

Detecting Emotional Stress during Typing Task with Time Pressure

Yee Mei Lim

Faculty of Applied Sciences and Computing
Tunku Abdul Rahman University College
Kuala Lumpur, Malaysia
Email: ymlim@acd.tarc.edu.my

Aladdin Ayesh

Faculty of Technology
De Montfort University
Leicester, United Kingdom
Email: aayesh@dmu.ac.uk

Martin Stacey

Faculty of Technology
De Montfort University
Leicester, United Kingdom
Email: mstacey@dmu.ac.uk

Abstract—Automated stress detection is important in developing adaptive e-learning systems. Empirical evidence suggests that mouse dynamics and keyboard dynamics analyses can be both effective in user behaviour modelling as well as emotion detection compared to physiological measures and facial expression recognitions, and yet they are far less expensive and considered non-intrusive. This paper investigates how mouse dynamics and keyboard dynamics can be affected by emotional stress, particularly stress induced by time pressure, text length and language familiarity. Our research findings show that longer text and unfamiliar language raise users' stress perceptions. Demanding job such as long typing task could result in anomalous behaviours once the users have lost motivation. Language familiarity mainly affects keyboard behaviour but text length change mouse behaviour. This shows that there is a good potential of developing an adaptive e-learning system by detecting learners' emotional stress from keyboard and mouse input.

Keywords—emotional stress; keyboard dynamics; mouse dynamics; language familiarity; text length

I. INTRODUCTION

Cognitive load theory (CLT) is an instructional theory that studies human knowledge or cognitive architecture to devise novel, instructional procedures [1]. Therefore it is one of the most influential theories in instructional design [2]. Based on CLT, we can design different measurement techniques to assess mental or cognitive efficiency of a learner or to diagnose the learner's instructional condition in order to enhance e-learning, such as providing adaptive learning materials, assessing the learner's performance, or to provide customized assessment [3]. This research presents a preliminary study assessing the potential of developing an automatic analysis of learner's emotional stress from keyboard and mouse input. It is important to utilize a non-intrusive and cost-effective method to measure learner's mental state or cognitive load so that the development of an adaptive e-learning system would become more feasible.

The difficulty of our research is there is lack of ground-truth to define emotional stress. Merely using user self-report survey could be easy but it can be erroneous and unreliable, since many people may not even be aware that they are stressed. To be able to measure human stress more effectively, some psychological theories suggest that in a task-specific environment, user stress levels can be varied according to two factors: demand (e.g. excessive demand on worker production, especially to meet a deadline) and control (e.g. lack of control over work processes) [4]. Besides, misfit between job

demands and individual capabilities intensifies the stress effect [5]. Therefore, by deliberately changing the workloads and control of tasks the user stress level can be changed.

To evoke users' stress in the experiments, there are some methods which are widely adopted, which include mental arithmetic, N-back number recall, time pressure, reading aloud, viewing affective picture or video, emotive text reading and story telling (see [6]–[9]). Among these methods, some are very useful to enable the job demands to be quantified or measured, for example mental arithmetic, N-back number recall and time pressure. However, in the e-learning environment, most of the jobs require text typing (e.g. post discussion). Therefore it is necessary to examine the text effect (the job demands can be varied by text length) on users' stress perception. Besides, familiarity of the subject matter can affect the students' performance during online-assessment. According to Tobias et al [10], "content with unfamiliar subjects require a more overt response for optimal learning than does familiar subject matter, as lack of familiarity implies that the responses required by the task may not be in the student's repertory". To simulate the familiar and unfamiliar environment, we utilize language familiarity in typing tasks. A research by Hulme et al [11] shows that memory spans for unfamiliar words was lower than the familiar words, due to a depression in the intercept of the recall-speech rate function. Therefore we believe that language familiarity may affect user behaviour (such as typing rhythms) to an extent, even the effect may be small.

To enable us to measure emotional stress that can be evoked in a typing task, we conduct a preliminary research on how keyboard and mouse dynamics could be affected by time pressure, increased job demand (text length) and language familiarity using empirical method. Section II presents the background study of keyboard and mouse dynamics analyses. Section III outlines the design and methodology of this research. Section IV shows the statistical results. Section V provides some in-depth discussions of the results. Lastly, Section VI concludes the paper.

II. BACKGROUND

Much research of human-computer interaction (HCI) has been done to measure or detect changes of stress level of a person. To classify stress versus non-stress conditions, besides the traditional social science research methods, the most common approaches are physiological measurements [8], [12]–[17] and

facial expressions recognition [9], [18]–[23]. Although both physiological measures and facial expression recognition have high accuracy rates, but the assessments could be obtrusive, requiring additional equipment (which can be costly), and are often labour or computationally intensive [6], [24]. To eliminate the downsides of both methods above, several research utilized mouse behaviour or keyboard behaviour analyses. To model user behaviour using keystroke dynamics, various attributes that can be extracted from the typing pattern of a user must be collected and examined, which include how quickly they type (keystroke speed), the duration of a key is pressed (Down-Up time), the latency between successive key-presses (Down-Down time), the key-press ordering, the usage of shift keys, deletion keys (backspace key and delete key) [25]. In a research by Vizer et al [6], which monitored users' cognitive and physical stress using keystroke dynamics and free-text analysis, the results show that users demonstrated different typing behaviours such as changes of patterns in key latency, keystroke speed, delete keys, navigation keys and other keys (such as letter and number keys). Therefore stress detection through keystroke dynamics and free-text analysis is possible. Some research has been carried to study the effects of stress on influencing users' physiological reactions when using a mouse. Research by Wahlström et al [14] showed that users would demonstrate increased physiological and psychological reactions using a computer mouse if they work under time pressure and verbal provocation. Similarly, Heiden et al [15] proved that if a user works under time pressure and precision demands, it would have caused a decrease in forearm muscle oxygenation when using a mouse. Some research use mouse dynamics analyses to model user behaviour. Pusara and Brodley [26] used mouse dynamics to detect anomalous behaviour through a user's mouse movement. Tsoulouhas et al [27] detected boredom of students caused by the presentation of a course. Their research demonstrated that mouse movement speed, inactivity occurrences and durations, and movement directions would change significantly for users who reported boredom in contrast to those who did not. The research also shows the best results (with FAR 2.7586%) in which the movements of the mouse are recorded with the intervals of 10 seconds. Although many people may not relate boredom to stress, nevertheless when users feel 'rustout', or under stimulation, it has a very negative effect, often producing boredom, fatigue and dissatisfaction, which is also known as understress [28], [29]. Therefore, we believe that user would demonstrate different behaviour when they work under stress conditions, and the changes of the mouse dynamics and keyboard dynamics could be detected.

Many limitations would exist when analysing mouse dynamics or keyboard dynamics alone. Unlike the human biometric measures such as finger print and iris, mouse and keyboard convey an unstructured and very small amount of information. Moreover, this information may vary not only because of the intrinsic variability of behavioural characteristics, but because different devices can be used, different environmental conditions exist, and when the users are working on one device, the other would be idle for long time [30]. In order to deal with the instability of the mouse movements and typing behaviours, there are several restrictions that must be done. First, the users must use normal, external and common mouse and keyboard devices during the experiments. Secondly, it is important to

control the text length in order to examine its effect on the typing rhythms. Most research in the field of keystroke analysis limit the experiments to produce samples from structured and predefined text. Many researchers strived to work with relatively short sample phrases, such as username and password [31]–[37]. Nevertheless there were a few research used free and long text in user authentication research. For instance Gunetti and Picardi [30] obtained samples of average 800 characters long and their result was reaching a False Alarm Rate (FAR) of less than 5% and an Impostor Pass Rate (IPR) of 0.00489%. Shimshon et al [38] used text samples of 433 to 1034 keystrokes in their research and they obtained FAR of 0.41% and IPR of 0.63%. Therefore keyboard dynamics analysis could be equally effective by using both fixed text and free text, regardless the length of the text (provided the text length produces enough data for analysis).

The other problem of analysing keyboard dynamics is to deal with the variability of individual typing skills. According to Davidson et al [39], typing speed has been closely related to how far ahead a typist looks in the text while typing. The typing speed will increase if the typist is able to look far ahead, which allows superior preparation and optimization of typing movement. Additionally, by exerting full concentration, individuals can increase their typing speed by 10-20%. When individuals engage in deliberated activities that are specially designed to increase their typing speed, such as setting time pressure, typists would usually find them too demanding. This practice involves actively pushing performance beyond its reliable limits, often lead to mistakes as old habits are broken.

III. METHODOLOGY

A. Formulation of User Behaviour, Mouse Behaviour and Keyboard Behaviour

Our research begins by hypothesizing that emotional stress can be detected through acquisition and analysis of three datasets, by inducing different stress levels through time pressure (Timing), and typing task (Question) with various text length (Text Length) and language familiarity (Familiarity). We identified the 3 datasets as user behaviour ($B(U)$), mouse behaviour ($B(M)$) and keyboard behaviour ($B(K)$).

Firstly, $B(U)$ is defined as follows:

$$B(U) = \langle TD, PA, Err, SP \rangle \quad (1)$$

TD = the duration to complete one task (milliseconds (ms))
 PA = Passive attempt that includes the attempt to give up
 Err = the number of typing errors, including missing words and punctuation marks, and spelling errors
 SP = the stress level score for each task, which is obtained from the user's self evaluation report (7-point Likert scale).

Secondly, $B(M)$, a dataset that captures the following mouse features in each task, is defined as follows.

$$B(M) = \langle MS, MID, MIO, MC \rangle \quad (2)$$

MS = Average mouse speed (pixels per ms)
 MID = Total mouse inactivity duration (ms)
 MIO = Total mouse inactivity occurrences
 $MC = \langle MCL, MCR^1 \rangle$, which is a dataset that consists of left click rate per ms (MCL) and right click rate per ms (MCR)

¹ MCR was removed later due to no data

Lastly we define the keyboard behaviour ($B(K)$) as a dataset that captures the keyboard features for each task, as follows:

$$B(K) = \langle KL, KS, EK \rangle \quad (3)$$

KL = Average keystroke latency (Down-Down time) (ms)

KS = Average keystroke speed per key (per second)

$EK = BSK + DK^2$, the total occurrences of error keys used (EK), which includes backspace (BSK) and delete (DK) keys

B. Experiment Setup

1) *Hardware and Software*: A program was written in Java to capture mouse raw data so that $B(M)$ can be computed. For every mouse click, the local time in milliseconds (ms) would be recorded. For every 10 ms, the mouse location would be captured and its respective time in milliseconds would be recorded. A separate program was written in VB.NET later in order to acquire the virtual-key codes generated by the Windows platform to compute $B(K)$. To protect user's privacy, the virtual-key codes were transformed into special codes automatically by the program. For instance, a number key or a letter key was recorded as "k", but the actual key-code pressed by the user was not stored. Once a key was pressed, the local time in milliseconds (ms) would be recorded. To simulate an online-assessment environment, an imitation of an online-assessment website was built. Six different typing tasks were set with different text length. Three questions were set in English (as familiar language), and the others were in German (as unfamiliar language). The requirements of the typing tasks are shown in Table I.

In terms of hardware, all the computers in the computer laboratory were equipped with Windows 7, 3.10GHz CPU, 4GB RAM, 17" monitor with the resolution of 1024x768 pixels, external standard QWERTY HID keyboard and external HID-compliant mouse. The imitation website would run on Google Chrome by default.

TABLE I. TYPING TASK REQUIREMENTS

| Question | Characteristics | | Text Length | |
|----------|-----------------|-------------|-------------|-------------------------|
| | Length | Familiarity | Words | Letters (without space) |
| 1 | short | familiar | 5 | 21 |
| 2 | short | unfamiliar | 5 | 25 |
| 3 | medium | familiar | 20 | 94 |
| 4 | medium | unfamiliar | 20 | 99 |
| 5 | long | familiar | 63 | 459 |
| 6 | long | unfamiliar | 63 | 451 |

2) *Procedures*: To determine the time limit to be given to the participants, we conducted a pilot test. From our initial pilot test of 13 samples, we obtained the average durations to complete Question 3 and Question 4 were about 26730 ms and 30602 ms respectively. The mean for them to complete Question 5 was 30247 ms, with 100% of them obtained more than 40 typing errors; while the mean to complete Question 6 was 24952 ms, with 76.92% of them obtained more than 40 typing errors, and the rest chose to give up. Therefore to enforce time pressure, we set 30 seconds time limit the experimental group to complete each task. The reason to set much longer text for Question 5 and Question 6 was to induce more stress under time pressure (which it is impossible to complete 63 words within 30 seconds without any error).

² DK was removed later due to no data

Longer text is also believed to lead to boredom, tiredness and fatigue, and this may ultimately results in ultimate stress point that the users can endure. We examine the chance of anomalies occurrences if ultimate stress point (or endurance limit) is reached.

All participants were required to run the experiments in a computer laboratory. They were given the same set of typing tasks. Before the participants started typing, instructions were displayed on the screen and they must provide their consensus in order to continue the experiments. Each question was displayed on individual page. On each page of the question, a "give up" button was given so that if the participants did not wish to continue the question they may choose to skip it. If a participant had chosen to give up, the number of PA would increase to one. When the participants were ready to start, they need to press the start button, and a unique code would be generated randomly and automatically by the website to represent their identities. Then the first question would be revealed so that they can start typing. The start time (in milliseconds) would be recorded automatically. Upon completion, the users must use the mouse device to submit the task. When the participants submitted the task, the end time (in millisecond) would be recorded, and a survey form would be displayed. They must complete the survey form to indicate their perceptions of how much stress they felt when typing the text (1 for strongly disagree and 7 for strongly agree). At the same time $B(U)$, $B(M)$ and $B(K)$ were computed based on the recorded raw data and stored in the database. The second question would be displayed next once they submitted the form, the start time would be recorded and the subsequent process would be repeated until the last question.

3) *The Control Group and Experimental Group*: Seventy-seven year-2 students from Bachelor Degree in Computer Science and Bachelor Degree in Information Technology were recruited based on voluntarily basis without any incentive. All of them passed the English test in Malaysian Certificate of Education, but none of them knew the German language. We filtered out 17 invalid samples (due to missing cases or outliers) and we analysed the data from 60 of them (aged between 18 to 24 years old, 90% male). The valid samples were divided into 2 groups. For control group, 30 of them were required to type the predefined texts of all 6 questions without any time constraint. The other 30 participants in the experimental group were given 30 seconds time limit for each question. The page would be submitted automatically if they could not complete the task on time.

IV. STATISTICAL RESULTS

We used Levene's test to ensure homogeneity between the two subject groups in our experiments. However due to Levene's test can be sensitive to detect even small departures from homogeneity and the assumption of normality [40], TD and MID were transformed using Log_{10} function. Besides, no one used right mouse click and delete key for the entire experiment. Therefore, we excluded the features of DK and MCR from the analyses. We started our analyses by examining the frequency of the users who had completed the tasks without any error ($Err = 0$). The frequency is given in Table II. The Analysis of Variance (ANOVA) [41] test shows that Err is significantly different according to Timing and Question (both

TABLE II. NUMBER OF USERS WITH ZERO MISTAKES ($Err = 0$)

| Question | 1 | 2 | 3 | 4 | 5 | 6 |
|-------------------------|----|----|----|----|----|----|
| Timing=no ($N = 30$) | 27 | 25 | 26 | 24 | 18 | 17 |
| Timing=yes ($N = 30$) | 19 | 16 | 15 | 8 | 12 | 14 |
| Total ($N = 60$) | 46 | 41 | 41 | 32 | 30 | 31 |

with $p = 0.00$). It is interesting to note that although we predicted that no one should be able to complete Question 5 and Question 6 in the experimental group, however the number of participants who scored $Err = 0$ is still high. In terms of the number of students who chose to give up during the typing task, we observed that the number did not increase until the participants reached Question 5, which 5% of them gave up Question 5 while 10% gave up Question 6 (all of them had attempted Question 1 to Question 4). It is also very interesting to note that we also observed that the students started to cheat from Question 5 onwards (where they copied-and-pasted the text directly to the text box instead of typing). This is further supported by statistical evidence. Based on the users who obtained perfect score ($Err = 0$) for Question 5 and Question 6, 95.08% of them used less than 20 seconds to complete the task (and 100% spent less than 30 seconds). Also, there were a total of 50 participants submitted Q5 and Q6 within 20 seconds, including those who were not given any time pressure. This phenomenon shows that when the users are given long text, even they are not pressured by time, the job will still be considered too demanding (SP is higher at Question 5 and 6). Therefore they are considered to have behaved anomalously at Question 5 and Question 6, possibly due to loss of motivation. Since PA and Err did not provide accurate data due to the anomalies in Question 5 and Question 6, therefore we excluded them from subsequent analyses.

We tested the main effects of the Question and Timing factors to SP using ANOVA. According to Figure 1, the lowest mean of SP is Question 1 ($\mu = 2.70$) and the highest is Question 6 ($\mu = 5.62$). SP increased when Question increased. However, Timing has no effect on SP and there is no interaction effect between Question and Timing ($p = 0.446$), which indicates that even the users were given time pressure, they had no different stress perceptions compared to those without time pressure. Although Timing does not change SP significantly, it does change some features of $B(M)$ and $B(K)$. Table III shows the effects of Timing factor to these features. Multivariate Analysis of Variance (MANOVA) [40] tests show that Timing significantly affects TD , MS , MIO , KS and KL , but it does not affect MID , MCL and EK . To analyse how

TABLE III. MANOVA TESTS OF THE BETWEEN-SUBJECTS EFFECTS

| Factor | Behaviour | Feature | p-value |
|--------|-----------|---------|---------|
| Timing | B(U) | TD | 0.0272 |
| | | SP | 0.6788 |
| | | MS | 0.0267 |
| | B(M) | MID | 0.6926 |
| | | MIO | 0.0000 |
| | | MCL | 0.4571 |
| | | KS | 0.0167 |
| | B(K) | KL | 0.0000 |
| | | EK | 0.1783 |

Highlighted cell indicates that the effect is significant at the level of $p < 0.05$

the variations between Question affects the $B(U)$, $B(M)$ and $B(K)$, we performed Tukey Post Hoc Test. The results show that Question is a significant effect on all features except MCL

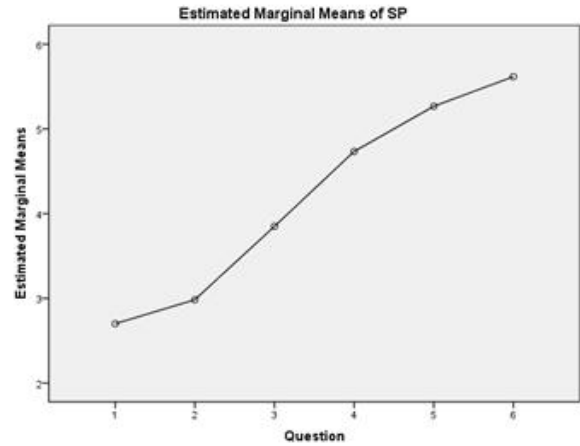


Fig. 1. Stress Perception (SP) increased according to Question. The differences between questions are significant at $p < 5e^{-18}$ level.

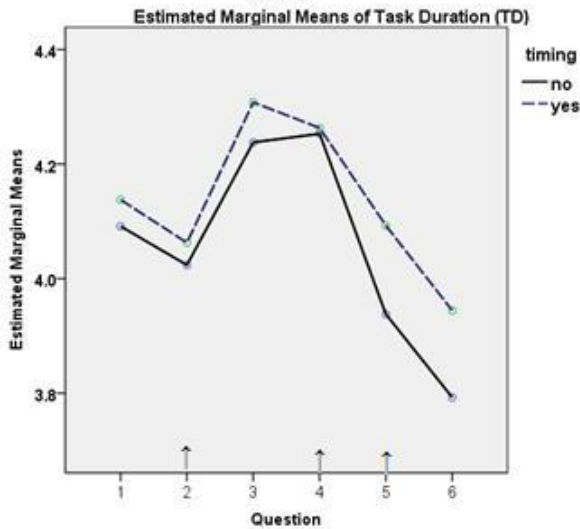
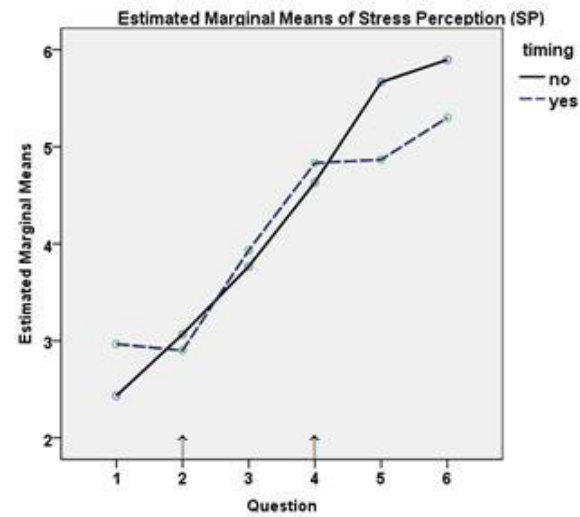


Fig. 2. Mean Plots of User Behaviour Features according to Question and Timing factors.

($p = 0.34$), as illustrated in Figure 2 to Figure 4. The arrow markers in each graph show the significant changeover point

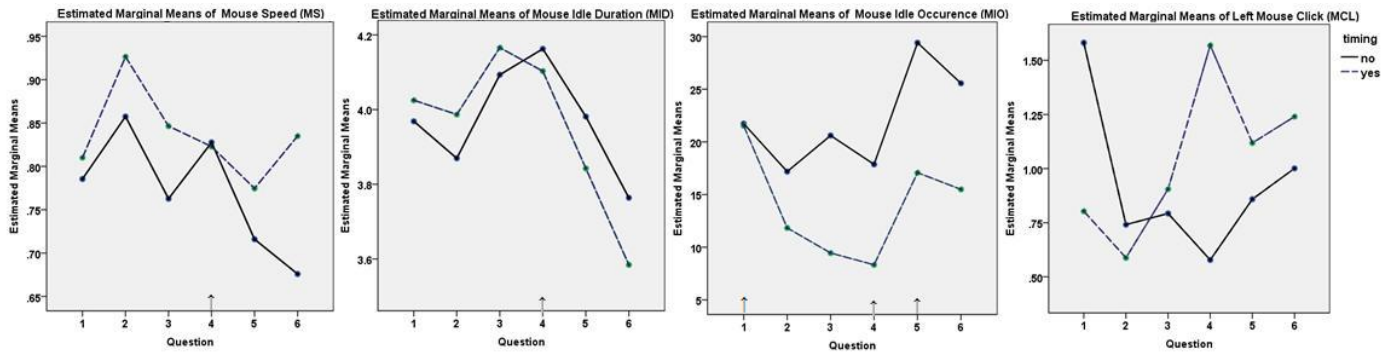


Fig. 3. Mean Plots of Mouse Behaviour Features according to Question and Timing.

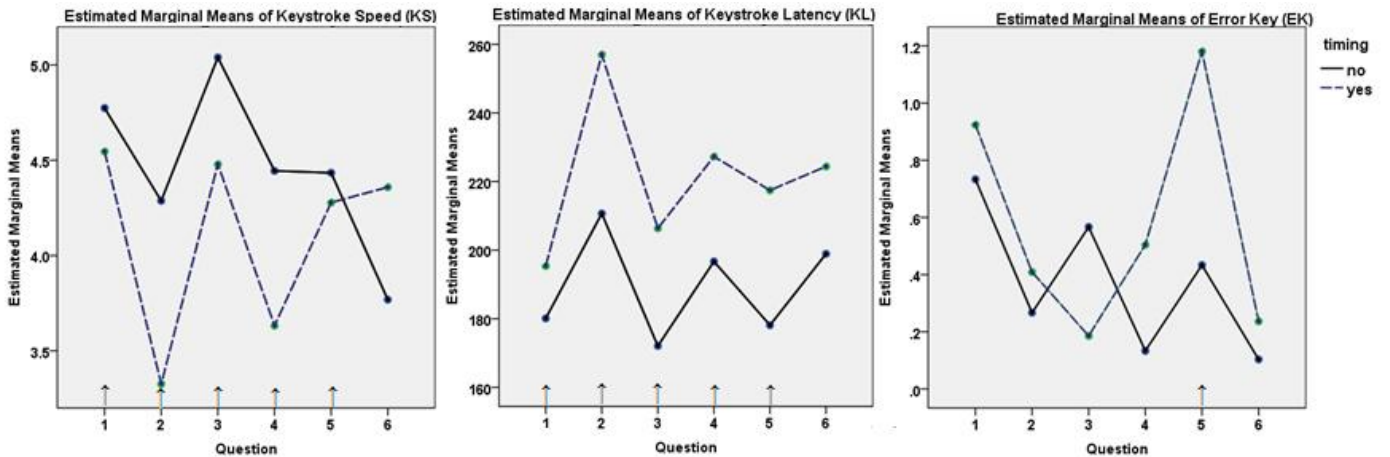


Fig. 4. Mean Plots of Keyboard Behaviour Features according to Question and Timing.

between classes. For instance, in the graph for *SP* in Figure 2, there are significant changes after Question 2 and Question 4. In the graph for *TD*, the duration increased significantly after Question 2, and then it decreased significantly after Question 4 and Question 5. To relate back to the anomalies occurred at Question 5, the graphs in Figure 2 and Figure 3 further supports our analysis. Besides anomalous results obtained in *PA* and *Err*, we could also observe anomalies occur after Question 4. As shown in Figure 2, *TD* drops despite increase of Text Length, Similarly in Figure 3, anomalous behaviours of *MS*, *MID* and *MIO* could be observed after Question 4. However no anomalies could not be seen from *B(K)*.

Since Question appears to be the only main effect that affects *SP*, we then analysed how the characteristics of the Question affect the behaviours. MANOVA tests were performed to verify the differences between Text Length and Familiarity. Besides verifying the significant effects of the two factors, we examined their interaction effects. Table IV shows the degree of confidence of each factor affecting *B(U)*, *B(M)* and *B(K)*. The effects of Text Length and Familiarity are significant, nevertheless there is no interaction between the 2 factors. In addition, high Wilks Lambda value [40] indicates that Familiarity only provides small amount of effects in changing the behaviours. However, this is sufficient for us to conclude that language familiarity does affect human behaviour, particularly in typing task, even the effect is small. Table V shows the results of the between-subjects effects. Text

Length significantly changes *B(U)* and *B(M)*, except *MCL*. But Text Length does not affect *B(K)* at all. This indicates that user's typing behaviour is not affected by the Text Length. This also explains the reason why anomalies could not be observed through *B(K)* despite increase of text length in Question 5 and Question 6. On the other hand, Familiarity affects *B(U)* and *B(K)*, but it does not affect *B(M)* (except *MIO*). To sum, *B(M)* is affected by Text Length but not Familiarity, while *B(K)* is affected by Familiarity but not Text Length. Figure

TABLE IV. UNIVARIATE TESTS FOR TEXT LENGTH AND FAMILIARITY TO *B(U)*, *B(M)* AND *B(K)*

| Factor | Significance, <i>p</i> -value | Wilk's Lambda, λ |
|-------------------------|-------------------------------|--------------------------|
| Text Length | 0.0000 | 0.5423 |
| Familiarity | 0.0000 | 0.8530 |
| Text Length*Familiarity | 0.3069 | 0.9430 |

Highlighted cell indicates that the effect is significant at $p < 0.00000005$ level

5 and Figure 6 demonstrate the responses of the *B(U)* and *B(M)* features according to Text Length and Familiarity. Note that *TD*, *MID*, and *MIO* demonstrate anomalous pattern when it reaches word length of 63 (which are Question 5 and Question 6). Nevertheless, by focusing only the short and medium word length (5 and 20 respectively), we could observe that when Text Length increased, *TD*, *SP* and *MID* increased, but *MS* and *MIO* decreased. Figure 7 illustrates the responses of the *B(K)* features according to Text Length and Familiarity. No significant changes can be observed when Text Length is increased. However in terms of Familiarity, *KS* and *EK* are

TABLE V. MANOVA TESTS OF THE BETWEEN-SUBJECTS EFFECTS

| Factor | Behaviour | Feature | p-value |
|-------------|-----------|---------|---------|
| Text Length | B(U) | TD | 0.0000 |
| | | SP | 0.0000 |
| | | MS | 0.0267 |
| | B(M) | MID | 0.0000 |
| | | MIO | 0.0000 |
| | | MCL | 0.7752 |
| | | KS | 0.5236 |
| | B(K) | KL | 0.5353 |
| | | EK | 0.4164 |
| | | | |
| Familiarity | B(U) | TD | 0.0221 |
| | | SP | 0.0162 |
| | | MS | 0.1461 |
| | B(M) | MID | 0.0729 |
| | | MIO | 0.0032 |
| | | MCL | 0.7083 |
| | | KS | 0.0000 |
| | B(K) | KL | 0.0003 |
| | | EK | 0.0074 |
| | | | |

Highlighted cell indicates that the effect is significant at the level of $p < 0.05$

TABLE VI. CORRELATION BETWEEN FEATURES

| | SP | TD | MS | MID | MIO | MCL | KS | KL | EK |
|-------------|----|----|----|-----|-----|-----|----|----|----|
| Question | ✓ | ✓ | | ✓ | ✓ | | | | ✓ |
| Length | ✓ | ✓ | | ✓ | ✓ | | | | |
| Familiarity | ✓ | | | | ✓ | | ✓ | ✓ | |
| SP | ■ | ✓ | | ✓ | ✓ | | ✓ | ✓ | |
| TD | ✓ | ■ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| MS | | | ■ | | ✓ | | | | |
| MID | ✓ | ✓ | | ■ | ✓ | ✓ | ✓ | ✓ | ✓ |
| MIO | ✓ | ✓ | ✓ | ✓ | ■ | | | ✓ | |
| MCL | | ✓ | | ✓ | | ■ | | | |
| KS | ✓ | ✓ | | | | | ■ | ✓ | |
| KL | ✓ | ✓ | | ✓ | ✓ | | ✓ | ■ | |
| EK | | ✓ | | ✓ | | | | | ■ |

Significant correlation exists between two features at $p < 0.05$ (2-tailed) level if it is ticked (✓). Highlighted cell indicates negative correlation coefficient.

significantly lower but *KL* is higher when the users typed in unfamiliar language. This explains that when users type the text in unfamiliar language, their average *SP* would be higher (see Figure 5). When *SP* increased, *KL* would become higher while *KS* and *EK* become lower. In terms of typing skill, the average completion duration per word dropped 62.35% from Question 2 to Question 3, as well as an increase of 25.03% of keystroke speed from Question 2 to Question 3, both signify increase of typing performance when Text Length increased.

To further analyse the relationships between all features, we ran Pearson Correlation Test. Due to the outliers given in Question 5 and Question 6, we only tested the correlation using Question 1 to Question 4 (see Table VI). From the results, *MCL* is not correlated to any features of *B(K)* nor *B(U)*, except *TD* and *MID*. When *TD* increased, *MCL* decreased. This shows that users would reduce the number of left mouse click when they spend longer duration on the typing task. *EK* is only correlated to *TD* and *MID*, but it is not correlated to *SP* and Question. Both Text Length and Familiarity are correlated to *SP*. Increase of Text Length will increase *SP*, but increase of Familiarity reduces *SP*. Besides, *SP* is correlated to *TD*, *MID*, *MIO*, *KS* and *KL*, but it is not correlated to *MS*, *MCL* and *EK*. When *SP* increased, *MID* and *KL* would increase but *MIO* and *KS* would drop relatively. Therefore this shows that cognitive stress can be detected by observing the *B(U)*, *B(M)* and *B(K)*. Besides, unifying both mouse and keyboard dynamics analyses is useful as they are actually correlated.

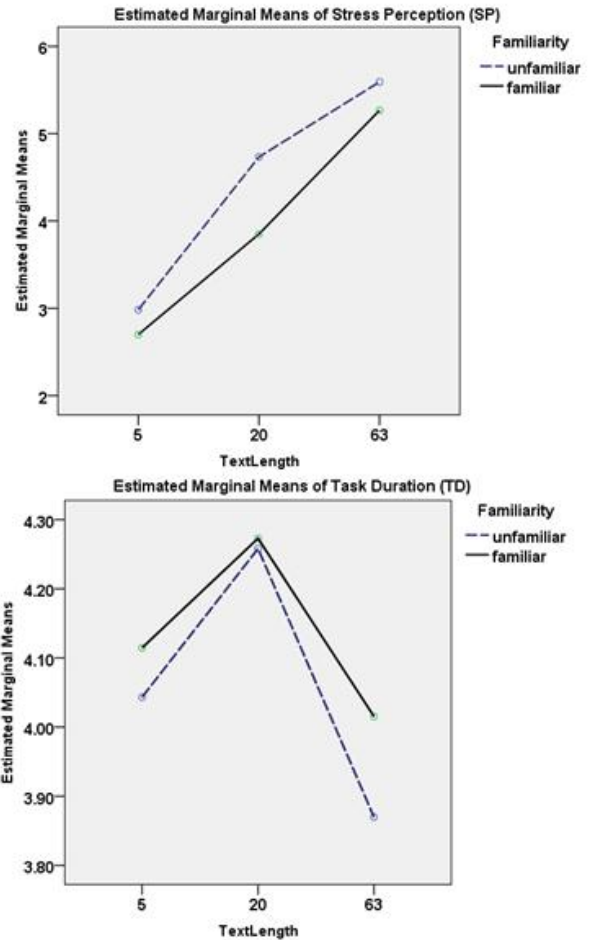


Fig. 5. Mean Plots of *B(U)* according to Text Length and Familiarity.

V. DISCUSSIONS

A summary of the observations based on the results are given below.

A. Main Factors

Question: Question produces significant effects on *SP*. The increment of text length in the Question always leads to higher *SP*. Users score highest *SP* when the Question is combined with long text and unfamiliar language.

Timing: Time pressure has no effect on *SP* and it has no interaction effect with Question, which indicates that even the users are given time pressure, they have no different stress perceptions to those without time pressure. However, Timing changes *TD* and *B(K)* significantly. This shows that time pressure pushes users to complete a task faster, although it may lead to more typing mistakes (increased *Err*).

Text Length: The increment of Text Length increases *SP*. Although Text Length has no significant effect on *B(K)*, but the average completion duration per word by the users dropped and *KS* increased when the number of words increased from 5 to 20. Both suggest that the performance of task completion increased when Text Length increased.

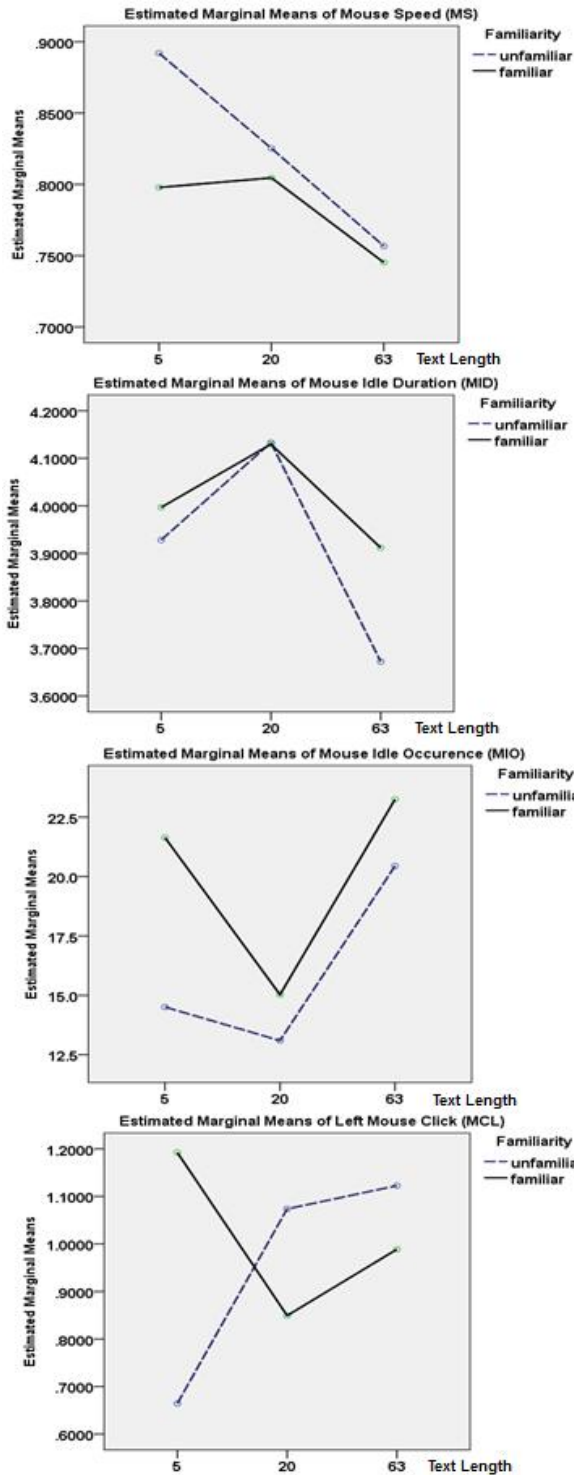


Fig. 6. Mean Plots of Mouse Behaviour Features according to Text Length (number of words) and Familiarity. Note that *MCL* is not significantly affected by Text Length and Familiarity.

Familiarity: Low familiarity of the language increases *SP* and reduces *MIO* and *KS*. When the users are unfamiliar with the language, they tend to type slower. Despite that, there is no significant difference in terms of *TD*. The possible reason is when the users are familiar with the language, the attempt to correct their typing errors before submission would increase

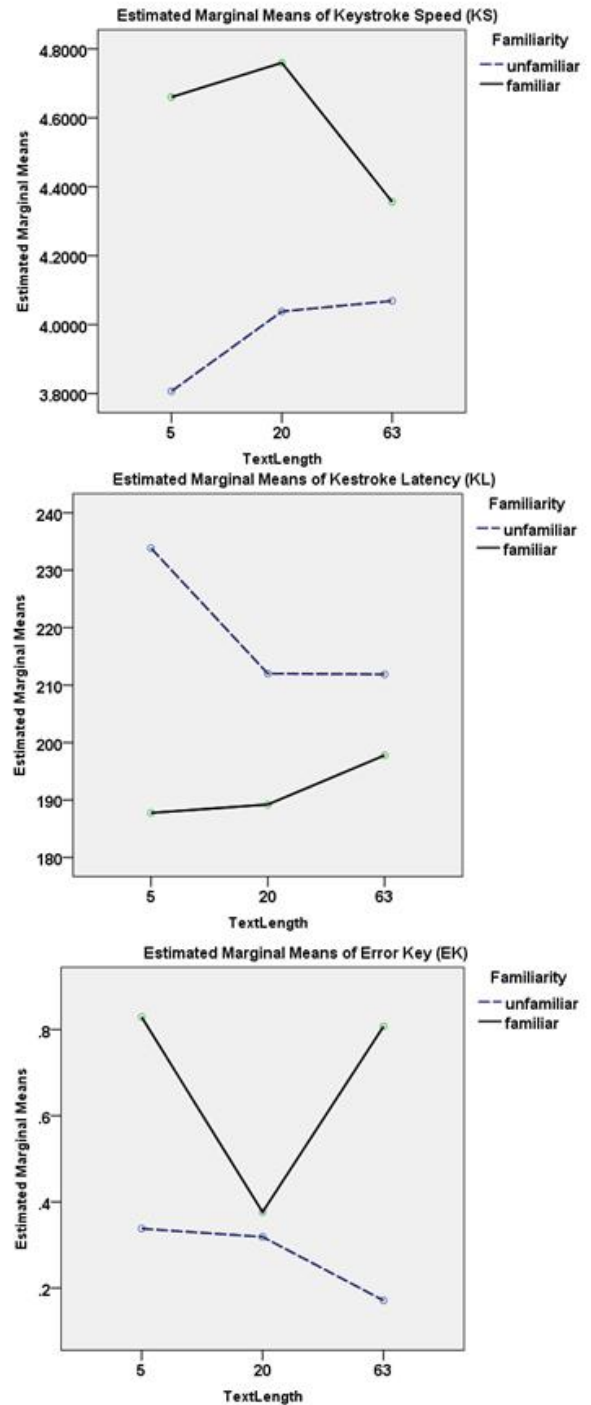


Fig. 7. Mean Plots of Keyboard Behaviour Features according to Text Length (number of words) and Familiarity.

(as *EK* increased), and this increases *TD*.

B. User Behaviour (*B(U)*)

SP: Users score higher *SP* for unfamiliar language. The increment of word length also leads to higher *SP*. If we relate the features of *B(M)* and *B(K)* to *SP*, higher *SP* leads to higher *MID* and *KL*, but lower *MIO*, *MS* and *KS*.

TD: The increment of *TD* leads to higher *SP*. It is also

correlated to almost all features except *MS*. This suggests that if the users need to spend longer time to complete a task, their perceptions of stress will increase, and their behaviours of using mouse and keyboard would change relatively.

PA: *PA* has no change until the participants reached Question 5. It is also very interesting to note that the students started to “cheat” from this point onwards, and this was further supported by statistical evidence. Therefore we consider that Question 5 is where the users have started losing motivation to continue the task. This can be explained by a few reasons, which include (1) high demand (of text length) that exceeds their estimated effort to complete the task, (2) time constraint and projected high *TD*, which reduce their estimated probability of success, (3) the task is beyond their acceptable effort to invest, and (4) the aversiveness of the task cause fatigue and tiredness to the students at the end of the experiments. However, we still believe that when the users are given long text, even they are not pressured by time constraint, the job will still be considered too demanding (*SP* increased).

Err: When the demand of job increased according to Question, the number of users with *Err* = 0 dropped. The number of users who did not make any mistakes is obviously lower when they were given time pressure. This suggests to us that when job demand increased, more users will make more errors, this is worse if they are given time pressure. However, even some users obtained perfect score for Question 5 and Question 6, most of them did not actually type the text (but copy-and-paste).

C. Mouse Behaviour (*B(M)*)

MS: *MS* decreased when users perceive higher *SP* although the correlation is not significant. *MS* is significantly affected by Timing and Text Length but not Familiarity. Increase of Text Length results in slower *MS*.

MID: *MID* is correlated to almost all features, except *MCL*. *MID* is affected by Text Length but not Timing and Familiarity. Increased Text Length results in higher *MID*.

MIO: *MIO* is mainly correlated to *B(U)* and other *B(M)* features (except *MCL*) but not *B(K)* (except *KL*). *MIO* is affected by Timing, Text Length and Familiarity. *MIO* decreased when Text Length increased. It would also decrease if Familiarity decreased.

MC: No one used *MCR* at all in the entire experiment. For *MCL*, it is not correlated to any features of *B(K)* nor *B(U)*, except *TD* and *MID*. When *TD* increased, *MCL* decreased. This shows that users reduce the number of left mouse click when they spend long duration on typing task. *MCL* decreased when *MID* increased.

D. Keyboard Behaviour (*B(K)*)

KS: *KS* is significantly affected by Timing. Interestingly the participants who were given time pressure typed slower than those without time pressure. One possible reason is the results may be affected by other uncontrolled factors. Examples include (1) the users in the control group may experience other kind of time pressure incurred by external environment; (2) the students in the experimental group probably were not taking the experiments seriously; or (3) the students in the control

group possessed better typing skills than those in the experimental group. *KS* is also affected by Familiarity. Users tend to type slower when they are unfamiliar with the language. This suggests that the users are not able to anticipate the upcoming words in unfamiliar language, such as German, and this affects their associated key presses. However, *KS* is not affected by Text Length. *KS* does not change significantly even Text Length has been increased. *KS* is lower when *SP* increased. *KS* is also correlated to *MID*, shows that when the users are busy pressing the keys, the mouse idle duration will be longer.

KL: Similar to *KS*, *KL* is affected by Timing and Familiarity but not Text Length. *KL* is also correlated to *SP*, *TD*, *MID*, *MIO* and *KS*. *KL* increased when *SP* increased.

EK: No one used *DK* at all in the entire experiment, therefore *EK* solely contains *BSK*. *EK* is only correlated to *TD* and *MID*, but it is not correlated to *SP* and Question. *EK* is high when Familiarity is high, indicates that users tend to make more corrections when they are familiar with the language. One of the reasons is the users can identify the errors easily when they are more familiar with the language. Another possible reason is that the capability of the browser (such as Google Chrome) that enables English spelling checking helps the users to spot the spelling error(s). Therefore it is important to switch off the capabilities of spelling and grammar checking before the experiment is conducted. Besides, *EK* is not affected by Text Length (as shown in Figure 7, the mean of a user using an error key is always less than one regardless the length of the text).

E. Using *B(U)*, *B(M)* and *B(K)* in Emotional Stress Detection

Text Length mainly affects *B(U)* and *B(M)* but not *B(K)*. Familiarity affects *B(U)* and *B(K)* but not *B(M)*. This information provides us some guides in order to design an automated human behaviour model. First we should mainly look into *B(U)* and *B(M)* features if the typing tasks involve changes in length. If *TD* increased, and at the same time *MID* increased but *MS* and *MIO* decreased, while there are no changes in *B(K)* features, then we could infer that the typing task demand has been increased. Secondly, to assess how familiar of the user with the task, we should look into *B(U)* and *B(K)* features. If the user is familiar with the typing task (such as language), this should show increment of *MIO*, *KS* and *EK*, drop in *KL* and no significant differences for other *B(M)* features. In our research, *MCL* may not be useful as an indicator in stress perception prediction, but the correlation results could still have provided a link, i.e. increase of *TD* results in decrease of *MCL*. If *TD* increases *SP*, then decrease of *MCL* may also mean increase of *SP* if *TD* also increased.

Generally, our research shows some important information in stress detection. Firstly, longer text length increases *SP*. Users also perceive higher stress if they are unfamiliar with the language. Time pressure does not necessarily affect how users perceive stress. Secondly, the correlations show that users perceive higher stress if they need to take longer time to complete the task. Therefore we can infer that *SP* should increase if *TD* increased. Thirdly, when the task complexity changed (such as Text Length and Familiarity), we could observe some changes in *B(M)* and *B(K)*. Higher *SP* generally

results in lower *MS*, *MIO* and *KS*, but higher *MID* and *KL*. Therefore, when the users are stressed, they would move the mouse slower and type slower. Lastly, the anomalies of the behaviours could suggest us when the users have reached an ultimate stress point or have lost motivation. For instance, *TD* and *MID* started to drop while *MIO* increased at Question 5, although the users perceived even higher stress for Question 5 and Question 6. Therefore the projection of the emotional stress level could only be valid as long as the students are still motivated to continue the task. Once anomalous behaviours are detected according to the projection of stress perception, then an adaptive system could be activated to motivate the students to continue their tasks.

VI. CONCLUSION

Our research findings show that users' emotional stress can affect keyboard dynamics and mouse dynamics. The correlation between mouse behaviour and keyboard behaviour indicates that the unification of both methods are useful than utilizing a single method alone. Longer text and unfamiliar language raise users' stress perceptions. Demanding job such as long typing task could result in anomalous behaviours once the user has lost motivation. Language familiarity does affect human behaviour and keyboard behaviour, even though the effect is small. Text length change mouse behaviour but not keyboard behaviour. Higher stress perception generally results in lower mouse speed, mouse idle occurrences and keystroke speed, but higher mouse idle duration and keystroke latency. This stress detection model will be used as the basis to construct a computational algorithm for detecting user's emotional stress in our future research.

However, our research is not without limitations. First, the sample size is small, which only consists of 30 students in each group, therefore we may not be able to generalize the findings. More rigorous experiments need to be conducted to verify the stress model. Secondly, homogeneity cannot be assumed between questions due to outliers and anomalies in Question 5 and Question 6. Thirdly, different users have different typing skills, and their typing abilities could have affected the results. Lastly, the results may also be affected by external environmental factors, such as external time pressure, participants' mood, and the motivation of the participants to continue the experiments. Lastly, our future work aims to design an automated stress detection system that can be implemented in an adaptive e-learning system.

ACKNOWLEDGEMENT

The authors would like to thank Tunku Abdul Rahman University College for allowing the experiments to be setup and conducted in one of its laboratory.

REFERENCES

- [1] J. Sweller, P. Ayres, and S. Kalyuga, *Cognitive load theory*. Springer, 2011, vol. 1.
- [2] J. L. Plass, R. Moreno, and R. Brünken, *Cognitive load theory*. Cambridge University Press, 2010.
- [3] T.-J. Chang, P.-S. Hsu, M.-H. Wu, and C.-P. Chuang, "A dynamic user-friendly interactive interface for adaptive e-learning: A real-time and non-interference diagnostic technique," in *Knowledge Acquisition and Modeling, 2008. KAM'08. International Symposium on*. IEEE, 2008, pp. 462–466.

- [4] R. A. Karasek, "Job demands, job decision latitude, and mental strain: Implications for job redesign." *Administrative science quarterly*, vol. 24, no. 2, 1979.
- [5] A. E. Rijk, P. M. L. Blanc, W. B. Schaufeli, and J. Jonge, "Active coping and need for control as moderators of the job demand–control model: Effects on burnout," *Journal of Occupational and Organizational Psychology*, vol. 71, no. 1, pp. 1–18, 1998.
- [6] L. M. Vizer, "Detecting cognitive and physical stress through typing behavior," *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems CHI EA 09*, p. 3113, 2009.
- [7] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Troster, and U. Ehlert, "Discriminating Stress From Cognitive Load Using a Wearable EDA Device," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 410–417, Mar. 2010.
- [8] R. Sloan, J. Korten, and M. Myers, "Components of heart rate reactivity during mental arithmetic with and without speaking," *Physiology & Behavior*, vol. 50, no. 5, pp. 1039–1045, Nov. 1991.
- [9] M. M. Khan, R. D. Ward, and M. Ingleby, "Classifying pretended and evoked facial expressions of positive and negative affective states using infrared measurement of skin temperature," *ACM Transactions on Applied Perception*, vol. 6, no. 1, pp. 1–22, 2009.
- [10] S. Tobias and T. Abramson, "Interaction among anxiety, stress, response mode, and familiarity of subject matter on achievement from programmed instruction." *Journal of educational psychology*, vol. 62, no. 4, p. 357, 1971.
- [11] C. Hulme, S. Maughan, and G. D. Brown, "Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span," *Journal of Memory and Language*, vol. 30, no. 6, pp. 685–701, 1991.
- [12] D. Carroll, *Health psychology: stress, behaviour and disease*. London: The Falmer Press, 1992.
- [13] N. M. Yusoff and S. S. Salim, "Scout and affective interaction design: Evaluating physiological signals for usability in emotional processing," in *2nd International Conference on Computer Engineering and Technology (IC CET)*, 2010, pp. V1–201–V1–205.
- [14] J. Wahlström, M. Hagberg, P. Johnson, J. Svensson, and D. Rempel, "Influence of time pressure and verbal provocation on physiological and psychological reactions during work with a computer mouse," *European journal of applied physiology*, vol. 87, no. 3, pp. 257–263, 2002.
- [15] M. Heiden, E. Lyskov, M. Djupsjöbacka, F. Hellström, and A. G. Crenshaw, "Effects of time pressure and precision demands during computer mouse work on muscle oxygenation and position sense," *European journal of applied physiology*, vol. 94, no. 1-2, pp. 97–106, 2005.
- [16] J. Zhai, A. Barreto, C. Chin, and C. Li, "User stress detection in human-computer interactions." *Biomedical sciences instrumentation*, vol. 41, pp. 277–282, 2004.
- [17] R. L. Hazlett, "Measuring emotional valence during interactive experiences: boys at video game play," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 2006, pp. 1023–1026.
- [18] P. Ekman and W. V. Friesen, "Facial action coding system: A technique for the measurement of facial movement. palo alto," 1978.
- [19] P. Branco, P. Firth, L. M. Encarnação, and P. Bonato, "Faces of emotion in human-computer interaction," in *CHI'05 Extended Abstracts on Human factors in computing systems*. ACM, 2005, pp. 1236–1239.
- [20] W. L. Cheong, C. M. Char, Y. C. Lim, S. Lim, and S. W. Khor, "Building a computation savings real-time face detection and recognition system," in *Signal Processing Systems (ICSPS), 2010 2nd International Conference on*, vol. 1. IEEE, 2010, pp. V1–815.
- [21] J. R. Cowell and A. Ayesh, "Extracting subtle facial expression for emotional analysis," in *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, vol. 1. IEEE, 2004, pp. 677–681.
- [22] A. Ayesh and J. Cowell, "Emotional analysis of facial expressions." *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, vol. 1, pp. 672–676, 2004.
- [23] J. Cowell and A. Ayesh, "Automatic analysis of facial expressions: problems and solutions to data collection," 2003.

- [24] L. M. Vizer, L. Zhou, and A. Sears, "Automated stress detection using keystroke and linguistic features: An exploratory study," *International Journal of Human-Computer Studies*, vol. 67, no. 10, pp. 870–886, 2009.
- [25] K. Revett, F. Gorunescu, M. Gorunescu, M. Ene, S. Magalhaes, and H. Santos, "A machine learning approach to keystroke dynamics based user authentication," *International Journal of Electronic Security and Digital Forensics*, vol. 1, no. 1, pp. 55–70, 2007.
- [26] M. Pusara and C. E. Brodley, "User re-authentication via mouse movements," in *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*, ser. VizSEC/DMSEC 2004. New York, USA: ACM, 2004, pp. 1–8.
- [27] G. Tsoulouhas, D. Georgiou, and A. Karakos, "Detection of Learners Affective State Based on Mouse Movements," *Journal of Computing*, vol. 3, no. 11, pp. 9–18, 2011.
- [28] H. Selye, *Stress in health and disease*. Butterworth, 1946.
- [29] —, *The stress in life*. McGraw-Hill, 1956.
- [30] D. Gunetti and C. Picardi, "Keystroke analysis of free text," *ACM Transactions on Information and System Security (TISSEC)*, vol. 8, no. 3, pp. 312–347, 2005.
- [31] J. A. Robinson, V. Liang, J. M. Chambers, and C. L. MacKenzie, "Computer user verification using login string keystroke dynamics," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 28, no. 2, pp. 236–241, 1998.
- [32] G. C. Boechat, J. C. Ferreira *et al.*, "Authentication personal," in *Intelligent and Advanced Systems, 2007. ICIAS 2007. International Conference on*. IEEE, 2007, pp. 254–256.
- [33] H.-R. Lv, Z.-L. Lin, W.-J. Yin, and J. Dong, "Emotion recognition based on pressure sensor keyboards," in *Multimedia and Expo, 2008 IEEE International Conference on*. IEEE, 2008, pp. 1089–1092.
- [34] P. S. Teh, S. Yue, and A. B. J. Teoh, "Improving keystroke dynamics authentication system via multiple feature fusion scheme," in *Cyber Security, Cyber Warfare and Digital Forensic (CyberSec), 2012 International Conference on*. IEEE, 2012, pp. 277–282.
- [35] N. Eswari, S. Sundarapandiyam, P. Vennila, R. Umamaheswari, and G. Jothilakshmi, "Keystroke biometrics with number-pad input using hybridization of adaboost with random forest," in *Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on*. IEEE, 2012, pp. 105–109.
- [36] R. Giot, M. El-Abed, and C. Rosenberger, "Web-based benchmark for keystroke dynamics biometric systems: a statistical analysis," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on*. IEEE, 2012, pp. 11–15.
- [37] R. Giot, C. Rosenberger, and B. Dorizz, "Can chronological information be used as a soft biometric in keystroke dynamics?" in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on*. IEEE, 2012, pp. 7–10.
- [38] T. Shimshon, R. Moskovitch, L. Rokach, and Y. Elovici, "Clustering di-graphs for continuously verifying users according to their typing patterns," in *Electrical and Electronics Engineers in Israel (IEEEI), 2010 IEEE 26th Convention of*. IEEE, 2010, pp. 000 445–000 449.
- [39] J. E. Davidson and R. J. Sternberg, *The psychology of problem solving*. Cambridge university press, 2003.
- [40] IBM, "Multivariate General Linear Modeling," 2012, available from <http://pic.dhe.ibm.com/infocenter/spssstat/v21r0m0/index.jsp?topic=/com.ibm.spss.statistics.cs/glm\intro.htm>.
- [41] IBM, "GLM Univariate Analysis," 2011, available from <http://pic.dhe.ibm.com/infocenter/spssstat/v20r0m0/index.jsp?topic=/com.ibm.spss.statistics.help/idh\glmu.htm>.