Behaviour & Information Technology, 2015
Vol. 34, No. 8, 787–798, http://dx.doi.org/10.1080/0144929X.2014.945962

Taylor & Francis
Taylor & Francis Group

Predicting numerical data entry errors by classifying EEG signals with linear discriminant analysis

Cheng-Jhe Lin[a,b] and Changxu Wu[b*]

[a]*Industrial Management, National Taiwan University of Science and Technology, Taipei, Taiwan, Republic of China;*
[b]*Industrial and Systems Engineering, State University of New York at Buffalo, Buffalo, New York, USA*

(*Received 16 December 2011; accepted 8 July 2014*)

Daily numerical data entry is subject to human errors, and errors in numerical data can cause serious losses in health care, safety and finance. Difficulty in detecting errors by human operators in numerical data entry necessitates an early error detection/prediction mechanism to proactively prevent severe accidents. To explore the possibility of using multi-channel electroencephalography (EEG) collected before movements/reactions to detect/predict human errors, linear discriminant analysis (LDA) classifier was utilised to predict numerical typing errors before their occurrence in numerical typing. Single trial EEG data were collected from seven participants during numerical hear-and-type tasks and three temporal features were extracted from six EEG sites in a 150-ms time window. The sensitivity of LDA classifier was revealed by adjusting the critical ratio of two Mahalanobis distances as a classification criterion. On average, the LDA classifier was able to detect 74.34% of numerical typing errors in advance with only 34.46% false alarms, resulting in a sensitivity of 1.05. A cost analysis also showed that using the LDA classifier would be beneficial as long as the penalty is at least 15 times the cost of inspection when the error rate is 5%. LDA demonstrated its realistic potential in detecting/predicting relatively few errors in numerical data without heavy pre-processing. This is one step towards predicting and preventing human errors in perceptual-motor tasks before their occurrence.

Keywords: human errors; linear discriminant analysis; electroencephalography

1. Introduction

Predicting numerical typing errors in data entry tasks has important real-life applications. Numerical data are often typed via a computer keyboard, a touch screen or other kinds of numerical input interfaces, such as keypads of handheld mobile devices. While making some errors is trivial, for example, dialling a wrong number on a mobile phone, certain errors in numerical data can cause severe accidents and economic loss, especially when system responses are critical and not reversible (prescription dosage in medical databases, target coordinates in combat systems during wartime, etc.). Arndt et al. (1994) collected 688 forms from seven medical centres and discovered that 2.4% of the received digital data were wrongly typed (Arndt et al. 1994). Fatalities actually happened due to wrong numerical entries in a popular drug delivery system used in hospitals (Thimbleby and Cairns 2010). Financial transactions, nuclear power plants, aviation traffic controls and military applications are examples where numerical typing errors can be crucial (Young 1996; Lyons 2007; Bohm et al. 2008; O'Hara, Higgins, and Brown 2008). Furthermore, low reliability of human operators in checking and verifying numerical data makes the situation worse. Kawado et al. (2003) investigated data management in clinical studies by comparing two data verification methods, double data entry (DDE) and read-aloud (RA) method (Kawado et al. 2003). Astoundingly, the DDE method obtained only a 68.2% error detection rate if the verification was performed by a different operator (and 45.5% if the data were checked by the same operator who input the data). The RA method had even lower detection rates, 50% and 40.9% by a different and the same data entry operator, respectively. The study implied that at least half of the errors may remain undetected in the system. Some may argue that using error check codes may avoid human errors, but this technique is controversial. Paul and Neal (1989) recommended not to use any check digits for they may cause some confusion to the operator (Paul and Neal 1989).

Although numerical typing is error prone and numerical data entry errors are less detectable, they received relatively less attention in experimental studies (Logan 1982; Rumelhart and Norman 1982). The difficulty of detecting errors in numerical typing by operators is attributable to lack of top-down error detection. In alphabetical data, operators can detect errors immediately based on contextual clues, for example, knowing WROD is wrong. In numerical data, unless operators remember exactly what the

*Corresponding author. Email: seanwu@buffalo.edu

numbers are, error detection is limited. In particular, when data are presented verbally, operators will have no persistent reference unless they ask for the inputs again. These kinds of hear-and-type tasks prevail in daily life, for example, hearing a phone number from a friend and dialling on a mobile phone, telling an account number to a bank teller who types it out, and customer service agents inputting parcel tracking numbers obtained via phone calls. Due to difficulties in detecting and correcting errors in numerical data by typists themselves and potentially severe consequences resulting from errors, it is important to set up an error prevention system which can predict the errors proactively to avoid the occurrence of accidents and economic losses.

Salthouse (1986) reviewed major experimental findings in alphabetical transcription typing, and the results underlined the importance of early detection of human typing errors. He found that only 40–70% of errors can be detected without reference to the master copy. That is, 30–60% of errors may not be detected if error detection is made only by typists themselves. More importantly, the fact that not all errors are self-detectable suggested that error detection by typists themselves was faulty. He speculated that self-error detection was handled by a translation process of typing which converted perception codes of characters to movement specifications for fingers and hands. The error detection mechanism in the translation process compared efferent movement specifications with afferent response feedback, and signalled the typist about potential errors (Salthouse 1986). This error detection mechanism and error correction phenomenon in alphabetical typing have been successfully modelled by Wu and Liu (2008) in their Queuing Network-Model Human Processor (Wu and Liu 2008). Self-error detection, however, is not perfect and unable to detect errors that are originated from earlier processes, for example, motor preparation. In addition, the detection of errors was late in the typing process. It would be too late for the typist or the system to do any correction upon self-detection of potential errors because the movement has been executed and the output commands through the human–computer interface would probably have been executed. Thus, it is not ideal to rely on the imperfect self-error detection or its post hoc psychophysical responses, for example, error-related negativity (ERN) (Parra et al. 2002, 2003), to prevent numerical typing errors.

Fortunately, neurological research has shown that early detection of data entry errors is possible by looking at patterns revealed in electroencephalography (EEG). Prediction systems based on recognising abnormal patterns of multi-channel EEG have been studied in medical and engineering psychology fields based on the belief that state transitions of hidden patterns in EEG were indicative of the onset of clinical symptoms and detectable through the quantitative analysis of brain dynamics, for example, epileptic seizure prevention system (Chaovalitwongse, Prokopyev, and Pardalos 2006). In previous studies, researchers found linkages between EEG patterns and mental activities, for example, motor preparation and mental workload. A certain preparatory process for voluntary movements existed in primary motor cortex (M1) and supplementary motor areas (SMAs). The process could be represented by a slow cortical negative potential which preceded motor movements by 500–600 ms (Ikeda et al. 1996; Slobounov et al. 2005). Suzuki et al. (2010) further found that the negative slope component of the motor-related cortical potential (MRCP) which occurred approximately 500 ms before the movement could be used to differentiate movement accuracy because it was directly related to the mental efforts devoted to planning of required accuracy (Suzuki et al. 2010). Since the movement accuracy is directly related to inclination of making typing errors, detecting EEG patterns that are associated with potential errors in advance of their occurrences could become a potential solution to predict errors and prevent serious consequences caused by typing errors.

Using different data mining techniques, including linear and non-linear classifiers, EEG patterns in a single trial of perceptual-motor tasks can be quantitatively analysed to reflect different mental activities (Lotte et al. 2007). Among all data mining techniques, linear discriminant analysis (LDA) was widely used in online detection of errors because in general linear classifiers are more robust with fewer parameters to tune and less prone to overfitting (Dornhege et al. 2004; Bashashati et al. 2007). Garrett et al. (2003) used an LDA classifier to differentiate four cognitive tasks (mental arithmetic, composing, rotation and counting) from the resting status; with six electrodes they achieved 66% accuracy on average (Garrett et al. 2003). The result of this study showed the potential of LDA classifiers to distinguish mental status of different cognitive processes, but the experiment involved only mental tasks and no motor responses. In a self-paced keying task, Blankertz, Curio, and Mller (2002) adopted Sparse Fisher Discriminant to classify EEG signals and differentiated index finger movements from small finger movements (Blankertz, Curio, and Mller 2002). The accuracy reached 96.7% and 93.6% at 120 ms before keystrokes for filtered and non-filtered data, respectively. Despite high accuracy, their classification dealt with distinction between different motor movements and their experimental task involved no errors and no complex cognitive process. As for error detection, Parra et al. (2003) used linear discrimination to detect response errors after their occurrence in a forced choice visual discrimination task (Parra et al. 2003). Using 64 electrodes and 2 time windows of 100 ms, they were able to reach 0.79 ± 0.05 accuracy without eye-blink removal. Although they developed an online classifier, the detection was made 200 ms after the erroneous keystrokes due to its utilisation of ERN. The post-detection reduced its practicability in that corrections might be too late to change what has happened. Therefore, the potential of classifying EEG patterns in a single trial of perceptual-motor tasks should be investigated under more realistic settings of experimental tasks, and the focus should be detecting errors in advance so that the outcomes could be applied in establishing error prevention systems for numerical data entry.

In this study, the potential of LDA to detect numerical typing errors before their occurrence was investigated and analysed in terms of sensitivity. Participants were asked to complete eight trials of realistic hear-and-type tasks during which EEG data were collected. LDA classifiers were first trained by half of experimental data and then used to differentiate erroneous keystrokes from correct responses in the other half of unseen data. Then, performance of LDA classifiers was analysed by manipulating the ratio of 2 Mahalanobis distances (Appendix 1) to find the best sensitivity that could be achieved in different time windows. The manipulation demonstrated characteristics of the LDA classifiers in a trade-off, that is, how the LDA classifiers achieved a higher hit rate while maintaining reasonable number of false alarms. Finally, feasibility of LDA classifiers was evaluated in a cost analysis where the least gain/cost ratio to make using the LDA classifier beneficial was revealed.

2. Method

2.1. *Participants*

Seven male participants without any hearing disability were recruited from the student body of State University of New York at Buffalo, USA. All participants were right-handed and required to perform a preliminary test with their right hands. The preliminary numerical typing test was used to assure participants' familiarity with numerical data entry. They were given 30 nine-digit numbers, for example, 235645891, and instructed to type out those numbers with a recommended multi-finger typing pattern (Figure 1). The numbers were randomly generated, and participants were required to finish all 30 numbers within 200 s with at least 80% accuracy in terms of digits. This requirement is comparable to the representative numerical typing performance of skilled typists in the literature (see detailed comparison in Section 3.1) (Seibel 1977; Marteniuk, Ivens, and Brown 1996). An informed consent was obtained from each participant before participation, and all participants were compensated for their time.

2.2. *Experiment*

After each participant passed the preliminary test, two practice trials of the experimental task were given prior to formal experimental trials. The experimental task was a typical hear-and-type task which emulated daily numerical data entry tasks performed by bank tellers or phone operators. A computer program read out 30 nine-digit, randomly generated numbers without decimals in each trial and the participants were told to type out those numbers. Every digit in the number was read out separately without chunking two or three digits in a number, for example, '123' was read as 'one, two, three' instead of 'one twenty-three' or 'one hundred twenty-three'. In addition, there was a 300-ms pause in-between every three digits. The numbers were read out in this way because, based on an observation and interview by the authors in a pilot study, it was found to be the most natural way to read out numbers for a person without any prior knowledge about their specific formats. The interval between digits was 500 ms (Raanaas, Nordby, and Magnussen 2002), and there was a short pause of 2.5 s after each number, during which the participants were reminded to press the enter key. The user interface did no error correction and participants were instructed to not press the delete key because error correction interfered with data collection and might delay their typing speed. If the participant did not show any inability in hear-and-type tasks in practice trials, he continued with eight formal trials during which EEG data were collected. Participants were allowed to adjust the volume, posture and other settings of typing environment to their preference

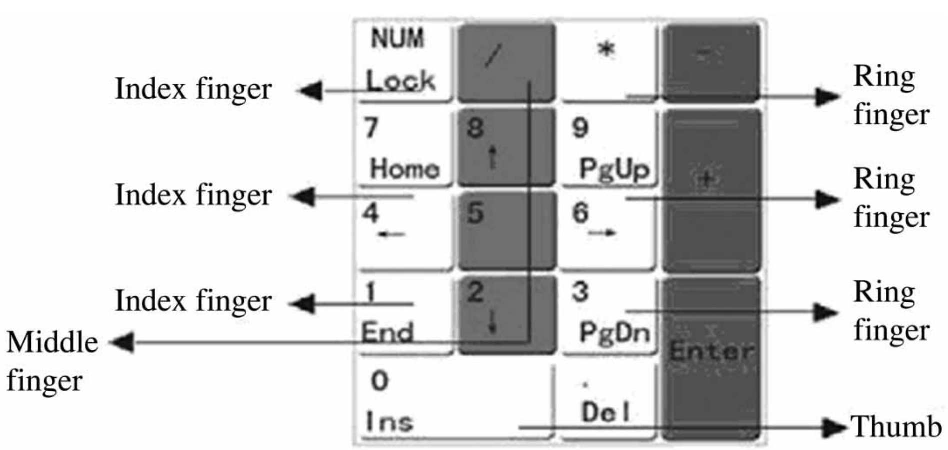

Figure 1. Recommended multi-finger typing pattern.

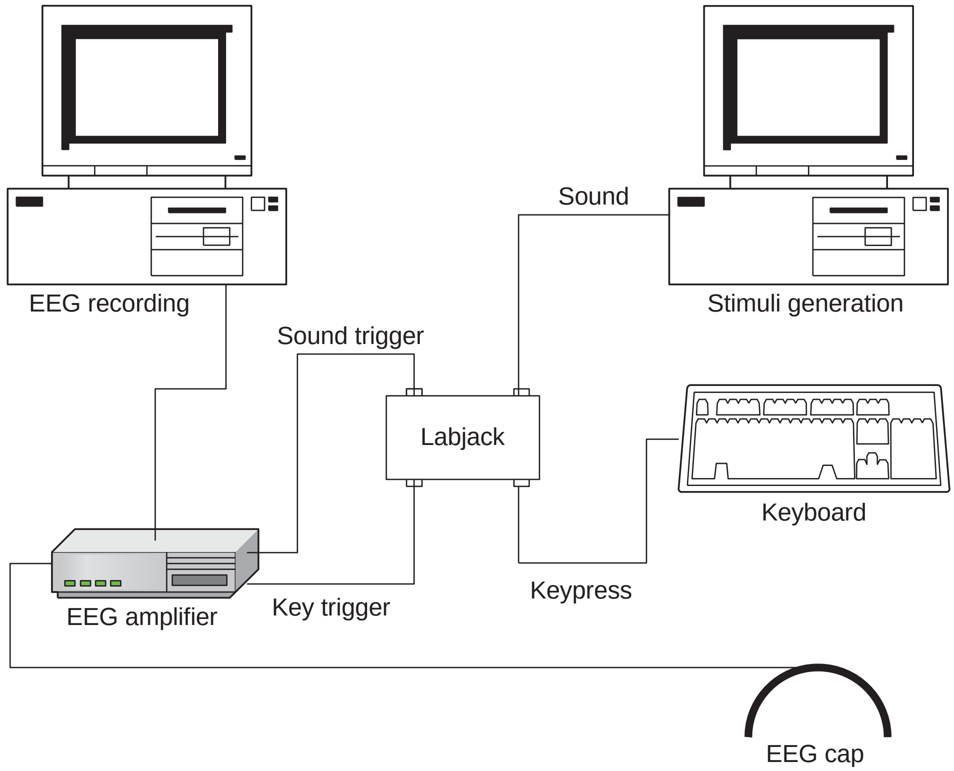

Figure 2. System structure for event recording.

before the experimental trials, and then the settings were kept constant through the whole experiment. A break of 10 min was provided to the participant after four trials. The participant's mental and physical fatigue level was monitored by using a subjective questionnaire after each trial. Controlling fatigue is necessary because experimental evidence showed that EEG patterns might be influenced by both physical and mental fatigue (Okogbaa, Shell, and Filipusic 1994; Ftaiti et al. 2010). The total experiment lasted about 40 min excluding preparation of EEG measurement, resting time and time spent in filling out questionnaires.

2.3. *Behavioural data collection and synchronisation*

Two separate computers (Figure 2) were used to generate auditory stimuli and to record EEG, respectively. During the experiment, the timing of each auditory stimulus was recorded by the stimuli generation computer, and a trigger signal coded as 'sound' was sent out simultaneously through a LabJack® interface (LabJack Corporation, Lakewood, CO, USA) to synchronise the EEG recording computer. Whenever a key was pressed, another synchronisation signal coded as 'key' was also sent to the EEG recording computer through the same interface. The timing of auditory stimuli, keystrokes and their correspondence (what number was presented and what key was pressed) were stored in a behavioural data file in the stimuli generation computer. The behavioural data were later analysed to differentiate erroneous keystrokes from correct responses. Whenever an auditory stimulus was responded by a wrong keystroke, that is, a sound 'one' was responded by pressing a '2' key, an error occurred. Those errors might be mistakes or slips according to Reason's categorisation, that is, the participants might interpret the auditory stimulus wrongly (mistakes) or inadvertently press the wrong key (slips) (Reason 1990). The erroneous keystrokes and correct responses were mapped to 'key' triggers in the continuous EEG recordings, and the 'key' triggers were further coded as either 'correct keystrokes' or 'erroneous keystrokes'. The 'key' triggers thus became time marks containing behavioural information based on which event-related discrete waves were extracted from the continuous EEG data.

2.4. *EEG data collections and processing*

During the experimental numerical typing task, EEG data were collected simultaneously with 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 NuAmps Express system (Neuroscan Inc., North Carolina, USA). Then, raw EEG data were processed through the following steps (Figure 3):

(1) The raw data collected at 1000 Hz sampling rate (i.e. 1000 data points measured in their amplitudes)

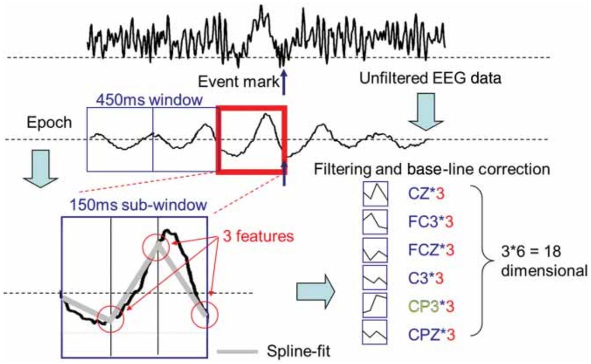

Figure 3. Pre-processing of EEG data.

were processed by a DC to 30 Hz bandpass filter (Pivik et al. 1993).

(2) The filtered data were epoched because only event-related potentials were of research interest. A segment of data (450-ms long, approximately the interval between stimuli) was extracted from each EEG channel, and the segment started at 450 ms before the keystroke (i.e. the end of the segment was right on the keystroke).

(3) The epoched data were baseline corrected (shifted based on the pre-stimulus EEG amplitude) to eliminate any signal drift because EEG signals can be possibly influenced by other electronic signals existing in the system or by individual factors that caused baseline amplitude to vary from person to person.

(4) The corrected epochs were spline fitted into 20 Hz, that is, using only 20 points to optimally capture the amplitude profile consisting of 1000 points (Blankertz, Curio, and Mller 2002). Therefore, each 450-ms window would have $450/1000 \times 20 = 9$ points.

(5) The down-sampled epochs (450-ms long) were divided into three sub-windows (150-ms long), each contained $9/3 = 3$ points. Those were three features used in the study, and only features from FC3, FCZ, C3, CZ, CP3 and CPZ electrodes were exported and further analysed by LDA.

FC3, FCZ, C3, CZ, CP3 and CPZ electrodes were chosen because they were located above the motor sensory cortex and close to the area in charge of right hand movements. Previous studies identified those electrodes as informational EEG sites in featuring space using linear discriminant analysis (Blankertz, Curio, and Mller 2002), and brain activities collected from those sites showed strong relations with MRCPs in both EEG (Kristeva-Feige et al. 2002; Fang et al. 2004) and focal transcranial magnetic stimulation studies (Mima et al. 2000). All data processes mentioned earlier were executed by Edit module of Scan 4.3 software (Neuroscan Inc., Charlotte, NC, USA).

The exported data files contained three features (in their time order) from each of six electrodes in 150-ms sub-windows and, therefore, formed a 3×6 matrix. A Visual Basic for Application (VBA®) program was coded to format each 3×6 matrix into a 1×18 column vector. Supposedly, if a classifier were used to classify later EEG data, the classifier should be trained by earlier EEG data. Based on this rationale, all vectors were equally[1] assigned into two different sets, a training set and a query set, based on their temporal order. Earlier data were assigned into the training set to train the classifier, while later data were assigned to the query set to validate the accuracy of the trained LDA classifier (see Appendix 1 for details of LDA analysis). All data in the query set were unseen by the LDA classifier, and classification results of the query set would be reported in terms of hit rates, false alarm rates and sensitivity. The hit rate was defined as the percentage of correctly classified data associated with erroneous responses, that is,

$$\text{Hit rate}(\%) = \left[\frac{\text{True positives}}{\begin{array}{c}(\text{True positives}+\\ \text{False negatives})\end{array}}\right] \times 100\%, \quad (1)$$

where true positives are events when erroneous responses are correctly classified as errors and false negatives are events when erroneous responses are incorrectly classified as correct reactions. In contrast, the false alarm was defined as the percentage of incorrectly classified data associated with correct responses, that is,

$$\text{False alarm}(\%) = \left[\frac{\text{False positives}}{\begin{array}{c}(\text{False positives}+\\ \text{True negatives})\end{array}}\right] \times 100\%, \quad (2)$$

where false positives are events when correct responses are incorrectly classified as errors and true negatives are events when correct responses are correctly classified as correct reactions. The sensitivity is the function of the hit rate and the false alarm, that is,

$$\text{Sensitivity} = Z(\text{hit}\%) - Z(\text{false alarm}\%), \quad (3)$$

where $Z(p)$ is the inverse function of standardised normal distribution, that is, $Z(p)$ returns the z-score that yields a cumulative probability of p. The sensitivity increases as hit rate increases and decreases as false alarm increases.

2.5. *Sensitivity analysis*

For each participant, EEG features were exported from three sub-windows (the 450-ms window in Figure 3 was segmented into three 150-ms sub-windows), representing psychophysiological responses from different time periods. EEG data from all three sub-windows were classified by LDA, but there was one sub-window where the LDA classifier had the best performance (the best sub-window). However, individual characteristics could affect the time period where the most relevant psychophysiological feedback could appear, and so the best sub-window to predict errors may vary. In addition, the decision boundary of the LDA classifier could be manipulated (Appendix 1). Increasing the sensitivity of the LDA classifier to the errors could raise the number of hits and false alarms at the same time. A sensitivity analysis was conducted to show the LDA classifier's performance in this trade-off, and to determine the best sub-window for the LDA classifier. The analysis found the highest d' value (the sensitivity for the receiver's operating characteristic curve, that is, $d' = Z(\text{hit\%}) - Z(\text{false alarm\%})$ (Wickens et al. 2012)), the LDA classifier could achieve in each sub-window by adjusting the ratio between the two Mahalanobis distances as a criterion of classification (Appendix 1). A VBA® program was coded to perform this analysis. The LDA classifier's performance in terms of d' as well as the best sub-window is reported in the result section.

3. Results

3.1. Pre-test and experimental typing performance

On average, participants achieved 96.6% accuracy while producing 131 keystrokes per minute in pretest. The typing speed of participants exceeded the representative number keying rates provided by Seibel (1977) (50–100 strokes per minute) (Seibel 1977), and the accuracy was comparable to another published study about numerical typing (about 2%) (Marteniuk, Ivens, and Brown 1996). Thus, the participants in this study could qualify as skilled typists. The average accuracy in experimental tasks was 98.8% and the average response time was 680 ms.

3.2. Sensitivity of LDA classifier

For the LDA classifier, the overall accuracy in training and query (prediction) phase was 80.8% and 65.64%, respectively (Table 1). The hit rate and the false alarm rate in predicting errors after adjustments of criterion based on sensitivity analysis are also listed in Table 1. The sensitivity of LDA classifier was calculated accordingly. The overall accuracy approached only about 66% because it was biased by the much greater number of correct keystrokes than erroneous ones. If high overall accuracy was pursued, the classifier could just classify all keystrokes into correct responses, resulting in nearly 99% accuracy. Apparently, however, all errors are missed in this case and the classifier is useless as an error predictor. On the other hand, the optimised LDA classifier helped to predict 74% errors in advance with only 35% of data checked. The sensitivity of the optimised LDA classifier was 1.05. Therefore, the LDA classifier was a much more effective error detection method than human operators using data verification methods such as RA or DDE, by which less than 70% of errors could be detected afterward after checking all data (Kawado et al. 2003).

3.3. Cost analysis of using LDA

The advantage of using LDA can also be manifested by a cost analysis. Suppose Co is the cost of inspecting a data point. Suppose also Cp and G are the penalty and the gain of missing and detecting an error, respectively. Accordingly, the expected profit (per keystroke) from using a classifier to predict errors can be formulated as

$$\begin{aligned} E(\text{Profits}) =& G \times \Pr(\text{Errors}) \times \Pr(\text{Hits}) \\ & - Cp \times \Pr(\text{Errors}) \times \Pr(\text{Misses}) \\ & - Co \times [\Pr(\text{Correct Responses}) \\ & \times \Pr(\text{False Alarms}) + \Pr(\text{Errors}) \\ & \times \Pr(\text{Hits})], \end{aligned} \tag{4}$$

where

Pr(Errors): Probability of errors, $0 < \Pr(\text{Errors}) < 1$

Table 1. Performance of LDA classifier.

Participant	Best sub-window[1] (ms)	CR[2]	Training accuracy (%)	Prediction hit rate (%)	Prediction false alarm (%)	Prediction accuracy (%)	Sensitivity (d')
1	−150 to 0	5.2	86.0	83.33	37.52	62.59	1.28
2	−150 to 0	2.6	87.3	71.43	22.46	77.50	1.32
3	−300 to −150	2.2	83.1	75.00	29.56	70.45	1.21
4	−150 to 0	1.1	65.4	64.29	40.79	59.41	0.60
5	−150 to 0	2.3	74.9	84.62	53.49	46.96	0.93
6	−450 to −300	1.7	87.4	75.00	20.6	79.37	1.49
7	−150 to 0	3.1	81.5	66.67	36.8	63.22	0.77
Average		–	80.8	74.34	34.46	65.64	1.05

[1]Negative values in this column signified the errors were detected in advance, for example, −150 to 0 ms means the errors were detected in the sub-window 150 ms before the keystrokes.
[2]CR: Criterion of classification in Section 2.5.

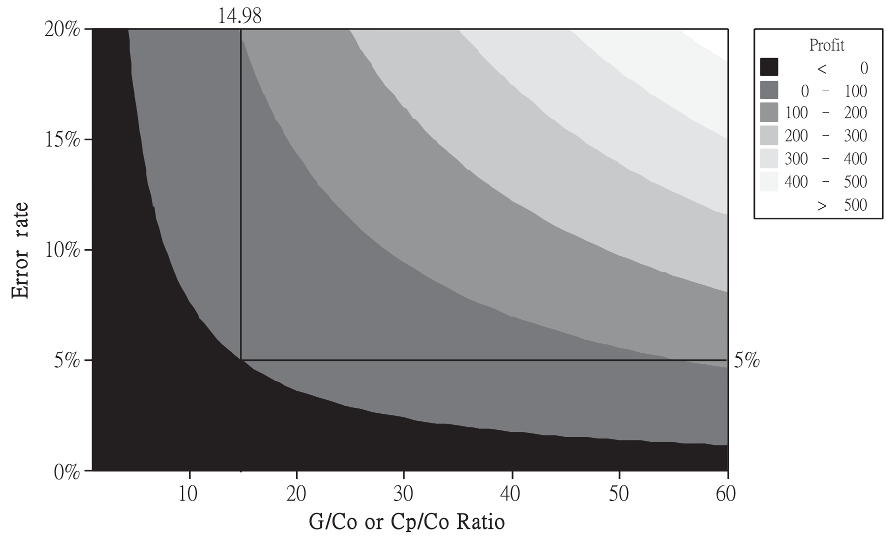

Figure 4. Contour plot of profits per 100 keystrokes using LDA classifier.

Pr(Correct Responses): Probability of correct responses, 0 < Pr(Correct Responses) < 1
Pr(Hits): Probability of hits, 0 < Pr(Hits) < 1
Pr(Misses): Probability of misses, 0 < Pr(Misses) < 1
Pr(False Alarms): Probability of false alarms, 0 < Pr(False Alarms) < 1

And the unit of E(Profits) can be any arbitrary unit, for example, money or time. Intuitively $G/\text{Co} = \text{Cp}/\text{Co}$ can be assumed (the ratio of gain over cost equals the ratio of penalty over cost). Given Pr(Hits), Pr(Misses) and Pr(False Alarms) for each classification method, E(Profits) is a function of typing accuracy (1-Pr(Errors)) and the ratio of penalty over cost. Figure 4 shows E(Profits) per 100 keystrokes as a function of Cp/Co and Pr(Errors).

Based on Figure 4, profits obtained of using LDA classifiers increases with the error rate and the ratio of penalty over cost. Even under a low error rate, the expected profit for using LDA can be still positive if the penalty of missing an error is high enough, that is, in a critical situation. For example, under the situation where typing accuracy is 95% and error rate is 5%, profits of using LDA will be positive if the penalty is at least 15 times the cost of inspecting a data point, i.e.

$$E(\text{Profits}) > 0 \text{ if } \frac{\text{Cp}}{\text{Co}} = \frac{G}{\text{Co}} > 15. \tag{5}$$

In contrast, taking a random guess would always result in a negative E(Profits) value since there was identical chance of hits and misses, but there was always a cost for inspecting data entries. The ratio (15 times) is not unreasonable since a typing error in a high digit number can easily produce a wrong number that is far bigger than its original value. For example, the cost of checking one digit may be 1 dollar. A wrongly typed second digit in a three-digit number, for example, 2‘5’0 is wrongly typed into 2‘8’0, can cause a penalty that is 30 times larger than the cost, and this could be common because the ‘8’ key is just next to the ‘5’ key.

4. Discussion

This study is one step towards predicting human errors in perceptual-motor tasks before their occurrence. Detecting errors in advance from such unbalanced data (rather few errors among numerous correct responses) itself is a very challenging topic, and an LDA classifier was proved satisfactory in predicting numerical data entry errors. Using EEG signals from six electrodes, LDA classifiers were able to detect 74% of numerical typing errors in advance by checking only less than 35% of data. Errors could be detected as early as 300 ms before the keystroke (for Participant 6). Although it may seem impractical to wear a 40-channel EEG cap while typing in a real work, the cap used in laboratory is not necessarily required for the implementation of the future error prediction device. The current study demonstrated that satisfactory results could be achieved by using only six electrodes, and a light-weight, headphone-like EEG device can be developed for this application. In fact, a wireless headband with four EEG electrodes has been devised for brain–computer interfacing (Lin et al. 2010). As technology advances, such wearable computing devices will become available very soon. Therefore, the simplicity of implementing LDA without heavy signal pre-processing demonstrated in this study makes it a possible proactive solution for typing error problems in reality.

Since this study focused on a new application of a linear data mining technique (LDA) to potentially detect error-associated EEG patterns before the occurrence of erroneous reactions, the methodology has some merits in regard to easy implementation to online applications.

While utilisation of more than 20 electrodes in other single trial EEG classification studies meant high dimensionality of data, the current study used only data from six EEG channels. Heavy pre-processing, including eye-blink removal and signal/noise ratio enhancement, was also required in other studies (Blankertz, Curio, and Mller 2002; Parra et al. 2002), but not in this study. In Parra et al. (2003), error detection was made at least 200 ms after keystrokes. Serious accidents could have happened if no online error detection warning systems were implemented because erroneous responses could have been executed at the time and not reversible. The high accuracy (over 90%) reported by Blankertz, Curio, and Mller (2002) can be partially attributable to their self-paced experimental task and high dimensionality of data. The motor readiness potential (BP) was found to be more pronounced in self-paced task (Ikeda et al. 1996), while our experimental task is a force-paced typing. In reality, non-self-paced tasks are commonly under time pressure and more subject to human errors. Also, Garrett et al. (2003) studied the effect of dimensionality by re-analysing data from Blankertz, Curio, and Mller (2002). With only six electrodes, the accuracy of classification based on the same data could have decreased from almost 95% to 76% in the best case (Garrett et al. 2003). Therefore, high accuracy achieved through high dimensionality in data analysis and low cognitive demand of experimental tasks would be subject to reduction in reality if high dimensionality is not available or greater task complexity is required. In contrast, the current study used low dimensional data and an emulated daily work, and, therefore, the outcome would be expected to be more applicable to real-world settings.

In terms of low dimensionality of data and high diversity of experimental tasks, a comparable work might be Garrett et al. (2003) in which six electrodes were also used and 66% accuracy was achieved. Our study used the same number of electrodes and obtained similar accuracy in a more realistic numerical typing task without any strenuous data processing. More electrodes could have been used to increase the accuracy. Better outcomes, however, are not guaranteed in that not all electrodes contribute pertinent information based on Blankertz, Curio, and Mller (2002). Also, it does not make much sense to include those electrodes associated with movements on the left side of the body, since all participants used their right hands to complete tasks. Indeed, removal of eye-blink artefacts or heavy noise filtering could have helped, but, again, pre-processing is deleterious to online applications and degrades benefits of timely detection. After all, pre-processing and high dimensionality are the causes of long latency. To facilitate development towards online error detectors/classifiers, higher prediction/detection rates should not be pursued at the expense of requiring heavy pre-processing or high dimensionality of collected data.

A potential application of the LDA classifier could be its integration to an online buffering system for checking suspicious data entry. Critical control systems such as nuclear power plants or missile launching systems may implement an input buffer to store suspicious data entries without executing them immediately because quick responses from human operators are error prone as the error probability declines over time (Andersen and Burns 1988). The results of this study showed that potential errors can be detected 300 ms before the key is pressed, that is, the error-associated EEG patterns are collected from a sub-window of 450–300 ms before the keystroke. Using this 300-ms lead time, the LDA classifier may be able to finish the classification, inform the system about possible errors in the incoming entries and enforce the input buffer. Upon acceptance of the input, the system still provides feedback to the operator acknowledging successful data entry, but the input will not be immediately executed unless the data entries in the input buffer are validated. During the validation latency, the operator may already have a second thought or become aware of errors that have been made through the signal, and, thus, the error is more likely to be corrected. Operators' vigilance and alertness can also be raised due to receiving correct-positive warnings and reacting to potential errors, resulting in higher robustness of the system.

On the other hand, if the system is designed to accept all responses (commands) at the time when they are entered, erroneous responses (commands) are likely executed at the time. Unless there is some kind of reverse mechanism to undo/cancel previously executed commands, the execution may not be stoppable; part of damage may have been done. For example, operators in a nuclear power plant may need to enter the depth of the controlling rods into the reactor to regulate its reaction speed, and a confirmation prompt may follow each data entry of the depth immediately. If the EEG after reactions was used for warning, the operator would confirm the command right away because there is no warning before reaction and would not find potential erroneous data entry until 200 ms later the online EEG data classifier says so. In this case, even if the execution is reversible, the process must take time and may cause instability of the system.

Several limitations associated with participants and real-time applications existed in the current study. First, our participants were recruited from the university student body and, thus, whether their perceptual-motor skills were comparable to a real work force is questionable. Based on the age (29.2-year old on average) and the expertise level (more than 100 keystrokes produced with more than 95% accuracy) of our participants, the current outcomes might be generalised to young skilled operators (around 30 years old). Older workers with reduced perceptual-motor capabilities were expected to have different psychophysiological responses which may cause discrepancy in outcome from the current study. Previous studies showed increased beta activity which is generally attenuated during active motor movements in EEG of older persons (Tucker et al.

1990; Polich 1997). Researcher also found delayed latency and altered lateralisation of MRCP that is indicative of motor preparation with ageing (Feve, Bathien, and Rondot 1991; Labyt et al. 2004). How these functional changes would influence the formulation and effectiveness of LDA classifiers, however, still needs investigation.

Second, number of participants (7) in this study seemed small at first glance. Yet, due to difficulties in recruiting participants and complex experimental settings, studies in EEG single trial analysis tended to use less than 10 participants (Blankertz, Curio, and Mller (2002), 3; Parra et al. (2003), 7; Garrett et al. (2003), 5). Seven participants recruited for the current study were comparable to other studies and should be satisfactory.

Finally, our method is an offline analysis and still subject to uncertainty of the best sub-window to choose. This uncertainty also implies a necessity to formulate a specific LDA classifier for each participant. However, several individual characteristics may affect the formulation of the LDA classifier such as motor control strategy in speed-accuracy trade-off, skill level revealed in response time and even surface condition of scalp. The approach demonstrated in this study did not suggest the best sub-window to be analysed and used for LDA training beforehand. One solution to this uncertainty is to establish a robust function by weighting classification results from different sub-windows so that the hit rate will not vary dramatically if a poor sub-window is chosen. However, given the consistency of skilled typists in inputting numbers, the best sub-window to predict future errors should be stable and could be obtained by collecting a typist's profile of classification accuracy over time. Selection of optimal criterion for distance ratio can be achieved through a similar procedure.

To keep practicality of the method, the current study used a relatively simple way (spline fitting) to capture the profile of EEG data. Statistical or geometric features may be considered in future research to better capture the EEG profile and may lead to superior classification results. Furthermore, LDA is not the only method used in the field of single-trial EEG analysis, and, thus, future studies could focus on more advanced pattern recognition methods such as Regularised FLDA, Support vector machine (SVM) and k-Nearest Neighbourhood (kNN). Other non-parametric methods are also being investigated in this field, but their applications to detect errors in real-world tasks are few. Future work might focus on benchmarking those advanced methods and improving the current method towards an online and robust classifier.

5. Conclusion

To investigate the potential of data mining techniques in predicting human errors, LDA was utilised to detect numerical data entry errors before keystrokes were made. Based on EEG data from FC3, FCZ, C3, CZ, CP3 and CPZ electrodes, LDA classifiers achieved a 74.34% detection rate on numerical typing errors by checking only less than 35% of data, resulting in a sensitivity of 1.05. A cost analysis revealed that using LDA is beneficial as long as the gain of finding an error is at least 15 times the cost of checking a data point. The use of LDA classifier to detect errors is also far more effective than data verification by human operators, in which only 70% of errors may be detected with 100% inspection. Based on the results of this study, online implementation of LDA classifiers could provide a proactive solution to error prevention for perceptual-motor tasks.

Acknowledgements

The authors thank an anonymous English speaker for her help with the preparation of the manuscript.

Funding

This research is partially funded by a National Science Foundation (NSF) grant.

Note

1. Whether there were equal numbers of data in the training and query the 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 was an odd number of correct/incorrect responses, the number of data in the query set would be one more than in the training set.

References

Andersen, V. M., and E. T. Burns. 1988. "Human Error Probability Models in the BWR Individual Plant Evaluation Methodology." Conference record for 1988 IEEE fourth conference on Human Factors and Power Plants, Monterey, CA, USA, June 5–9. IEEE. 323–342.

Arndt, S., G. Tyrrell, R. F. Woolson, M. Flaum, and N. C. Andreasen. 1994. "Effects of Errors in a Multicenter Medical Study: Preventing Misinterpreted Data." *Journal of Psychiatric Research* 28 (5): 447–459.

Bashashati, A., M. Fatourechi, R. K. Ward, and G. E. Birch. 2007. "A Survey of Signal Processing Algorithms in Brain–Computer Interfaces Based on Electrical Brain Signals." *Journal of Neural Engineering* 4 (2): R32–R57.

Blankertz, B., G. Curio, and K. R. Mller. 2002. "Classifying Single Trial EEG: Towards Brain Computer Interfacing." *Advances in Neural Information Processing Systems (NIPS 2001)* 14 (1): 157–164.

Bohm, N., B. Gladman, I. Brown, C. Schaufler, and J. F. V. Schiller. 2008. "Banking and Bookkeeping." In *Security Engineering: A Guide to Building Dependable Distributed Systems*, edited by Anderson, R. J. 2nd ed., 313–364, Indianapolis, IN: Wiley.

Chaovalitwongse, W. A., O. A. Prokopyev, and P. M. Pardalos. 2006. "Electroencephalogram (EEG) Time Series Classification: Applications in Epilepsy." *Annals of Operations Research* 148 (1): 227–250.

Dornhege, G., B. Blankertz, G. Curio, and K. R. Muller. 2004. "Boosting Bit Rates in Noninvasive EEG Single-Trial

Classifications by Feature Combination and Multiclass Paradigms." *Biomedical Engineering, IEEE Transactions on* 51 (6): 993–1002.

Duda, R. O., P. E. Hart, and D. G. Stork. 2000. "Maximum-Likelihood and Bayesian Parameter Estimation." *Pattern Classification*, edited by Duda, R. O., P. E. Hart and D. G. Stork. 2nd ed., 84–159. New York, NY: John Wiley & Sons.

Fang, Y., V. Siemionow, V. Sahgal, F. Xiong, and G. H. Yue. 2004. "Distinct Brain Activation Patterns for Human Maximal Voluntary Eccentric and Concentric Muscle Actions." *Brain Research* 1023 (2): 200–212.

Feve, A. P., N. Bathien, and P. Rondot. 1991. "Les potentiels corticaux liés au mouvement de l'homme âgé." *Neurophysiologie Clinique/Clinical Neurophysiology* 21 (4): 281–291.

Ftaiti, F., A. Kacem, N. Jaidane, Z. Tabka, and M. Dogui. 2010. "Changes in EEG Activity Before and After Exhaustive Exercise in Sedentary Women in Neutral and Hot Environments." *Applied Ergonomics* 41 (6): 806–811.

Garrett, D., D. A. Peterson, C. W. Anderson, and M. H. Thaut. 2003. "Comparison of Linear, Nonlinear, and Feature Selection Methods for EEG Signal Classification." *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 11 (2): 141–144.

Ikeda, A., H. O. Lüders, T. F. Collur, R. C. Burgess, H. H. Morris, T. Hamano, and H. Shibasaki. 1996. "Subdural Potentials at Orbitofrontal and Mesial Prefrontal Areas Accompanying Anticipation and Decision Making in Humans: A Comparison with Bereitschaftspotential." *Electroencephalography and Clinical Neurophysiology* 98 (3): 206–212.

Kawado, M., S. Hinotsu, Y. Matsuyama, T. Yamaguchi, S. Hashimoto, and Y. Ohashi. 2003. "A Comparison of Error Detection Rates Between the Reading Aloud Method and the Double Data Entry Method." *Controlled Clinical Trials* 24 (5): 560–569.

Kristeva-Feige, R., C. Fritsch, J. Timmer, and C.-H. Lu¨Cking. 2002. "Effects of Attention and Precision of Exerted Force on Beta Range EEG-EMG Synchronization During a Maintained Motor Contraction Task." *Clinical Neurophysiology* 113 (1): 124–131.

Labyt, E., W. Szurhaj, J. L. Bourriez, F. Cassim, L. Defebvre, A. Destée, and P. Derambure. 2004. "Influence of Aging on Cortical Activity Associated with a Visuo-Motor Task." *Neurobiology of Aging* 25 (6): 817–827.

Lin, C. T., L. W. Ko, M. H. Chang, J. R. Duann, J. Y. Chen, T. P. Su, and T. P. Jung. 2010. "Review of Wireless and Wearable Electroencephalogram Systems and Brain-Computer Interfaces—A Mini-Review." *Gerontology* 56 (1): 112–119.

Logan, G. D. 1982. "On the Ability to Inhibit Complex Movements: A Stop-Signal Study of Typewriting." *Journal of Experimental Psychology: Human Perception and Performance* 8 (6): 778–792.

Lotte, F., M. Congedo, A. Lecuyer, F. Lamarche, and B. Arnaldi. 2007. "A Review of Classification Algorithms for EEG-Based Brain–Computer Interfaces." *Journal of Neural Engineering* 4 (2): R1–R13.

Lyons, P. B. 2007. "The Human Factor in Nuclear Safety." *Joint IEEE conference on human factors and power plants and workshop on human performance.* Washington, DC: US Nuclear Regulatory Commission, Office of Public Affairs, 7. http://pbadupws.nrc.gov/docs/ML0724/ML072410234.pdf

Marteniuk, R. G., C. J. Ivens, and B. E. Brown. 1996. "Are There Task Specific Performance Effects for Differently Configured Numeric Keypads? *Applied Ergonomics* 27 (5): 321–325.

Mima, T., J. G. Steger, A. E. Schulman, C. Gerloff, and M. Hallett. 2000. "Electroencephalographic Measurement of Motor Cortex Control of Muscle Activity in Humans." *Clinical Neurophysiology* 111 (2): 326–337.

O'Hara, J. M., J. C. Higgins, and W. S. Brown. 2008. *Human Factors Considerations with Respect to Emerging Technology in Nuclear Power Plants.* Washington, DC: US Nuclear Regulatory Commission, Office of Public Affairs.

Okogbaa, O. G., R. L. Shell, and D. Filipusic. 1994. "On the Investigation of the Neurophysiological Correlates of Knowledge Worker Mental Fatigue Using the EEG Signal." *Applied Ergonomics* 25 (6): 355–365.

Parra, L., C. Alvino, A. Tang, B. Pearlmutter, N. Yeung, A. Osman, and P. Sajda. 2002. "Linear Spatial Integration for Single-Trial Detection in Encephalography." *NeuroImage* 17 (1): 223–230.

Parra, L. C., C. D. Spence, A. D. Gerson, and P. Sajda. 2003. "Response Error Correction-a Demonstration of Improved Human-Machine Performance Using Real-time EEG Monitoring." *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 11 (2): 173–177.

Paul, P., and R. W. Neal. 1989. "Error Detecting Decimal Digits." *Communications of the ACM* 32 (1): 106–110.

Pivik, R. T., R. J. Broughton, R. Coppola, R. J. Davidson, N. Fox, and M. R. Nuwer. 1993. "Guidelines for the Recording and Quantitative Analysis of Electroencephalographic Activity in Research Contexts." *Psychophysiology* 30 (6): 547–558.

Polich, J. 1997. "EEG and ERP Assessment of Normal Aging." *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* 104 (3): 244–256.

Raanaas, R. K., K. Nordby, and S. Magnussen. 2002. "The Expanding Telephone Number Part 2: Age Variations in Immediate Memory for Multiple-Digit Numbers." *Behaviour & Information Technology* 21 (1): 39–45.

Reason, J. 1990. *Human Error*. New York, NY: Cambridge University Press.

Rumelhart, D. E., and D. A. Norman. 1982. "Simulating a Skilled Typist: A Study of Skilled Cognitive-Motor Performance." *Cognitive Science: A Multidisciplinary Journal* 6 (1): 1–36.

Salthouse, T. A. 1986. "Perceptual, Cognitive, and Motoric Aspects of Transcription Typing." *Psychological Bulletin* 99 (3): 303–319.

Seibel, R. 1977. "Data Entry Devices and Procedures." In *Human Engineering Guide to Equipment Design*, edited by Kinkade, R. G. and H. S. V. Cott., 311–344. Washington, DC: US Government Printing Office.

Slobounov, S., M. Hallett, S. Stanhope, and H. Shibasaki. 2005. "Role of Cerebral Cortex in Human Postural Control: An EEG Study." *Clinical Neurophysiology* 116 (2): 315–323.

Suzuki, S., T. Matsui, Y. Sakaguchi, K. Ando, N. Nishiuchi, and M. Ishihara. 2010. "The Possibility of Determination of Accuracy of Performance Just Before the Onset of a Reaching Task Using Movement-Related Cortical Potentials." *Medical & Biological Engineering & Computing* 48 (9): 845–852.

Thimbleby, H., and P. Cairns. 2010. "Reducing Number Entry Errors: Solving a Widespread, Serious Problem." *Journal of The Royal Society Interface* 7 (51): 1429–1439.

Tucker, D. M., J. G. Penland, H. H. Sandstead, D. B. Milne, D. G. Heck, and L. M. Klevay. 1990. "Nutrition Status and Brain Function in Aging." *The American Journal of Clinical Nutrition* 52 (1): 93–102.

Wickens, C. D., J. G. Hollands, R. Parasuraman, and S. Banbury. 2012. *Signal Detection and Absolute Judgment. Engineering Psychology and Human Performance.* 4th ed., Upper Saddle River, NJ: Pearson.

Wu, C., and Y. Liu. 2008. "Queuing Network Modeling of Transcription Typing." *ACM Trans. Comput.-Hum. Interact.* 15 (1): 1–45.

Young, D. M. 1996. *Data Inaccuracy in the Global Transportation Network*. Wright-Patterson Air Force Base, OH: Air Force Institute of Technology Wright-Patterson.

Appendix 1. Fundamentals of LDA

Figure A1 shows the concept of an LDA classifier in a two-dimensional space and the concept should be generalised in a $k \times p$-dimensional feature space. In this study, an LDA classifier is a hyper-plane in $k \times p$-dimensional space, where $k = 3$ (features in a sub-window) and $p = 6$ (channels) which separates all data points (x) optimally into two subsets: one subset associated with correct numerical entries, and the other subset associated with erroneous ones.

Given k features for p electrodes in an epoch of a 150-ms sub-window, EEG data relevant to an event (a keystroke in this study) could be represented by a $k \times p$-dimensional column vector $\mathbf{x}$ ($k = 3$ and $p = 6$ in this study). Now, if there are n_1 correct responses in the first set $\mathbf{D_c} = \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_{n1}}\}$ and n_2 error responses in the second set $\mathbf{D_e} = \{\mathbf{x_{n1+1}}, \mathbf{x_{n1+2}}, \ldots, \mathbf{x_{n1+n2}}\}$, the definition of $\mathbf{S_B}$, the 'between classes scatter matrix', can be given as

$$\mathbf{S_B} = (\mathbf{m_c} - \mathbf{m_e})(\mathbf{m_c} - \mathbf{m_e})^{\mathbf{t}}, \tag{A1}$$

where $\mathbf{m_c}$ is the $k \times p$-dimensional sample mean of the $\mathbf{x}$ in the set $\mathbf{D_c}$, that is,

$$\mathbf{m_c} = \frac{1}{n_1} \sum_{\mathbf{x} \in \mathbf{D_c}} \mathbf{x}. \tag{A2}$$

And similarly, $\mathbf{m_e}$ is the $k \times p$-dimensional sample mean of the $\mathbf{x}$ in the set $\mathbf{D_e}$:

$$\mathbf{m_e} = \frac{1}{n_2} \sum_{\mathbf{x} \in \mathbf{D_e}} \mathbf{x}. \tag{A3}$$

Also the definition of the 'within classes scatter matrix' $\mathbf{S_W}$ can be given as follows:

$$\mathbf{S_W} = \mathbf{S_c} + \mathbf{S_e}, \tag{A4}$$

where

$$\mathbf{S_c} = \sum_{\mathbf{x} \in \mathbf{D_c}} (\mathbf{x} - \mathbf{m_c})(\mathbf{x} - \mathbf{m_c})^{\mathbf{t}} \tag{A5}$$

and

$$\mathbf{S_e} = \sum_{\mathbf{x} \in \mathbf{D_e}} (\mathbf{x} - \mathbf{m_e})(\mathbf{x} - \mathbf{m_e})^{\mathbf{t}}. \tag{A6}$$

The objective of LDA is then to maximise the following criterion function:

$$J(\mathbf{w}) = \frac{\mathbf{w^t S_B w}}{\mathbf{w^t S_W w}}, \tag{A7}$$

where $\mathbf{w}$ is the direction of the hyper-plane which separates data into two subsets. This criterion function is a matrix representation of the following function:

$$J = \frac{|\tilde{m}_c - \tilde{m}_e|}{\tilde{S}_c^2 + \tilde{S}_e^2} \tag{A8}$$

where $\tilde{m}_c$ and $\tilde{m}_e$ are, respectively, the sample means for the projected points in the correct response subset and the error subset, i.e.

$$\tilde{\mathrm{m}}_{\mathrm{c}} = \frac{1}{n_1} \sum_{\mathbf{x} \in \mathbf{D}_c} \mathbf{w}^t \mathbf{x}, \tag{A9}$$

$$\tilde{\mathrm{m}}_{\mathrm{e}} = \frac{1}{n_2} \sum_{\mathbf{x} \in \mathbf{D}_e} \mathbf{w}^t \mathbf{x}, \tag{A10}$$

And $\tilde{S}_c$ and $\tilde{S}_e$ are, respectively, the scatters for the projected points in the correct response subset and the error subset, that is,

$$\tilde{\mathrm{S}}_{\mathrm{c}} = \sum_{\mathbf{x} \in \mathbf{D}_c} (\mathbf{w}^t \mathbf{x} - \tilde{\mathrm{m}}_{\mathrm{c}})^2 \tag{A11}$$

$$\tilde{\mathrm{S}}_{\mathrm{e}} = \sum_{\mathbf{x} \in \mathbf{D}_e} (\mathbf{w}^t \mathbf{x} - \tilde{\mathrm{m}}_{\mathrm{e}})^2 \tag{A12}$$

Therefore, a $\mathbf{w}$ maximising $J(\mathbf{w})$ actually means a hyper-plane that maximises the distance between the projected sample means of the two subsets while minimising the scatters. Then, it can be proved that the optimal $\mathbf{w}^*$ which maximises criterion function $\boldsymbol{J}$ will be (Duda, Hart, and Stork 2000)

$$\mathbf{w}^* = \mathbf{S_W^{-1}}(\mathbf{m_c} - \mathbf{m_e}). \tag{A13}$$

And given $\mathbf{w}^*$, the optimal decision boundary has the equation:

$$\mathbf{w^t x} + \mathbf{w_0} = 0. \tag{A14}$$

The abovementioned procedure can be visualised in Figure A2. The separating hyper-plane is obtained by seeking the projection that maximises the distance between the two class means and minimises the intra-class variance. If the direction of the plane can be rotated in a way that the mean of projected data points on the plane would be maximal between groups and the variation of

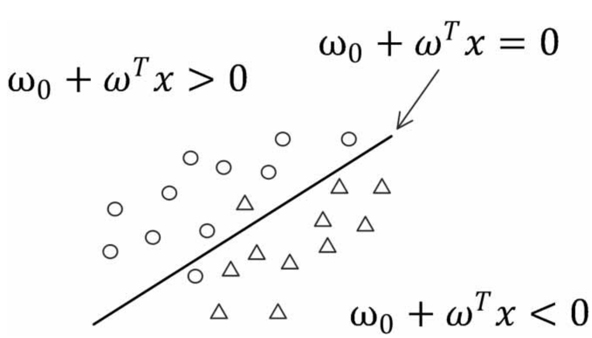

Figure A1. LDA classifier (Lotte et al. 2007).

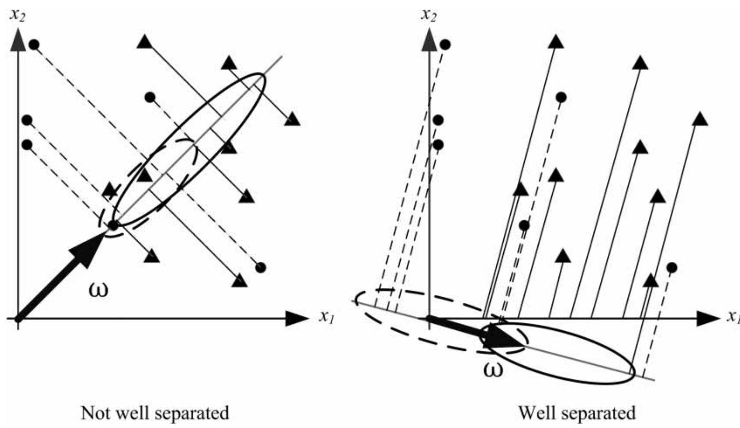

Figure A2. Visualisation of LDA procedure (Duda, Hart, and Stork 2000).

projected points would be minimal within groups, the orientation of the hyper-plane would be optimal, and the criterion of classification can be determined accordingly. For more details, refer to Duda, Hart, and Stork (2000).

To obtain an initial LDA classifier, half of the data points are first assigned into a training set. The data in the training set are used to train the LDA classifier and the trained LDA classifier is used to classify data in the query set for validation.

In the current study, the training set contained data from the first half of the experiment (data collected earlier in time). A statistical software package (Minitab Inc., Pennsylvania, USA) was used to produce the discriminant function for each participant from the training data. The data in the query set were then classified by calculating the values of the discriminant function for each observation $\mathbf{x}$ and assigning each data point to the group where it had the highest functional value. The software also reported the Mahalanobis distance (a distance that takes into account the correlations of the data set) of each data point to the group center, that is,

$$D(\mathbf{x}, \mathbf{m_i}) = (\mathbf{x} - \mathbf{m_i})^{\mathbf{t}} \Sigma^{-\mathbf{1}} (\mathbf{x} - \mathbf{m_i}); i = \mathbf{c} \text{ or } \mathbf{e}, \qquad \text{(A15)}$$

where Σ is the pooled covariance matrix for two groups. Since the Mahalanobis distance represents the similarity of observation $\mathbf{x}$ to the group centre $\mathbf{m}_i$, the rule of classification can be expressed by the following:

Classify $\mathbf{x}$ *into* $\mathbf{D}_c$ *if*
$D(\mathbf{x}, \mathbf{m_e})/D(\mathbf{x}, \mathbf{m_c}) > criterion$
else
Classify $\mathbf{x}$ *into* $\mathbf{D_e}$

Initially, the criterion of classification was set to 1.