Integrating Human Behavior Modeling and Data Mining Techniques to Predict Human Errors in Numerical Typing

Cheng-Jhe Lin, Changxu Wu, *Member, IEEE*, and Wanpracha A. Chaovalitwongse, *Senior Member, IEEE*

***Abstract*—Numerical typing errors can lead to serious consequences, but various causes of human errors and the lack of contextual clues in numerical typing make their prediction difficult. Human behavior modeling can predict the general tendency in making errors, while data mining can recognize neurophysiological feedback in detecting cognitive abnormality on a trial-by-trial basis. This study suggests integrating human behavior modeling and data mining to predict human errors because it utilizes both 1) top-down inference to transform interactions between task characteristics and conditions into a general inclination of an average operator to make errors and 2) bottom-up analysis in parsing psychophysiological measurements into an individual's likelihood of making errors on a trial-by-trial basis. Real-time electroencephalograph (EEG) features collected in a numerical typing experiment and modeling features produced by an enhanced human behavior model (queuing network model human processor) were combined to improve error classification performance by a linear discriminant analysis (LDA) classifier. Integrating EEG and modeling features improved the results of LDA classification by 28.3% in keenness (d') and by 10.7% in the area under ROC curve (AUC) from that of using EEG only; it also outperformed the other three benchmarking scenarios: using behaviors only, using apparent task features, and using task features plus trial information. The AUC was significantly increased from using EEG along only if EEG + Model features were used.**

***Index Terms*—Behavior modeling, data mining, electroencephalograph (EEG), human errors, linear discriminant analysis, numerical typing.**

I. Introduction

HUMAN errors in numerical typing tasks can induce financial loss, safety threats, or even fatalities in critical systems [1]–[3]. Error prevention with numerical data entry, however, is particularly difficult. First, there are no unified theories concerning error mechanisms [4]. Errors may reflect the cumulative effects of many different factors and involve preresponse conditions and postresponse processes [5]. Thus, behavioral, psychophysiological, and biomechanical evidence should converge to explain how humans commit errors. Although Salthouse [6] investigated alphabetical typing errors and concluded that typing errors could occur in either input (perception of inputs), parsing (recognition of inputs), translation (formation of movement specifications) or execution (production of movements) phases of the cognitive process, Lin and Wu [7] found that error mechanisms for alphabetical typing did not necessarily translate to numerical typing. Furthermore, automatic error detection for numerical typing is difficult due to the lack of contextual clues. For example, "CAK" can be detected as a potential error because there is no such word, but numerical data may only have patterns with respect to format, maximum, minimum, or ranges. The various causes of human errors and the difficulty in detecting human errors by contextual clues in numerical typing suggest a need for proactive error prevention methods based on psychophysiological measurements that reflect real-time status of various cognitive processes and functional changes in brain activity.

Multichannel electroencephalograph (EEG) is a psychophysiological measurement of brain activity in a waveform that is sensitive to functional changes and indicative of mental status. EEG of different spectra can provide valuable insights to states of vigilance [8], fatigue [9], [10], emotion [11], attention/concentration [12], and even different cognitive tasks [13], [14]. For example, wakefulness was negatively related to alpha activity (8–13 Hz) [8]. EEG signals collected from different sites of the scalp also provide functional mappings to cortical areas. For example, error-related brain potentials are most characterized by the EEG signals along the frontal-central part of the midline (e.g., Cz and Fz electrode according to international 10–20 system) as the anterior cingulate cortex is the major area responsible for error detection [4], [15]. Whenever there is a hand movement, EEG displays an event-related desynchronization (ERD) on the contralateral side of the parietal lobes where the somatosensory cortex is located [16]. More importantly, motor-related cortical components (MRCPs) in EEG have been used to predict movement accuracy because they indicate motor preparedness [17]. Although the information in MRCPs was a reliable indicator of error occurrence, insufficient motor preparedness at the planning stage in the central nervous system may result in low movement accuracy when motor commands are carried out by peripheral motor systems.

Although multichannel EEG was informative for functional changes in different cortical areas and for motor accuracy, EEG

Manuscript received August 13, 2013; revised March 29, 2014 and July 1, 2014; accepted August 17, 2014. Date of publication December 31, 2014; date of current version January 13, 2015. This work was supported by the National Science Foundation under Grant 0916883 and Grant 0916580. This paper was recommended by Associate Editor M. Dorneich.

C.-J. was with the Department of Industrial and Systems Engineering, State University of New York at Buffalo in completing this work. He is now with Department of Industrial Management, National Taiwan University of Science and Technology, Taipei 10607, Taiwan (email: Robert_cjlin@mail.ntust.edu.tw)

C. Wu is with the Department of Industrial and Systems Engineering, State University of New York, Buffalo, NY 14214 USA (email: changxu.wu@gmail.com; seanwu@buffalo.edu).

W. A. Chaovalitwongse is with the Department of Industrial and Systems Engineering, University of Washington, Seattle, WA 98195 USA (e-mail: 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/THMS.2014.2357178

analysis is an inherently challenging task. First, EEG contains nonstationary, stochastic signals [10]. An EEG pattern in association with severe abnormality such as an epileptic seizure is rather unpredictable. Even medical experts can confuse what event is happening [18]. Distinguishing between less salient differences between EEGs associated with correct and error responses is difficult. Furthermore, prolonged collection of online EEG data creates large datasets, the majority of which may be irrelevant to the event of interest, and may require tedious analysis processes [10]. Finally, EEG responses are subject to artifacts (e.g., muscular artifacts due to eye movements [19]), noise (e.g., electromagnet interference [20]), and subject-dependent EEG characteristics. For example, the position of ERD may vary between subjects, and the optimal frequency band for analysis is also subject-specific [16]. Those challenges call for a robust methodology to recognize mental states and identify functional changes by analyzing multichannel EEG.

Using data mining techniques to classify multichannel EEG signals into normal and error-related states may be an effective solution to prevent errors in numerical typing if error-related states can be detected beforehand. Although EEG can reflect motor preparedness of movements to a certain extent [21], its predictive power for final motor execution is limited. Wang *et al.* [22] obtained 0.35 and 0.47 for keenness (d') of error detection using support vector machine (SVM) and linear discriminant analysis (LDA) classifiers, respectively. The relatively low keenness revealed that EEG alone was insufficient for capturing information relevant to typing errors due to peripheral processes beyond cognition. For example, neuromotor noise theory states that a perceptual-motor system is a stochastic system. Therefore, psychological stressors would enhance its inherent noise and biomechanical filtering needs to control resultant physiological tremors [23]–[25]. When speedy movements are pushed beyond a certain optimal point, errors may be caused by spatial variability of the motoric movements even if the cognitive process is well controlled. Furthermore, one cannot obtain good results in classifying EEG data when error rates are generally low and variation is large across subjects [4] as training with EEG data from error and correct trials is difficult.

To better predict human errors, a computational behavior model that involves conceptual, cognitive, and motor processes may be necessary. In successive responses during numerical typing, a slow reaction in one keystroke can delay another keystroke, and typing with single or multiple fingers may imply different difficulty. Time pressure also impacts the spatial variability of movements negatively. Those trends are not captured by analyzing EEG in single trials but well reflected in the modeling results. Therefore, combining human behavioral modeling features that transform observations to possibilities of human errors and EEG features that give real-time human cognitive status can improve results of error detection. Human behavioral modeling can complement single trial EEG analysis by supplementing the general inclination of making errors given task conditions while EEG features reflect functional changes in cortical activities on a trial-by-trial basis. In the center of Fig. 1, an error occurs when the task is difficult (symbolized by a thick circle) and the human is not well prepared (signified by noisy EEG). Data mining learns this by bottom-up training, but it only captures the human part of the nature (psychophysiological feedback). In contrast, a human behavior model is constructed top-down, establishing relationships between tasks and behaviors without knowing the human's mental status. Integration is compatible with the analysis by synthesis [26] approach based on the belief that, if the underlying model for production of outcomes (e.g., human errors) can be determined, then the actions to produce outcomes (e.g., behaviors that are observed before the errors) can be classified. In other words, if a model can predict final behavior outcomes (e.g., human errors) based on environmental inputs, such as external task demands and time stress, then the intermediate outputs of the model that are used to determine the final outcomes should be classifiable, and their combination with EEG data should render patterns more distinguishable.

This study presents a new numerical typing error prediction framework based on integrated psychophysiological and behavioral modeling features. This study also demonstrates that behavior modeling features can improve the error prediction by data classification better than apparent task features (i.e., characteristics of tasks that are influential to the performance but can be obtained without a human behavior model) does. This paper presents a framework to integrate real-time brain activity features through analysis of EEG signals before the typing responses with behavior modeling features based on prior knowledge of error mechanisms generated offline by an enhanced queuing network-model human processor (QN-MHP) model. An LDA classifier is trained using three integrated feature sets (EEG + Modeling, EEG + Task, and EEG + Task + Condition), and its error prediction performance is compared with the performance when only EEG features are used.

The rest of the paper is as follows. Section II highlights related research in data mining with error-related EEG components. Section III includes the EEG data collection methods and EEG feature extraction. Section IV provides the generation of behavioral modeling features, the extraction of task-related features, and data classification by an LDA classifier. The classification results are in Section V. The discussion is in Section VI.

II. Related Work

When EEG was analyzed with data mining technique, researchers used to focus on its applications to brain–computer interfaces (BCIs) [27], where EEG classifiers served as feature translators that convert EEG features into logical control commands. In general, linear classifiers (e.g., LDA) are more robust than nonlinear ones [e.g., artificial neural network (ANN)] because they have fewer parameters to tune and are less prone to overfitting [27], [28]. Averaging/rejecting activations of low certainty, debounce blocking (i.e., the interface is deactivated during a refractory period), and event-related negativity were often used as postprocessing methods to reduce errors in using BCIs. Error detection was generally conducted posteriorly, while proactive error prevention (i.e., predicting errors before the commands were sent or when the user was unaware of errors) was absent. Chavarriaga *et al.* [29] assessed whether recognizing error-related EEG potentials during gesture-based human–computer interaction could improve the

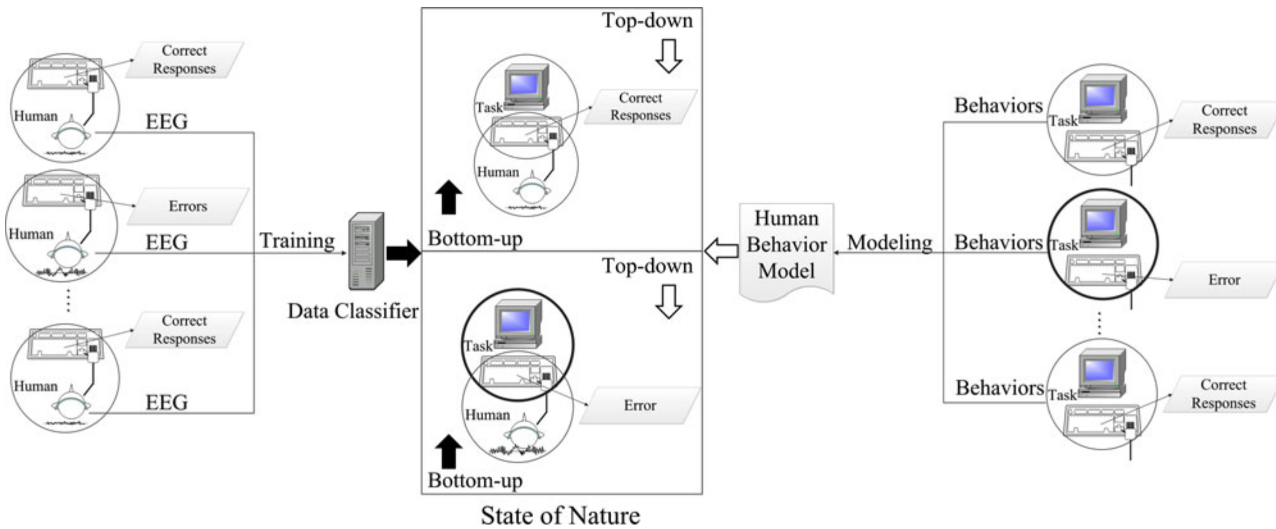

Fig. 1. Conceptual diagram of integration.

system (i.e., only trials that were not classified as errors would be considered correct recognitions of gestures). With a recursive Bayesian estimation technique, the classifier's performance reached above random level in six out of seven subjects, and it improved the gesture recognition rate of a k-nearest neighbor classifier by 6.4%. Blankertz *et al.* [30] used a Fisher discriminant classifier to detect error-related potentials in a two-choice reaction time task and obtained above a 78% detection (hit) rate from seven out of eight subjects. Parra *et al.* [31], [32] utilized an LDA classifier to detect error-related potentials in a visual detection task (flanker task) and the classification achieved 0.79 AUC (area under ROC curve) with 64-channel EEG. Their online application showed 21.4% ± 21.7% error reduction rate (two out of seven subjects showed degradation in performance). Autoerror recovery in BCI operations using event-related potential such as P300 as a response verification mechanism was also reported [33].

The aforementioned research utilized error-related or event-related brain potentials elicited after responses, but neurophysiological literature shows that some EEG components that occur before responses may indicate motor preparedness and thus are predictive of final response accuracy. For example, Bereitschaftspotential (BP) and negative slope (NS) of MRCPs were associated with preparation and/or execution of voluntary movements [21]. They were differentiated in that BP was affected by psychological factors such as level of intention, preparation, and movement selection, but NS was related to physical factors such as precision, discreteness, and complexity of the movement itself [34]. BPs of professional athletes were short in duration with smaller amplitudes and earlier onsets than those of non-professionals [35], [36]. The NS component was predictive of movement accuracy in aiming tasks [17]. Furthermore, planning and execution of hand and/or finger movement desynchronize the mu rhythms, and the power decrease of this ERD can be used to discriminate four motor imagery tasks (left-hand, right-hand, foot, and tongue) based on classification of single EEG trials [37]. LDAs have also been used to differentiate spatial features of multichannel EEGs processed by different independent components analysis algorithms from four motor imagery tasks [38]. The evidence showed a promising potential for using LDA to classify EEG components before responses as method to predict motor accuracy in advance.

Studies have attempted to adopt data mining techniques, including feature combination, utilization of inverse models, and contrast enhancing. Dornhege *et al.* [28] combined MRCP features and ERD features by a joint probabilistic modeling (PROB) method and used LDA classifiers tailored specifically for those complementary features to obtain bit rate gains as high as 50%. Ferrez [39] claimed that inverse models can reconstruct neuronal sources of brain activities measured at the scalp (e.g., EEG) so that better estimates of intracranial activity can be provided. He tested the cortical current density (CCD) model and the estimated local filed potential (eLFP) model which provided estimated intracranial activity and compared them with temporal EEG features. The CCD model features increased the recognition rate of error trials in a BCI interface by 2%, while the eLFP model features degraded the performance of a Gaussian classifier by 2%. Other studies tried to enhance contrast by reconstructing (e.g., wavelet transform [40]) or recovering (e.g., blind source separation [41]) EEG features from noises and artifacts. Thus, while different modeling techniques have been used to enhance or reconstruct psychophysiological features to improve BCI effectiveness, the potential of combining behavioral modeling features and EEG features to prevent human errors in generic human–computer interaction, however, has never been studied.

III. Methods for Electroencephalograph Collection and Extraction

A. Task

The task was a simulated numerical hear-and-type [7] task. A desktop computer played 30 random combinations of nine digits in each trial and subjects were instructed to press corresponding keys one-by-one when they heard the verbally presented numbers.

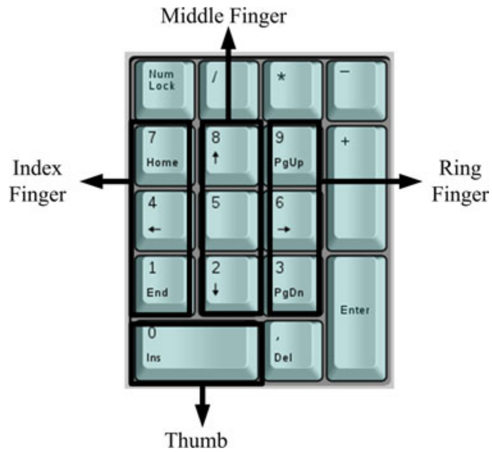

Fig. 2. Recommended fingering pattern.

B. Subjects

Eight right-handed university community members volunteered to participate in this study, including four males and four females. The average age of subjects was 22.3 $\pm$ 1.8 and their time was compensated.

C. Procedure

Subjects were pretested following the procedure in [7] to assure their numerical typing skills. If they qualified, they continued eight experimental trials of different speech rates, fingering strategies, and urgency after a five minute break. In each trial, subjects were asked to type both quickly and accurately. To avoid accumulating fatigue, subjects were asked to rest for five minutes after four consecutive trials.

D. Trial Design

Three variables were used to create the trials.

1) *Speech Rate*: To simulate different speech rates, the desktop PC played the random numbers at two different speeds: high (2 digit/s) and low (1 digit/s).
2) *Fingering Strategy*: There were two possible fingering strategies: single finger (index finger) or multiple fingers (based on a recommended fingering pattern in Fig. 2).
3) *Urgency*: Subjects were asked to hurry while maintaining accuracy in the urgent trials. To assure their compliance to the instruction, a performance-based bonus was given in urgent trials. The amount of bonus was determined by the percentage of fast and accurate responses, i.e., correct keystrokes within 600 ms.

Each subject performed one trial at each combination of factor levels [2 (speech rate) $\times$ 2 (fingering strategy) $\times$ 2 (urgency)] for a total of eight trials.

E. Data Collection and Extraction

During the experiment, EEG data were collected by an EEG cap containing 40 Ag/AgCl electrodes according to the international 10–20 system. The signals were sampled at 1000 Hz and amplified by the NuAmps Express system (Neuroscan Inc., Charlotte, NC, USA). Raw EEG data were first processed by a dc to 30-Hz bandpass filter (EEG above 30 Hz greatly overlapped with the muscular artifacts that might incur due to typing movements [20]). Filtered EEG data in a time window from 450 ms before keystrokes to 0 ms at keystrokes were then extracted as epochs. 450-ms length was chosen because otherwise the segments of different keystrokes would overlap. All epochs were baseline corrected to eliminate signal drifts and downsampled to 20 Hz. The amplitudes of downsampled EEG were used as temporal features and were further decomposed to 150-ms subsets, resulting three features in three subsets for each electrode [42]. Only features from six electrodes (FC3, FCZ, C3, CZ, CP3, and CPZ) were used for further data classification (those electrodes located above the motor sensory cortex and close to the area in charge of right hand movements). All data processing was conducted by batch functions in the Edit module of the Scan 4.3 software (Neuroscan Inc., Charlotte, NC, USA).

IV. Other Feature Generation, Extraction, and Data Classification

A. Modeling Feature Generation

Of the human behavior models proposed to account for typing behaviors [43]–[45], the QN-MHP covers the majority of behavior phenomena observed from skillful typists [6], [43]. It is a computational architecture consisting of perceptual, cognitive, and motor subnetworks where different servers represent information-processing units that simulate cortical functions. Links connect servers and represent neurological pathways between cortical areas (see Fig. 3). See [43] for mapping of the QN-MHP architecture to neurological research findings.

An enhanced QN-MHP model [46] was used in this study to generate modeling features for numerical typing. It incorporated several mechanisms to account for changes in performance time and accuracy under different pacing (speech rates), motor control (fingering strategies), and urgency conditions. A top-down control routed information entities actively to simulate an enforced delay directed by the prefrontal cortex during multitasking due to cognitive-bottleneck [47], resulting a two-task interference (see Fig. 3). Close-loop control was assumed to replace ballistic movement control under urgency [48], resulting in a speed–accuracy tradeoff. A change of the fingering strategy (one finger or multiple) could affect performance time and accuracy through a NGOMSL-style response-selection procedure and a neuromotor noise modeling. Multifinger typing (see Table I) was assumed to take one additional cycle for deciding which finger to use.

The enhanced QN-MHP also quantified neuromotor noise

$$\sigma_u = \sqrt{k_{\mathrm{SDN}}^2 \cdot \frac{u^2}{c^2} + k_{\mathrm{CN}}^2 \cdot I^2 + k_{\mathrm{TN}}^2 \cdot c^2} \qquad (1)$$

where σ_u was extent of noise added to muscle activation level u in a movement duration $c{*}MT_0$ ($MT_0 = 420$ *ms*, average movement time from [49]). k_{SDN}, k_{CN}, and k_{TN} were constants for relative weights of three noise components: signal-dependent noise (SDN), constant noise (CN), and temporal noise (TN). Their settings can be found in [46]. From (1), noise increased with the muscle activation level (u) during multifinger typing, because it required more motor unit recruitment to compensate

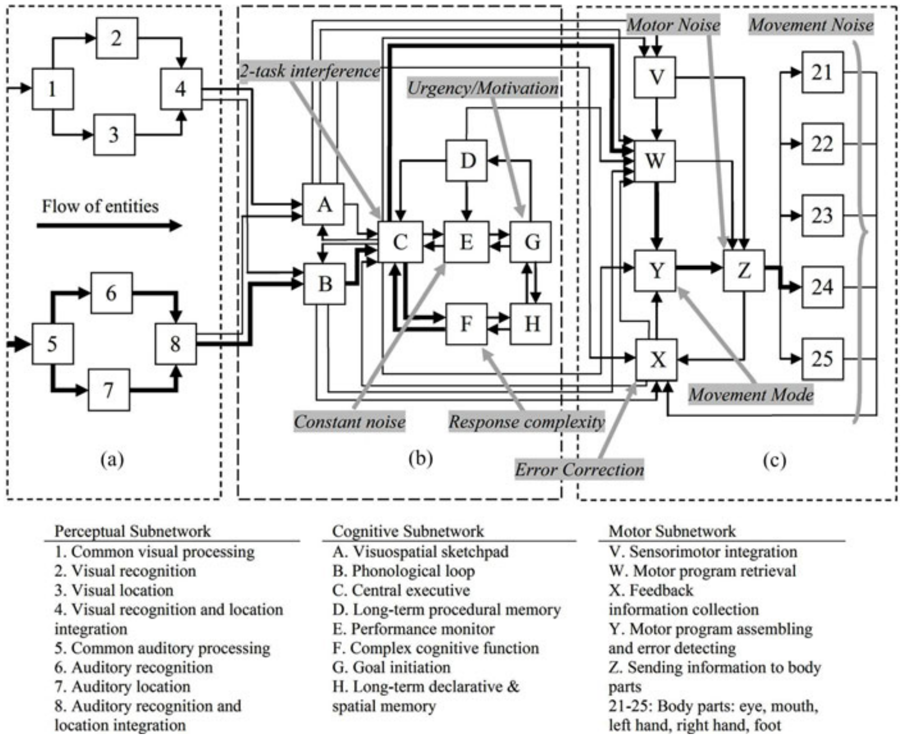

Fig. 3. QN-MHP architecture and server functions.

TABLE I
NGOMSL-STYLE TASK PROCEDURE FOR NUMERICAL TYPING AT SERVER F

GOAL: Do numerical typing task method for GOAL: Press <key> at <location> with <finger>	
Step 1	Retrieve <location> of <number> on a numerical keyboard from LTM
Step 2	Watch for <label> around <location>
Step 3	Recall <finger strategy> from LTM
Step 4	Decide if <finger strategy> matches <single finger> IF MATCH go to step 5 ELSE go to step 7
Step 5	Press <key> at <location> with <index finger>
Step 6	Return with goal accomplished
Step 7	Decide <finger to use> based on <location>
Step 8	Press <key> at <location> with < finger to use >
Step 9	Return with goal accomplished

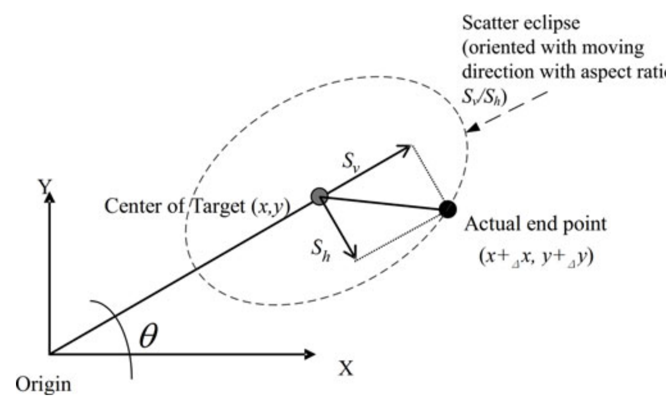

Fig. 4. Endpoint variability in two orthogonal directions.

for a finger force deficit caused by finger-enslaving effects [50] and thus generated more neuromotor noise.

Based on neuromotor noise theory, the spatial variability of movements is proportional to neuromotor noise. Therefore, the endpoint variability increased when multifinger typing was used. During fast pacing, CN was assumed to be magnified by interference index (I)

$$I = DL/(T_{i,C} + T_{i,F} + T_{i,C}) \tag{2}$$

where DL is the delay caused by the two-task interference; $T_{i,C}$ and $T_{i,F}$ are processing times for information entity i at server C and server F in the cognitive subnetwork (see Fig. 2). The calculation of DL and settings of $T_{i,C}$ and $T_{i,F}$ parameters are described in [46]. When hurried (the urgent trials), the typists were assumed to alter their movement mode from visually guided aiming to ballistic aiming (regulated by server Y in Fig. 3), resulting in loss of correction at the homing phase and higher spatial variability. Based on those mechanisms and other QN-MHP assumptions (e.g., the response time to an information entity can be computed by summarizing the processing time of the entity at each server on its path in Fig. 2 plus a keyclosure time), the response time for different typing conditions can be generated. In addition, the spatial variability can be computed and further decomposed into two orthogonal components that are parallel and perpendicular to the movement direction (see Fig. 4). The predicted response time, the aiming shifts estimated in x-direction ($_\Delta x$) and in y-direction ($_\Delta y$), together with interference index (I) from equation (2) were used as modeling features.

B. Task Feature Extraction

Combining modeling features with EEG features may benefit data classification because the integrated feature set contains apparent task characteristics that may imply task difficulty and are thus indicative of susceptibility to errors. Four apparent task features were selected to test if combining them with EEG features alone can produce similar results to that of combining modeling features with EEG features.

1) *Number to be pressed (N)*: Different numbers to be pressed mean different reaching difficulty and were thus relevant to errors. For example, "3" and "7" keys were more difficult to reach on a standard numerical keypad [7].
2) *Quickness of previous keystroke (Q)*: In numerical hear-and-type tasks, keystrokes were made successively and rhythmically. When a keystroke was made too quickly or too slowly, it might disturb the typing rhythm and cause errors for the following keystroke. The quickness of the previous keystroke was quantified as the average velocity of the previous keystroke, i.e.

$$Q_{i-1} = D(N_{i-2}, N_{i-1})/PT_{i-1} \tag{3}$$

   where Q_{i-1} was the quickness of the previous keystroke (the current keystroke is the ith keystroke) and the D(N_{i-2}, N_{i-1}) was the travel distance of the previous keystroke (from key N_{i-2} to N_{i-1}). PT is the response time of the previous keystroke that could be obtained by online behavior feedbacks, i.e., reaction time measurements.
3) *Fitts' difficulty index (DI)*: Fitts' DI was used to quantify movement difficulty in aiming

$$DI = \log_2(2 * D/S) \tag{4}$$

   where *D* is the travel distance of the keystroke, and *S* is the target size (width of the key).
4) *Movement angle of the keystroke (A)*: In 2-D aiming task, the movement angle was found to be influential for aiming performance [49], [51]. The movement angles were quantified in a 360° polar coordinate where moving right A was 0°, left 180°, upward 90°, and downward 270°.

C. Feature Combination and Data Classification

Three subsets of 18-D EEG features (3 features/electrode × 6 electrodes = 18 features for each keystroke) were collected through the experiment in Section III. The three types of features (EEG features, modeling features, and task features) were further combined into four datasets (see Table II):

1) an EEG only set contained only 18-D EEG features;
2) a behaviors only set contained both apparent task features and advanced modeling features derived from interaction between tasks and behaviors;
3) an EEG + Task set contained 18-D EEG features and 4-D task features;
4) an EEG + Model set contained 18-D EEG features and 4-D modeling features.

Since there were three subsets of EEG features representing psychophysiological responses in different time windows before each keystroke, they were combined to task features and modeling features alternately and the combined data went through the LDA training process one by one. All data points in a given set were equally[1] divided into two nonoverlap subsets: a training set consisting of data collected earlier in an experimental trial and a query set consisting of data collected later. It was assumed that the best training performance will indicate the best time window to extract EEG features, e.g., if the training data extracted in a window from 450 to 300 ms before the keystroke yielded the best training accuracy, the query set should contain EEG features extracted from the same time window (−450 to −300 ms). The LDA training process was completed by using Minitab software package (Minitab Inc., Pennsylvania, USA) and the concept of determining the query set is depicted in Fig. 5. After the training, the software produced the decision boundary of the trained LDA classifier as well as the Mahalanobis distance[2] of each data point to the group center

$$D(\mathbf{x}, \mathbf{m}_i) = (\mathbf{x} - \mathbf{m}_i)^t \mathbf{\Sigma}^{-1} (\mathbf{x} - \mathbf{m}_i), \quad i = c \text{ or } e \tag{5}$$

where $\mathbf{x}$ denotes a p-dimensional feature vector that represents all features relevant to a keystroke; $\mathbf{m_c}$ and $\mathbf{m_e}$ are the p-dimensional sample mean of the classified correct response group and erroneous response group, respectively. $\mathbf{\Sigma}$ is the pooled covariance matrix for two groups. Then, the rule of classification can be expressed by the following:

$$\begin{aligned} &\text{Classify } \mathbf{x} \text{ into } D_c \text{ if} \\ &\qquad \mathbf{D}(\mathbf{x}, \mathbf{m_e})/\mathbf{D}(\mathbf{x}, \mathbf{m_c}) > \text{criterion} \\ &\text{else} \\ &\qquad \text{Classify } \mathbf{x} \text{ into } \mathbf{D_e}. \end{aligned} \tag{6}$$

The criterion of classification should be set to 1. The criterion, however, could be manipulated to increase the keenness of LDA classifier to the errors. Raising the criterion value would increase both the hit rate and the false alarm rate. A keenness analysis was, thus, needed to determine optimal criterion for the ratio $\boldsymbol{D}(\mathbf{x}, \boldsymbol{m_e})/\boldsymbol{D}(\mathbf{x}, \boldsymbol{m_c})$ greater than 1 (the baseline) that provides the highest d-prime for the receiver's operating characteristic (ROC) curve, i.e., $d' = Z(Hit) - Z(FalseAlarm))$ and AUC. A VBA program was coded to perform this analysis by adjusting the ratio between the two Mahalanobis distances as a criterion of classification.

D. Data Analysis

The LDA classification results for EEG only, behaviors only, EEG + Task, and EEG + model models in terms of training

[1]Whether there were equal numbers of data in the training and query set depended on whether there was an even or odd number of correct/incorrect responses in each trial. If there was an even number of correct/incorrect responses, they were equally assigned into training and query set. If there were an odd number of correct/incorrect responses, the training set would have one more data point than the query set.

[2]The Mahalanobis Distance is the distance between the test point $\mathbf{x}$ to the center of a set of given sample points that definitely belong to that set (e.g., the set of correct responses or errors), taking into account the correlations of the data set. When data points are distributed in p-dimensional space, whether the test point is close to the average or center of the sample points determines its membership to the data set. In those directions where data points are spread out over a large range, the test point can be relatively further away from the center but still belongs to the set. Therefore, Mahalanobis distance corresponds to Euclidean distance in the transformed space based on the shape of the distribution.

TABLE II
FEATURES USED IN EEG ONLY, EEG + TASK, AND EEG + MODEL SET

Feature	EEG only	Behaviors only	EEG + Task	EEG + Model
EEG temporal	EEG amplitude at FC3, FCZ, C3, CZ, CP3, and CPZ	N/C	EEG amplitude at FC3, FCZ, C3, CZ, CP3, and CPZ	EEG amplitude at FC3, FCZ, C3, CZ, CP3, and CPZ
Task	N/C*	Number to be pressed Quickness of previous keystroke Fitts' difficulty index Movement angle	Number to be pressed Quickness of previous keystroke Fitts' difficulty index Movement angle	N/C
Modeling	N/C	Predicted response time Predict shift in X direction Predict shift in Y direction Estimated Interference %	N/C	Predicted response time Predict shift in X direction Predict shift in Y direction Estimated Interference %

*N/C: not included.

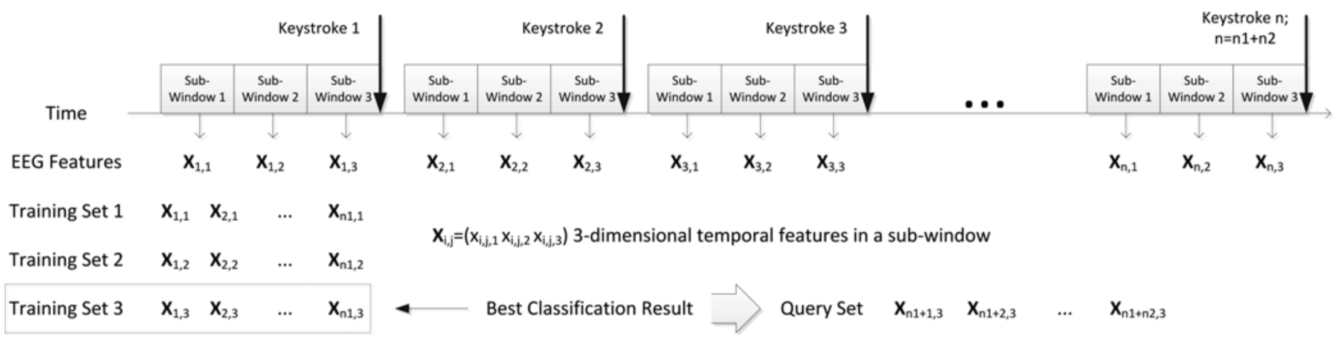

Fig. 5. Determining the query set.

accuracy, hit rates, and false alarm rates were calculated by the VBA program. First, the ROC curve was obtained by manipulating the classification criterion in (6). Then, the best keenness of detection in query and AUC obtained through the manipulation were computed. Finally, pair-wise comparisons for the AUC achieved by different models were conducted using the Minitab statistical software.

V. RESULTS

The classification results for using EEG only dataset are listed in Table III. After an initial inspection of typing performance, data from two subjects (subjects 1 and 4) were excluded due to their extremely high accuracy in numerical typing (insufficient samples of errors in the training set). Using only temporal EEG features from the remaining six subjects, the LDA classifier achieved keenness $d' = 0.60$ and area under ROC curve $AUC = 0.56$. The AUC for all subjects were above chance level (50%) except for subject 8, and errors could be detected as early as 300 ms before keystrokes (the training accuracy was found in the subwindow 300 to 450 ms before keystrokes).

The classification results for using the "behaviors only" set are listed in Table IV. On average, the keenness decreased 3.3% and AUC decreased 8.9% from the classification results using EEG only. For four out of six subjects, the AUC decreased but overall the decrease was not significant statistically. The only pair-wise comparison for AUC that was significant was behaviors only versus EEG + Model ($\mathrm{t}_5 = 3.15$; $\mathrm{p} = 0.025$). The results showed that relying on behavioral data only was not as effective as using psychophysiological feedback because behavioral data provided only general inclinations of making errors but contained no real-time information for motor preparation and execution as psychophysiological feedback did.

The classification results for using EEG + Task set are listed in Table V. On average, the keenness of the LDA classifier increased 3.2% and AUC increased 3.6%, comparing with classification results when EEG only set was used. There was a trend for one pair-wise comparison for AUC: EEG + task versus EEG + Model ($\mathrm{t}_5 = 2.27$; $\mathrm{p} = 0.072$). In addition, AUC of two out of six subjects became worse. The results showed that adding features derived from apparent task characteristics did not necessarily benefit error classification in numerical typing.

The classification results for using EEG + model set are listed in Table VI. The LDA classifier obtained 28.3% increase in keenness and 10.7% increase in AUC. The EEG only versus EEG + Model comparison for AUC was significant ($\mathrm{t}_5 = 3.99$; $\mathrm{p} = 0.010$). All six subjects obtained better AUC. Therefore, modeling features effectively improved error classification results using LDA and EEG + Model features were relatively robust than EEG + Task features.

It might be arguable that modeling features had advantages over task features because experimental conditions, such as

TABLE III
CLASSIFICATION RESULTS FOR EEG ONLY SET

Subject	Best Interval[1]	CR[2]	Training Accuracy	Prediction (Query)			
				Hit Rate	False Alarm	Keenness[3]	AUC[4]
2	(−150, 0)	1.25	70.40%	95.65%	83.62%	0.73	0.56
3	(−450, −300)	1.14	65.30%	97.37%	80.88%	1.06	0.58
5	(−150, 0)	1.02	69.50%	52.00%	32.80%	0.50	0.61
6	(−150, 0)	1.12	67.60%	88.57%	70.45%	0.67	0.59
7	(−450, −300)	1.19	69.30%	94.44%	82.38%	0.66	0.57
8	(−450, −300)	1.05	65.00%	58.62%	58.36%	0.01	0.45
Average		-	67.34%	81.11%	68.08%	0.60	0.56

[1]Best interval indicates which of the three subsets of EEG temporal features was selected. Negative value means before keystroke in units of milliseconds. The best interval was selected based on the training results with cross-validation, i.e., the interval during which the training accuracy with cross-validation is the highest was selected.
[2]CR is the best criterion (CR $>$ 1) set in decision rule (6) which generated the highest d-prime. A CR $>$ 1 rendered the classifier more sensitive to error detection (higher hit rate) and susceptible to false alarms.
[3]Keenness (*d'*) varied using different CR. Here, the best *d'* (the point at which ROC curve was best separated from chance level) was listed with its corresponding hit and false alarm rates.
[4]AUC is the area under the ROC curve; the $+/-$ sign after AUC value signifies whether the AUC increased/decreased comparing with when EEG only set was used.

TABLE IV
CLASSIFICATION RESULTS FOR BEHAVIORS ONLY SET

Subject	Best Interval[1]	CR[2]	Training Accuracy	Prediction (Query)			
				Hit Rate	False Alarm	Keenness[3]	AUC[4]
2	-	1.62	65.50%	100.00%	100.00%	0.50	0.52−
3	-	1.15	68.90%	96.15%	84.47%	0.76	0.62+
5	-	1.00	56.60%	45.45%	38.33%	0.18	0.44−
6	-	0.60	65.50%	5.26%	0.49%	0.96	0.58−
7	-	1.35	66.10%	90.91%	83.58%	0.28	0.44−
8	-	1.11	63.90%	90.48%	68.94%	0.82	0.48+
Average		-	64.42%	71.38%	62.64%	0.58	0.51−

[1]For behavioral features, there was only one subset because they are not collected in real time and contained no temporal information, i.e., they did not vary with time.
[2]CR is the best criterion (CR $>$ 1) set in decision rule (6) which generated the highest d-prime. A CR $>$ 1 rendered the classifier more sensitive to error detection (higher hit rate) and susceptible to false alarms.
[3]Keenness (*d'*) varied using different CR. Here, the best *d'* (the point at which ROC curve was best separated from chance level) was listed with its corresponding hit and false alarm rates.
[4]AUC is the area under the ROC curve; the $+/-$ sign after AUC value signifies whether the AUC increased/decreased comparing with when EEG only set was used.

TABLE V
CLASSIFICATION RESULTS FOR EEG + TASK SET

Subject	Best Interval[1]	CR[2]	Training Accuracy	Prediction (Query)			
				Hit Rate	False Alarm	Keenness[3]	AUC[4]
2	(−300, 150)	1.16	74.90%	93.75%	66.34%	1.11	0.59+
3	(−450, −300)	1.05	68.30%	57.69%	47.56%	0.26	0.58−
5	(−150, 0)	1.00	73.00%	54.55%	25.33%	0.78	0.64+
6	(−150, 0)	1.15	71.40%	80.00%	71.88%	0.26	0.54−
7	(−150, 0)	1.21	87.60%	81.82%	41.98%	1.11	0.63+
8	(−450, −300)	1.05	68.10%	61.90%	53.12%	0.22	0.47+
Average		-	73.88%	71.62%	51.04%	0.62	0.58+

[1]Best interval indicates which of the three subsets of EEG temporal features was selected. Negative value means before keystroke in units of milliseconds. The best interval was selected based on the training results with cross-validation, i.e., the interval during which the training accuracy with cross-validation is the highest was selected.
[2]CR is the best criterion (CR $>$ 1) set in decision rule (6) which generated the highest d-prime. A CR $>$ 1 rendered the classifier more sensitive to error detection (higher hit rate) and susceptible to false alarms.
[3]Keenness (*d'*) varied using different CR. Here, the best *d'* (the point at which ROC curve was best separated from chance level) was listed with its corresponding hit and false alarm rates.
[4]AUC is the area under the ROC curve; the $+/-$ sign after AUC value signifies whether the AUC increased/decreased comparing with when EEG only set was used.

TABLE VI
CLASSIFICATION RESULTS FOR EEG + MODEL SET

Subject	**Best Interval**[1]	**CR**[2]	**Training Accuracy**	**Prediction (Query)**			
				Hit Rate	**False Alarm**	**Keenness**[3]	**AUC**[4]
2	(−150, 0)	1.33	97.10%	87.50%	53.70%	1.06	0.68+
3	(−300, −150)	1.17	91.50%	88.46%	66.09%	0.78	0.65+
5	(−450, −300)	1.11	94.20%	61.54%	39.18%	0.56	0.62+
6	(−150, 0)	1.26	89.70%	95.00%	76.66%	0.92	0.64+
7	(−300, −150)	2.18	99.20%	75.00%	42.74%	0.86	0.64+
8	(−450, −300)	1.09	70.30%	80.95%	67.21%	0.43	0.49+
Average		-	90.33%	81.41%	57.60%	0.77	0.62+

[1]Best interval indicates which of the three subsets of EEG temporal features was selected. Negative value means before keystroke in units of milliseconds. The best interval was selected based on the training results with cross-validation, i.e., the interval during which the training accuracy with cross-validation is the highest was selected.
[2]CR is the best criterion (CR $>$ 1) set in decision rule (6) which generated the highest d-prime. A CR $>$ 1 rendered the classifier more sensitive to error detection (higher hit rate) and susceptible to false alarms.
[3]Keenness (*d'*) varied using different CR. Here, the best *d'* (the point at which ROC curve was best separated from chance level) was listed with its corresponding hit and false alarm rates.
[4]AUC is the area under the ROC curve; the $+/-$ sign after AUC value signifies whether the AUC increased/decreased comparing with when EEG only set was used.

TABLE VII
CLASSIFICATION RESULTS FOR EEG + T&C SET

Subject	**Best Interval**[1]	**CR**[2]	**Training Accuracy**	**Prediction (Query)**			
				Hit Rate	**False Alarm**	**Keenness**[3]	**AUC**[4]
2	(−300, 150)	1.17	75.80%	93.75%	71.90%	0.95	0.58+
3	(−450, −300)	1.15	68.90%	96.15%	82.84%	0.82	0.58+
5	(−150, 0)	1.00	74.30%	54.55%	24.76%	0.80	0.65+
6	(−150, 0)	1.20	71.60%	94.74%	86.52%	0.52	0.58−
7	(−150, 0)	1.24	87.70%	81.82%	49.91%	0.91	0.64+
8	(−450, −300)	1.20	68.60%	95.24%	92.81%	0.21	0.49+
Average		-	74.48%	86.04%	68.12%	0.70	0.59+

[1]Best interval indicates which of the three subsets of EEG temporal features was selected. Negative value means before keystroke in units of milliseconds. The best interval was selected based on the training results with cross-validation, i.e., the interval during which the training accuracy with cross-validation is the highest was selected.
[2]CR is the best criterion (CR $>$ 1) set in decision rule (6) which generated the highest d-prime. A CR $>$ 1 rendered the classifier more sensitive to error detection (higher hit rate) and susceptible to false alarms.
[3]Keenness (*d'*) varied using different CR. Here, the best *d'* (the point at which ROC curve was best separated from chance level) was listed with its corresponding hit and false alarm rates.
[4]AUC is the area under the ROC curve; the $+/-$ sign after AUC value signifies whether the AUC increased/decreased comparing with when EEG only set was used.

speech rates, finger typing strategies, and urgency levels of typing, were inputs to the behavioral model but not used to generate task features. Thus, an extra dummy feature (C) was coded to signify experimental conditions (e.g., 1 = fast speech, multifinger typing, and nonurgent condition) and added to EEG + Task set. The classification results using this new dataset (EEG + T&C) are shown in Table VII. Adding information of experimental conditions improved keenness by 16.6% and increased AUC by 5.4%. There was a trend for the pair-wise comparison: EEG only versus EEG + T&C ($\mathrm{t_{df1,df2}} = -2.22$; $\mathrm{p} = 0.077$). The ROC curves of the LDA classifier using five different feature sets are plotted in Fig. 6. From Fig. 6, the ROC curves of the EEG + Model sets are separated from the chance level (diagonal line in the graphs) and the curves of the EEG only sets, while the ROC curves of the EEG + T&C sets almost overlap the curves of the EEG + Task sets. ROC curves of LDA classifiers using the "behaviors only" set generated poor results. Therefore, the enhanced QN-MHP model transformed task-relevant information into distinct features that were otherwise unavailable from EEG and task features. Combining modeling features with EEG features improves classification results because interactions between experimental conditions and task characteristics were considered based on advanced top-down inference and integrated with bottom-up real-time analysis on psychophysiological feedback.

VI. DISCUSSION

Integrating modeling features produced by an advanced human behavior model with real-time psychophysiological (EEG) features could produce better error prediction results by an LDA classifier. In contrast, adding features derived from apparent task characteristics and even environmental conditions barely benefitted data mining. Integration of data mining and human

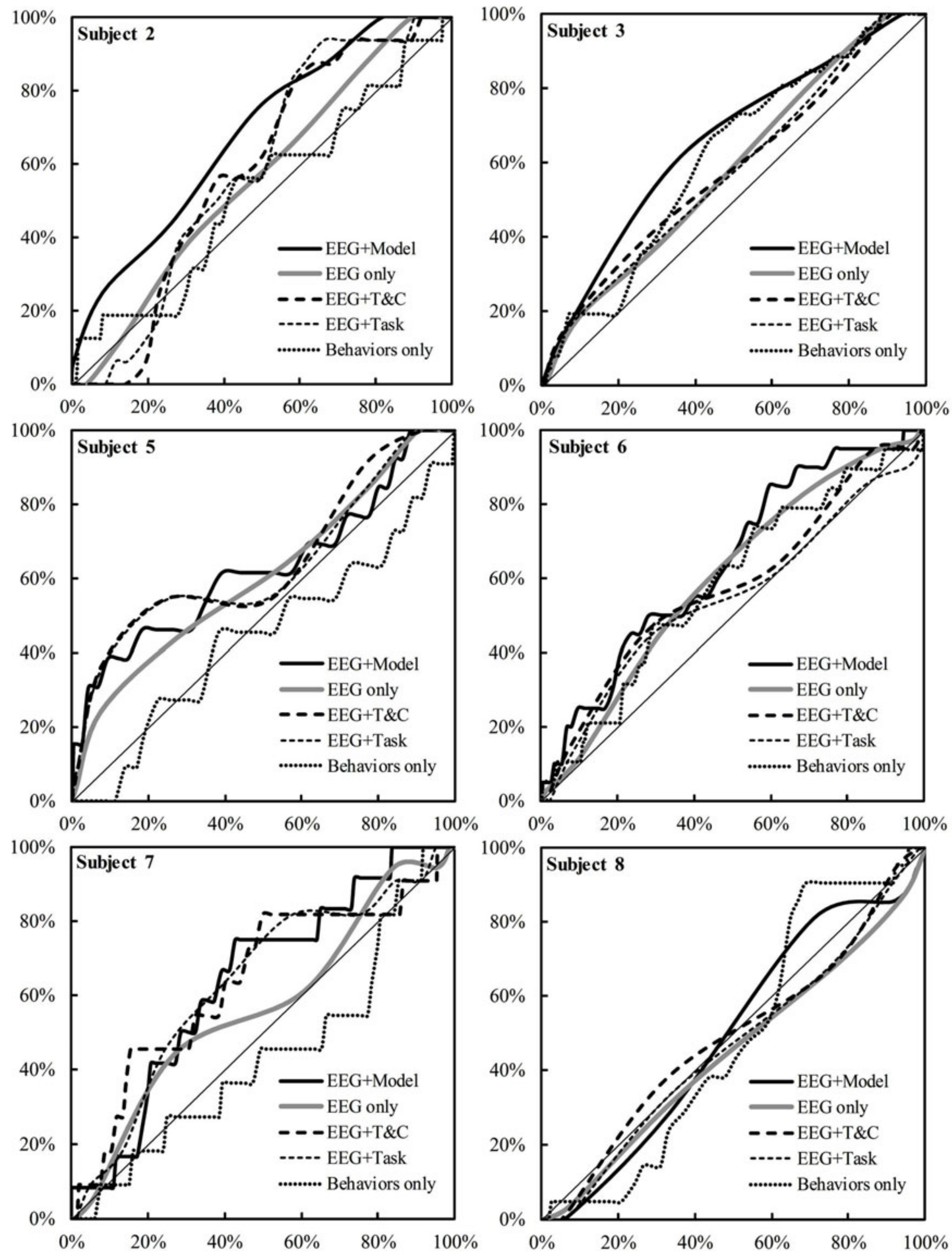

Fig. 6. ROC curves of the LDA classifier using different feature sets.

behavior modeling utilized both 1) top-down inference to transform interactions between experimental conditions and task characteristics into inclination of an average operator to make errors and 2) bottom-up analysis to parse real-time EEG that is indicative of psychophysiological status into likelihood of making errors on a trial-by-trial basis to generate informative features. The advantage of integration was observed by comparing results from using EEG only, behaviors only, and EEG + Model feature sets. Neither using EEG only nor behaviors only features alone was enough for error prediction. A combination of both real-time psychological feedback (EEG) and derived behavioral signals (modeling features) produced better results.

The integration is compatible with the notion of analysis by synthesis approach, and it was attempted before by using inverse models for reconstructing intracranial activity to help increase accuracy in EEG recognition for BCI applications [39]. Our enhanced QN-MHP is somewhat like an inverse model for EEG classification associated with errors, because it reconstructed what happened in perceptual, cognitive, and motor subnetworks (see Fig. 3) given task conditions and generated informational indices (predicted movement time, interference levels, and spatial variability) for error making in general. EEG, on the other hand, provided real-time information about motor preparation and execution as they varied with time. Our novelties, comparing with using inverse models, were to utilize both observable human behaviors (predicted movement time and spatial variability) and nonobservable behaviors (cognitive interference level) by prediction and compare their effectiveness with apparent features available without having a model. Applications of data mining techniques to analyze error-related EEG patterns generally provide 5 to 10% improvements in performance [29], [30], and this study demonstrated integration of modeling and EEG features can give another 28.3% increase in keenness and 10.7% increase in AUC, superior to 4% increase when an inverse model was used to reconstruct brain activities [39]. Therefore, the human behavior model essentially provided data mining classifier some "internal messages" that could not be obtained through simply processing task characteristics or environmental conditions.

The current study used a linear classifier and achieved a comparable performance to [22] in terms of AUC (the current study: 0.63 versus the previous study: 0.64). The previous study, however, achieved the best performance by using a sophisticated SVM classifier and all 36 EEG channels. The current study was able to achieve a similar performance by using straightforward temporal features (EEG amplitudes) from only six EEG channels with help of modeling features. The results thus demonstrated that modeling features possess considerable potential as supplementary information resource to the formation of human errors even if a simple linear data classifier is used. One limitation of the current study is that the enhanced QN-MHP for typing may not work for other tasks because it was tuned for predicting typing behaviors on physical keyboards. Typing on touch-based keyboards, for example, may cause alternation in behaviors in that lack of tactile feedback from key press may increase visual demand, and our model in its current status will not make good predictions for that. This does not necessarily mean that integration cannot be generalized to other tasks, but another effective human behavior model must exist for a particular situation, or the enhanced QN-MHP needs to be expanded to cover predictions of other tasks.

Nevertheless, advanced morphological features and frequency-related features, as well as other data mining techniques such as SVMs or ANNs, may be worth trying with other contrast enhancing methods to investigate the potential of integration of modeling features. In other BCI application studies, error detection/prediction mechanism did not work for some subjects (success in six out of seven subjects in [29]; seven out of eight subjects in [30]; five out of seven subjects in [31], [32]). Compared with those application studies, our method worked for all six subjects, i.e., using EEG + Model increased AUC from using EEG only. Subject 8 was the only one whose AUC was less than 0.5. This subject had 2.5 times as many typing errors as the most accurate subject. His relatively poor performance might indicate carelessness which could have caused less distinguishable psychophysiological feedback and resulted in outlier results in classification. Whether the performance was degraded due to decreasing engagement or increasing number of errors, however, needs to be confirmed in future studies where subjects' engagement can be measured.

References

[1] H. Thimbleby and P. Cairns, "Reducing number entry errors: Solving a widespread, serious problem," *J. Roy. Soc. Interface*, vol. 7, pp. 1429–1439, Oct. 2010.

[2] J. M. O'Hara, J. C. Higgins, and W. S. Brown, "Human factors considerations with respect to emerging technology in nuclear power plants," Brookhaven National. Laboratory., Upton, N.Y., USA, Rep. NUREG/CR-6947, 2008.

[3] N. Bohm, B. Gladman, I. Brown, C. Schaufler, and J. F. v. Schiller, "Banking and bookkeeping," in *Security Engineering: A Guide to Building Dependable Distributed Systems*, R. J. Anderson, Ed. 2nd ed. New York, NY, USA: Wiley, 2008, pp. 313–364.

[4] M. Falkenstein, J. Hoormann, S. Christ, and J. Hohnsbein, "ERP components on reaction errors and their functional significance: A tutorial," *Biol. Psychol.*, vol. 51, pp. 87–107, 2000.

[5] J. R. Fedota and R. Parasuraman, "Neuroergonomics and human error," *Theor. Issues Ergonom. Sci.*, vol. 11, pp. 402–421, 2010.

[6] T. A. Salthouse, "Perceptual, cognitive, and motoric aspects of transcription typing," *Psychol. Bull.*, vol. 99, pp. 303–319, 1986.

[7] C.-J. Lin and C. Wu, "Factors affecting numerical typing performance of young people in a hear-and-type task," *Ergonomics*, vol. 54, pp. 1159–1174, 2011.

[8] A. Subasi, "Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients," *Expert Syst. Appl.*, vol. 28, pp. 701–711, 2005.

[9] 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, pp. 1524–1533, 2008.

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

[11] W. J. Ray and H. W. Cole, "EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes," *Science*, vol. 228, pp. 750–752, 1985.

[12] T. Fernández, T. Harmony, M. Rodríguez, J. Bernal, J. Silva, A. Reyes, and E. Marosi, "EEG activation patterns during the performance of tasks involving different components of mental calculation," *Electroencephalogr. Clin. Neurophysiol.*, vol. 94, pp. 175–182, 1995.

[13] C. W. Anderson, E. A. Stolz, and S. Shamsunder, "Discriminating mental tasks using EEG represented by AR models," in *Proc. IEEE 17th Annu. Conf. Eng. Med. Biol. Soc.*, 1995, pp. 875–876.

[14] C. W. Anderson and Z. Sijerčić, "Classification of EEG signals from four subjects during five mental tasks," in *Proc. Conf. Eng. Appl. Neural Netw.*, 1996, pp. 407–414.

[15] P. Ferrez and J. d. R. Millán, "You are wrong!—Automatic detection of interaction errors from brain waves," in *Proc. Int. Joint Conf. Artif. Intell.*, 2005, pp. 1413–1418.

[16] 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 Recog.*, 2004, pp. 391–394.

[17] S. Suzuki, T. Matsui, Y. Sakaguchi, K. Ando, N. Nishiuchi, and M. Ishihara, "The possibility of determination of accuracy of performance just before the onset of a reaching task using movement-related cortical potentials," *Med. Biol. Eng. Comput.*, vol. 48, pp. 845–852, 2010.

[18] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model," *Expert Syst. Appl.*, vol. 32, pp. 1084–1093, 2007.

[19] K.-Q. Shen, C.-J. Ong, X.-P. Li, Z. Hui, and E. P. V. Wilder-Smith, "A feature selection method for multilevel mental fatigue EEG classification," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 7, pp. 1231–1237, Jul. 2007.

[20] 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, pp. 547–558, 1993.

[21] H. Shibasaki and M. Hallett, "What is the Bereitschaftspotential?" *Clin. Neurophysiol.*, vol. 117, pp. 2341–2356, 2006.

[22] S. Wang, C.-J. Lin, C. Wu, and W. A. Chaovalitwongse, "Early detection of numerical typing errors using data mining techniques," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 6, pp. 1199–1212, Nov. 2011.

[23] A. W. A. Van Gemmert and G. P. Van Galen, "Auditory stress effects on preparation and execution of graphical aiming: A test of the neuromotor noise concept," *Acta Psychologica*, vol. 98, pp. 81–101, 1998.

[24] A. W. A. Van Gemmert and G. P. Van Galen, "Stress, neuromotor noise, and human performance: A theoretical perspective," *J. Exp. Psychol. Human Perception Perform.*, vol. 23, pp. 1299–1313, 1997.

[25] G. P. Van Galen and M. Van Huygevoort, "Error, stress and the role of neuromotor noise in space oriented behavior," *Biol. Psychol.*, vol. 51, pp. 151–171, 2000.

[26] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, NY, USA: Wiley, 2000.

[27] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch, "A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals," *J. Neural Eng.*, vol. 4, pp. R32–R57, 2007.

[28] G. Dornhege, B. Blankertz, G. Curio, and K. R. Muller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun. 2004.

[29] R. Chavarriaga, A. Biasiucci, K. Forster, D. Roggen, G. Troster, and J. R. Millan, "Adaptation of hybrid human–computer interaction systems using EEG error-related potentials," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2010, pp. 4226–4229.

[30] B. Blankertz, C. Schäfer, G. Dornhege, and G. Curio, "Single trial detection of EEG error potentials: A tool for increasing BCI transmission rates," in *Artificial Neural Networks*, J. Dorronsoro, Ed. Berlin, Germany: Springer , 2002, pp. 138–138.

[31] L. Parra, C. Alvino, A. Tang, B. Pearlmutter, N. Yeung, A. Osman, and P. Sajda, "Linear spatial integration for single-trial detection in encephalography," *NeuroImage*, vol. 17, pp. 223–230, 2002.

[32] 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, Jun. 2003.

[33] S. A. Inverso, "Automatic error recovery using P3 response verification for a brain-computer interface," M.S. thesis, Comput. Sci. Dept., Rochester Inst. Technol., Rochester, NY, USA, 2004.

[34] H. Masaki, N. Takasawa, and K. Yamazaki, "Enhanced negative slope of the readiness potential preceding a target force production task," *Electroencephalogr. Clin. Neurophysiol.*, vol. 108, pp. 390–397, 1998.

[35] Y. C. A. Kita, A. Mori, and M. Nara, "Two types of movement-related cortical potentials preceding wrist extension in humans," *Neuroreport*, vol. 12, pp. 2221–2225, 2001.

[36] A. Hatta, Y. Nishihira, T. Higashiura, S. R. Kim, and T. Kaneda, "Long-term motor practice induces practice-dependent modulation of movement-related cortical potentials (MRCP) preceding a self-paced non-dominant handgrip movement in kendo players," *Neurosci. Lett.*, vol. 459, pp. 105–108, 2009.

[37] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. H. Lopes da Silva, "Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks," *NeuroImage*, vol. 31, pp. 153–159, 2006.

[38] C. Brunner, M. Naeem, R. Leeb, B. Graimann, and G. Pfurtscheller, "Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis," *Pattern Recog. Lett.*, vol. 28, pp. 957–964, 2007.

[39] P. Ferrez, "Error-related EEG potentials in brain–computer interfaces," Ph.D. dissertation, section de gÉnie Électrique et Électronique Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland, 2007.

[40] W.-Y. Hsu, C.-C. Lin, M.-S. Ju, and Y.-N. Sun, "Wavelet-based fractal features with active segment selection: Application to single-trial EEG data," *J. Neurosci. Methods*, vol. 163, pp. 145–160, 2007.

[41] A. Tang, M. Sutherland, and Y. Wang, "Contrasting single-trial ERPs between experimental manipulations: Improving differentiability by blind source separation," *NeuroImage*, vol. 29, pp. 335–346, 2006.

[42] B. Blankertz, G. Curio, and K. R. Mller, "Classifying single trial EEG: Towards brain computer interfacing," *Adv. Neural Inf. Process. Syst.*, vol. 14, pp. 157–164, 2002.

[43] C. Wu and Y. Liu, "Queuing network modeling of transcription typing," *ACM Trans.: Comput. Human Interaction*, vol. 15, pp. 1–45, 2008.

[44] D. E. Rumelhart and D. A. Norman, "Simulating a skilled typist: A study of skilled cognitive-motor performance," *Cognitive Sci.: Multidisciplinary J.*, vol. 6, pp. 1–36, 1982.

[45] B. E. John, "TYPIST: A theory of performance in skilled typing," *Human Comput. Interaction*, vol. 11, pp. 321–355, 1996.

[46] C.-J. Lin and C. Wu, "Mathematically modeling the effects of pacing, finger strategies and urgency on numerical typing performance with queuing network model human processor," *Ergonomics*, vol. 55, pp. 1180–1204, 2012.

[47] E. Koechlin and A. Hyafil, "Anterior prefrontal function and the limits of human decision-making," *Science*, vol. 318, pp. 594–598, Oct. 2007.

[48] P. Donkelaar and I. M. Franks, "The effects of changing movement velocity and complexity on response preparation: Evidence from latency, kinematic, and EMG measures," *Exp. Brain Res.*, vol. 83, pp. 618–632, 1991.

[49] R. J. van Beers, P. Haggard, and D. M. Wolpert, "The role of execution noise in movement variability," *J. Neurophysiol.*, vol. 91, pp. 1050–1063, Feb. 2004.

[50] M. L. Latash, S. Li, F. Danion, and V. M. Zatsiorsky, "Central mechanisms of finger interaction during one- and two-hand force production at distal and proximal phalanges," *Brain Res.*, vol. 924, pp. 198–208, 2002.

[51] A. Murata and H. Iwase, "Extending Fitts' law to a three-dimensional pointing task," *Human Movement Sci.*, vol. 20, pp. 791–805, 2001.

Cheng-Jhe Lin received the M.S. degree from the National Tsinghua University, Hsinchu, Taiwan, in 2002, and the Ph.D. degree from the State University of New York, Buffalo, NY, USA, in 2012.

He is currently an Assistant Professor of industrial management with the National Taiwan University of Science and Technology, Taipei, Taiwan. His research interests include applications of data mining techniques in cognitive engineering and integrated physical neuroergonomics methods in assessing usability of mobile computing devices.

Changxu Wu (S'04–M'07) received the B.S. degree in psychology with the focus on engineering and mathematical psychology from Zhejiang University, Hangzhou, China, in 1999, and the M.S. degree in engineering psychology and human–computer interaction from the Chinese Academy of Sciences, Beijing, China, in 2002. He also received the M.S. and Ph.D. degrees in industrial and operational engineering from the University of Michigan, Ann Arbor, MI, USA, in 2004 and 2007, respectively.

He is currently an Associate Professor with the Department of Industrial and System Engineering, State University of New York, Buffalo, NY, USA. His research interests include integrating cognitive science and engineering system design, especially modeling human cognition system with its applications in system design, improving transportation safety, promoting human performance in human–computer interaction, and inventing innovative sustainable and smart energy systems with human in the loop.

Wanpracha A. Chaovalitwongse (SM'11) received the Ph.D. degree in industrial and systems engineering from the University of Florida, Gainesville, FL, USA.

He is currently an Associate Professor of industrial and systems engineering and radiology (joint), and a Core Member of the Integrated Brain Imaging Center, University of Washington, Seattle, WA, USA. His research interests include optimization and machine learning in neurophysiological and imaging data.