	3	0.6	Morp.	0.58	Temp.	0.58	Wave.	0.57	Temp.	0.59	Temp.
	4	0.67	Temp.	0.64	Temp.	0.73	Wave.	0.75	Temp.	**0.78**	Temp.
	5	0.6	Temp.	0.56	Morp.	0.57	Temp.	0.6	Temp.	0.64	Temp.
	6	0.64	Morp.	0.64	Morp.	0.64	Morp.	0.66	Morp.	0.63	Morp.
	7	0.57	Morp.	0.6	Wave.	0.61	Morp.	0.56	Morp.	0.61	Morp.
	8	0.55	Temp.	0.56	Temp.	0.59	Temp.	0.6	Temp.	0.58	Wave.
	9	0.57	Morp.	0.6	Morp.	0.58	Morp.	0.6	Morp.	0.57	Temp.
	ave.	**0.59**	-	**0.59**	-	**0.60**	-	**0.62**	-	**0.62**	-

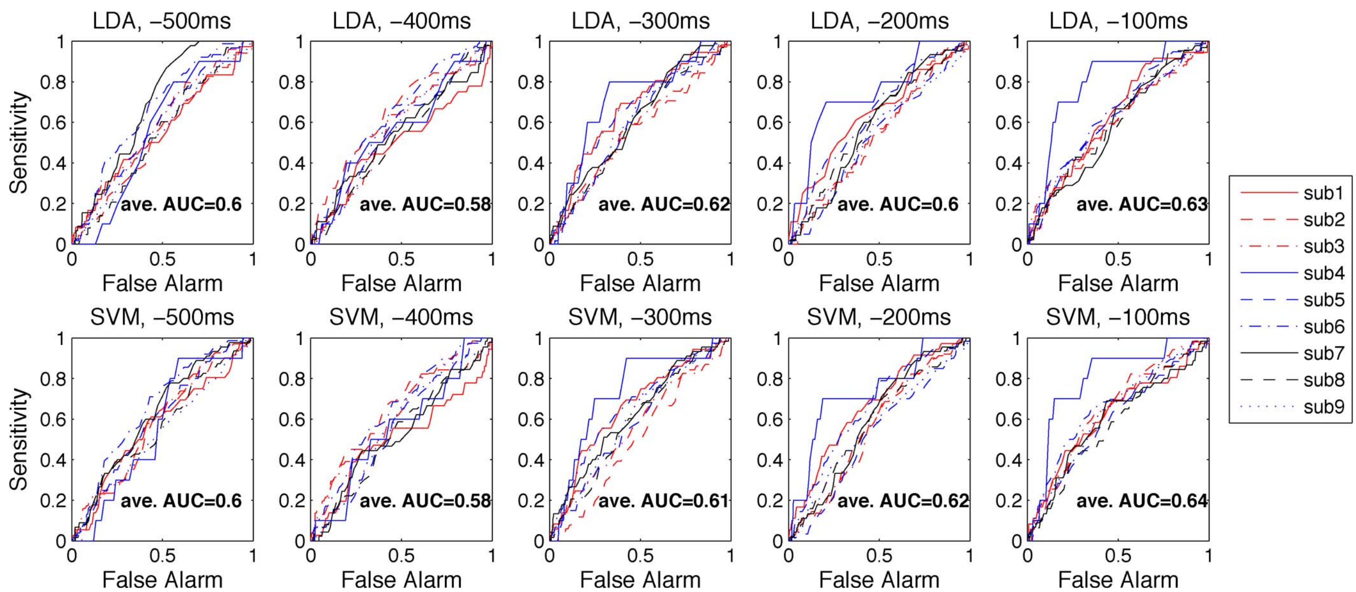

Fig. 5. In-subject ROC curves of the nine subjects at each time interval for LDA and PSVM based on their best choice of features. The averaged AUC value over the nine subjects is denoted in the bottom part of each subplot.

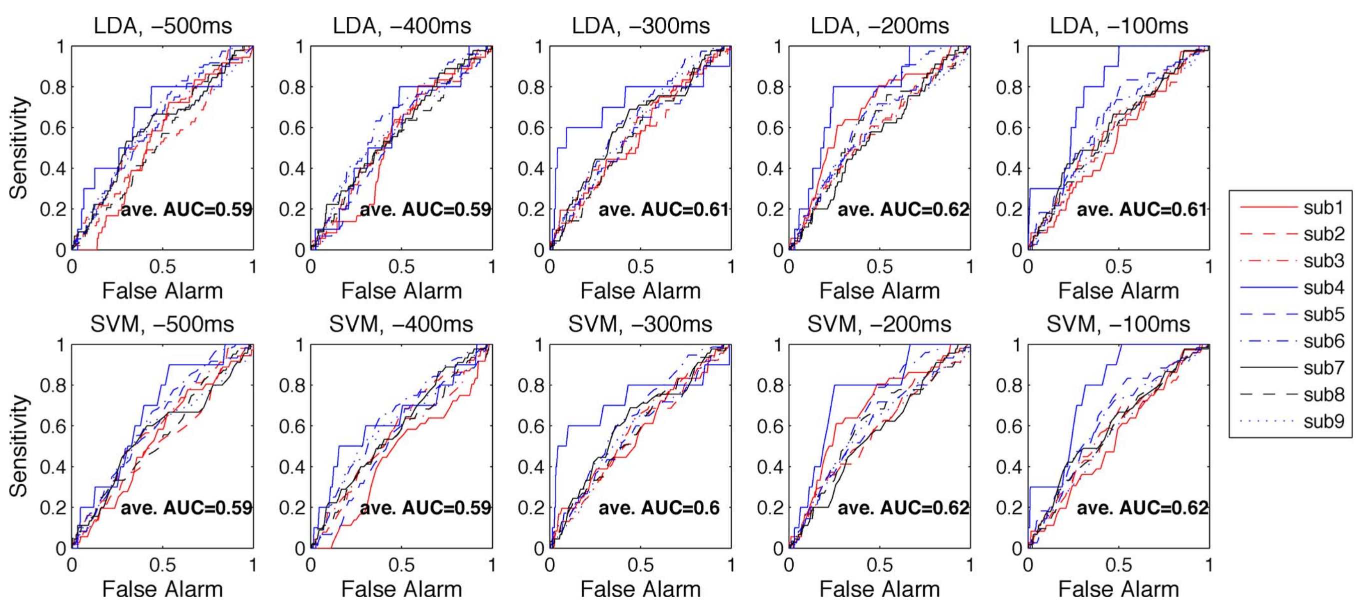

Fig. 6. Cross-subject ROC curves of the nine subjects of LDA and PSVM at each time interval based on their best choice of features. The averaged AUC value over the nine subjects is denoted in the bottom part of each subplot.

value of LDA was 0.62, achieved at $-200 \sim -100$ ms, and the best averaged AUC value of PSVM was also 0.62, achieved at both time intervals of $-200 \sim -100$ ms and $-100 \sim 0$ ms. It is noted that the averaged in- and cross-subject AUC values were all above 0.60 at the last three time intervals of $-300 \sim 200$, $-200 \sim 100$, and $-100 \sim 0$ ms. In addition, we notice that the classification accuracy on subject 4 was generally higher than those of the other nine subjects. When excluding subject 4, we still can get an averaged AUC values of around 0.60 at the last three time intervals. These results further confirmed our hypothesis that the most salient information of the brain activity associated with erroneous keystrokes may be gained within 300 ms prior to keystrokes. The AUC values indicated that the distributions of erroneous and correct EEG patterns might be different. As a result, erroneous keystrokes might be predictable based on EEG recordings.

VI. Conclusion and Discussion

Given that numerical typing errors are detectable in advance, some error prevention systems can be implemented (e.g., decimal-oriented error detection scheme [33]). For example, whenever there is an input digit being detected as a potential erroneous entry with relatively high chance, the error prevention system can halt the input and check the digit to see whether the check equation is satisfied or not. Without error detection beforehand, the system may have to check every digit and put the process on halt until the whole data set goes through. With error prediction, in contrast, maybe only half of the data need to be checked online, and then, around 80% of the errors might be detected before processing, saving much of idle time. The rest can still be processed offline by traditional data-checking mechanisms or methods, as described in the introduction part. Obviously, the error detection can provide some early filtering mechanisms to prevent erroneous data from being input into the system and thus improves system safety and performance, as well as reduces cost or loss related to typing errors.

In this paper, we have applied data mining techniques to investigate EEG patterns during numerical typing. The temporal, morphological, and wavelet-based time–frequency features were extracted. Popular data mining tools, namely, LDA and PSVM, were employed in this binary classification task. Since the number of erroneous EEG samples of each subject was too few to train the classifiers, we designed the in-subject leave-one-pattern-error-out and the cross-subject leave-one-subject-out cross-validation methodology to achieve an unbiased estimate of classification performance. The experimental results of this study are promising. The averaged in- and cross-subject AUC values were both above 0.60 at the last three time intervals of $-300 \sim 200$, $-200 \sim 100$, and $-100 \sim 0$ ms. These results indicated that the distribution of erroneous EEG patterns may be considerably different from that of the correct ones, particularly at the last 300 ms prior to keystrokes. The results are very encouraging considering that the classification problem of this study is extremely challenging due to the highly imbalanced data structure and that we only used a very simple and straightforward classification framework. This study confirmed our hypothesis that it is possible to proactively predict erroneous keystrokes in advance of error occurrence based on EEG recordings.

All our experiments were performed based on nine subjects. The number of subjects is limited due to difficulties in recruiting subjects and complex experimental settings. Although this study based on the limited data pool might not represent a generalized result for all people, the concept of automated early prediction of erroneous keystrokes seems to be conceivable based on the classification results of this study. It implied that data mining techniques, which have advanced classification and prediction capabilities, could facilitate detection of transient changes in brain dynamics from EEG recordings. To the best of our knowledge, this study is among the first attempts to investigate EEG-based brain-wave patterns for numerical typing errors. The results of this study may pave a new way of developing an early detection system of erroneous keystrokes to assist people in various typing tasks.

Finally, it is noted that we only focused on the early detection of typing errors in this study. As a result, only the EEG data prior to keystrokes were used. However, for less crucial tasks, we also believe that the EEG data a little (such as a few hundreds of milliseconds) after the keystrokes can also be very useful for typing error detection. It is still able to provide error warnings to the user within 1 s, and the user can conveniently correct the error right away easily. The using of data after keystrokes will be one of the future works in this study.

References

[1] N. Addison, *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine, and Finance*. New York: Taylor & Francis, 2002.

[2] R. Agarwal and J. Gotman, "Adaptive segmentation of electroencephalographic data using a nonlinear energy operator," in *Proc. IEEE Int. Symp. Circuits Syst.*, 1999, vol. 4, pp. 199–202.

[3] C. W. Anderson and Z. Sijercic, "Classification of EEG signals from four subjects during five mental tasks," in *Solving Eng. Problems with Neural Netw.: Proc. Int. Conf. Eng. Appl. Neural Netw.*, 1996, pp. 407–414.

[4] C. W. Anderson, E. A. Stolz, and S. Shamsunder, "Discriminating mental tasks using EEG represented by AR models," in *Proc. IEEE Eng. Med. Biol. Annu. Conf.*, 1995, pp. 875–876.

[5] S. Arndt, G. Tyrrell, R. F. Woolson, M. Flaum, and N. C. Andreasen, "Effects of errors in a multicenter medical study: Preventing misinterpreted data," *J. Psychiat. Res.*, vol. 28, no. 5, pp. 447–459, Sep./Oct. 1994.

[6] G. A. Barreto, R. A. Frota, and F. N. S. de Medeiros, "On the classification of mental tasks: A performance comparison of neural and statistical approaches," in *Proc. IEEE Workshop Mach. Learn. Signal Process.*, 2004, pp. 529–538.

[7] B. Blankertz, G. Curio, and K. R. Müller, "Classifying single trial EEG: Towards brain computer interfacing," *Adv. Neural Inf. Process. Syst.*, vol. 14, no. 2, pp. 157–164, Jan. 2002.

[8] W. Chaovalitwongse, Y. J. Fan, and R. C. Sachdeo, "Novel optimization models for abnormal brain activity classification," *Oper. Res.*, vol. 56, no. 6, pp. 1450–1460, Nov. 2008.

[9] W. Chaovalitwongse and P. Pardalos, "On the time series support vector machine using dynamic time warping kernel for brain activity classification," *Cybern. Syst. Anal.*, vol. 44, no. 1, pp. 125–138, Jan. 2008.

[10] W. A. Chaovalitwongse, Y. J. Fan, and R. C. Sachdeo, "On the time series k-nearest neighbor classification of abnormal brain activity," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 37, no. 6, pp. 1005–1016, Nov. 2007.

[11] R. Esteller, J. Echauz, T. Cheng, B. Litt, and B. Pless, "Line length: An efficient feature for seizure onset detection," in *Proc. 23rd Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2001, vol. 2, pp. 1707–1710.

[12] R. Esteller, J. Echauz, and T. Tcheng, "Comparison of line length feature before and after brain electrical stimulation in epileptic patients," in *Proc. 26th Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2004, pp. 4710–4713.

[13] T. P. Exarchos, A. T. Tzallas, D. I. Fotiadis, S. Konitsiotis, and S. Giannopoulos, "EEG transient event detection and classification using association rules," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 3, pp. 451–457, Jul. 2006.

[14] M. Falkenstein, J. Hoormann, S. Christ, and J. Hohnsbein, "ERP components on reaction errors and their functional significance: A tutorial," *Biol. Psychol.*, vol. 51, no. 2/3, pp. 87–107, Jan. 2000.

[15] S. Fazli, F. Popescu, M. Danoczy, B. Blankertz, K. R. Muller, and C. Grozea, "Subject-independent mental state classification in single trials," *Neural networks*, vol. 22, no. 9, pp. 1305–1312, Nov. 2009.

[16] K. Fukunaga, *Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1990.

[17] G. N. Garcia, T. Ebrahimi, and J. M. Vesin, "Support vector EEG classification in the Fourier and time–frequency correlation domains," in *Conf. Proc. 1st Int. IEEE EMBS Conf. Neural Eng.*, 2003, pp. 591–594.

[18] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda, "Response error correction-a demonstration of improved human-machine performance using real-time EEG monitoring," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 173–177, 2003.

[19] T. C. Handy, *Event-Related Potentials: A Methods Handbook*. Cambridge, MA: MIT Press, 2004.

[20] J. F. Kaiser, "On a simple algorithm to calculate the energy of a signal," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 1990, vol. 1, pp. 381–384.

[21] M. Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner, and H. Ritter, "BCI competition 2003-data set IIb: Support vector machines for the P300 speller paradigm," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1073–1076, Jun. 2004.

[22] M. Kawado, S. Hinotsu, Y. Matsuyama, T. Yamaguchi, S. Hashimoto, and Y. Ohashi, "A comparison of error detection rates between the reading aloud method and the double data entry method," *Control. Clin. Trials*, vol. 24, no. 5, pp. 560–569, Oct. 2003.

[23] T. N. Lal, T. Hinterberger, G. Widman, M. Schröer, J. Hill, W. Rosenstiel, C. E. Elger, B. Schökopf, and N. Birbaumer, "Methods towards invasive human brain computer interfaces," in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2005, pp. 737–744.

[24] C.-J. Lin and C. Wu, "Detecting typing errors in a numerical typing task with linear discriminant analysis of single trial EEG," *Human Factors Ergonomics Soc. Annu. Meet. Proc.*, vol. 53, no. 10, pp. 595–599, 2009.

[25] S. Lu, C. Guan, and H. Zhang, "Subject-independent brain computer interface through boosting," in *Proc. 19th Int. Conf. Pattern Recognit.*, Tampa, FL, Dec. 2008, pp. 1–4.

[26] O. L. Mangasarian and E. W. Wild, "Proximal support vector machine classifiers," in *Proc. Knowl. Discov. Data Mining*, 2001, pp. 77–86.

[27] S. Mika, "Kernel Fisher discriminants," M.S. thesis, Dept. Computer Sci., Univ. Technol., Berlin, Germany, 2002.

[28] A. Moran, I. Bar-Gad, H. Bergman, and Z. Israel, "Real-time refinement of subthalamic nucleus targeting using Bayesian decision-making on the root mean square measure," *Movement Disorders*, vol. 21, no. 9, pp. 1425–1431, Sep. 2006.

[29] K. Natarajan, R. Acharya, F. Alias, T. Tiboleng, and S. K. Puthusserypady, "Nonlinear analysis of EEG signals at different mental states," *Biomed. Eng. Online*, vol. 3, p. 7, Mar. 2004.

[30] S. Nieuwenhuis, K. Ridderinkhof, J. Blom, G. Band, and A. Kok, "Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task," *Psychophysiology*, vol. 38, no. 5, pp. 752–760, Sep. 2001.

[31] H. Ocak, "Optimal classification of epileptic seizures in EEG using wavelet analysis and genetic algorithm," *Signal Process.*, vol. 88, no. 7, pp. 1858–1867, Jul. 2008.

[32] D. Olsen, R. Lesser, J. Harris, R. Webber, and J. Cristion, "Automatic detection of seizures using electroencephalographic signals," U.S. Patent 5 311 876, May 17, 1994.

[33] P. Putter and N. R. Wagner, "Error detecting decimal digits," *Commun. ACM*, vol. 32, no. 1, pp. 106–110, Jan. 1989.

[34] R. T. Pivik, R. J. Broughton, R. Coppola, R. J. Davidson, N. Fox, and M. R. Nuwer, "Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts," *Psychophysiology*, vol. 30, no. 6, pp. 547–558, Nov. 1993.

[35] K. Polat and S. Gües, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform," *Appl. Math. Comput.*, vol. 187, no. 2, pp. 1017–1026, Apr. 2007.

[36] P. W. Ferrez and J. D. R. Millán, "Error-related EEG potentials generated during simulated brain–computer interaction," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 923–929, Mar. 2008.

[37] A. Rakotomamonjy, V. Guigue, G. Mallet, and V. Alvarado, "Ensemble of SVMs for improving brain computer interface P300 speller performances," in *Artificial Neural Networks: Biological Inspirations—ICANN 2005*. Berlin, Germany: Springer-Verlag, 2005, pp. 45–50.

[38] I. J. M. D. Rampil and M. J. D. V. M. Laster, "No correlation between quantitative electroencephalographic measurements and movement response to noxious stimuli during isoflurane anesthesia in rats," *Anesthesiology*, vol. 77, no. 5, pp. 920–925, Nov. 1992.

[39] O. A. Rosso, M. T. Martin, A. Figliola, K. Keller, and A. Plastino, "EEG analysis using wavelet-based information tools," *J. Neurosci. Methods*, vol. 153, no. 2, pp. 163–182, Jun. 2006.

[40] T. A. Salthouse, "Perceptual, cognitive, and motoric aspects of transcription typing," *Psychol. Bull.*, vol. 99, no. 3, pp. 303–319, May 1986.

[41] Discussion Paper 08 015 S. Scholtus, Algorithms for Correcting Some Obvious Inconsistencies and Rounding Errors in Business Survey Data, The Hague/Heerlen, The Netherlands: Statistics Netherlands2008, Discussion Paper 08 015.

[42] Discussion Paper 09 046 S. Scholtus, Automatic Correction of Simple Typing Errors in Numerical Data With Balance Edits, The Hague/Heerlen, The Netherlands: Statistics Netherlands2009, Discussion Paper 09 046.

[43] K. Q. Shen, X. P. Li, C. J. Ong, S. Y. Shao, and E. P. V. Wilder-Smith, "EEG-based mental fatigue measurement using multi-class support vector machines with confidence estimate," *Clin. Neurophysiol.*, vol. 119, no. 7, pp. 1524–1533, Jul. 2008.

[44] V. Srinivasan and C. Eswaran, "Approximate entropy-based epileptic EEG detection using artificial neural networks," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 3, pp. 288–295, May 2007.

[45] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *J. Roy. Stat. Soc. Ser. B, Methodological*, vol. 36, no. 2, pp. 111–147, 1974.

[46] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model," *Expert Syst. Appl.*, vol. 32, no. 4, pp. 1084–1093, May 2007.

[47] S. Trewin, "An invisible keyguard," in *Assets: Proc. 5th Int. ACM Conf. Assistive Technol.*, 2002, pp. 143–149.

[48] S. Wong, G. H. Baltuch, J. L. Jaggi, and S. F. Danish, "Functional localization and visualization of the subthalamic nucleus from microelectrode recordings acquired during DBS surgery with unsupervised machine learning," *J. Neural Eng.*, vol. 6, no. 2, p. 026006, Apr. 2009.

[49] C. Wu and Y. Liu, "Queuing network modeling of transcription typing," *ACM Trans. Comput.-Human Interact.*, vol. 15, no. 1, pp. 1–45, May 2008.

[50] W. Xu, C. Guan, C. E. Siong, S. Ranganatha, M. Thulasidas, and J. Wu, "High accuracy classification of EEG signal," in *Proc. 17th Int. Conf. Pattern Recognit.*, 2004, pp. 391–394.

[51] D. M. Young, "Data inaccuracy in the global transportation network," M.S. thesis, Air Force Inst. Technol., Wright-Patterson AFBase, OH, 1996.

Shouyi Wang (S'09) received the B.S. degree in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2003 and the M.S. degree in systems and control engineering from the Delft University of Technology, Delft, The Netherlands, in 2005. He is currently working toward the Ph.D. degree in the Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ.

His current research interests include data mining and machine learning, multivariate time series modeling and forecasting, and online intelligent decision-making system.

Cheng-Jhe Lin received the M.S. degree in industrial engineering and engineering management from National Tsing Hua University, Hsinchu, Taiwan, in 2002. He is currently working toward the Ph.D. degree in industrial and systems engineering at the University at Buffalo, The State University of New York, Buffalo.

His research interests include cognitive modeling for human perceptual-motor skills, application of data mining techniques in human error prevention, and quantitative human–computer interface design methods.

Changxu Wu received the B.S. degree in psychology with focus on engineering and mathematical psychology from Zhejiang University, Hangzhou, China, in 1999, the M.S. degree in engineering psychology and human–computer interaction from the Chinese Academy of Sciences, Beijing, China, in 2002, and the M.S. and Ph.D. degrees in industrial and operational engineering from the University of Michigan, Ann Arbor, in 2004 and 2007, respectively.

Since August 2007, he has been an Assistant Professor with the Department of Industrial and System Engineering, University at Buffalo, The State University of New York, Buffalo. The main theme of his current research is the development of computational models of human performance and mental workload, addressing both fundamental and neurological issues of perceptual-motor behavior and human cognition with their applications in designing intelligent systems. He has published papers on *Psychological Review*, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A, *ACM Transactions on Computer–Human Interaction*, *International Journal of Human–Computer Studies*, as well as several other journals.

Dr. Wu is a member of the Human Factors and Ergonomics Society, the Society of Automobile Engineers, the Cognitive Science Society, and the American Society of Engineering Education (ASEE). He served IEEE and other academic organizations as a Reviewer of IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, *Applied Ergonomics*, and IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A. He was the recipient of the Outstanding Student Instructor Award from the ASEE at the University of Michigan in 2006.

Wanpracha Art Chaovalitwongse (M'05–SM'11) received the B.S. degree in telecommunication engineering from the King Mongkut Institute of Technology Ladkrabang, Bangkok, Thailand, in 1999 and the M.S. and Ph.D. degrees in industrial and systems engineering from the University of Florida, Gainesville, in 2000 and 2003, respectively.

In 2004, he was with Corporate Strategic Research, ExxonMobil Research and Engineering, where he managed research in developing efficient mathematical models and novel statistical data analyses for upstream and downstream business operations. He is currently an Associate Professor of industrial and systems engineering with Rutgers University, Piscataway, NJ, a Visiting Associate Professor of operations research and financial engineering with Princeton University, Princeton, NJ, and an Affiliate Associate Professor of industrial and systems engineering with the University of Washington, Seattle.