Ergonomics

Reactions, accuracy and response complexity of numerical typing on touch screens

Cheng-Jhe Lin [a] [b] & Changxu Wu [a]

[a] Department of Industrial and Systems Engineering, State University of New York at Buffalo, 414 Bell Hall, Buffalo, NY, USA

[b] Department of Industrial Management, National Taiwan University of Science and Technology, Taipei, Taiwan Republic of China (R.O.C.)
Version of record first published: 19 Apr 2013.

PLEASE SCROLL DOWN FOR ARTICLE

Reactions, accuracy and response complexity of numerical typing on touch screens

Cheng-Jhe Lin[a,b1] and Changxu Wu[a]*

aDepartment of Industrial and Systems Engineering, State University of New York at Buffalo, 414 Bell Hall, Buffalo, NY, USA;
bDepartment of Industrial Management, National Taiwan University of Science and Technology, Taipei, Taiwan Republic of China (R.O.C.)

Touch screens are popular nowadays as seen on public kiosks, industrial control panels and personal mobile devices. Numerical typing is one frequent task performed on touch screens, but this task on touch screen is subject to human errors and slow responses. This study aims to find innate differences of touch screens from standard physical keypads in the context of numerical typing by eliminating confounding issues. Effects of precise visual feedback and urgency of numerical typing were also investigated. The results showed that touch screens were as accurate as physical keyboards, but reactions were indeed executed slowly on touch screens as signified by both pre-motor reaction time and reaction time. Provision of precise visual feedback caused more errors, and the interaction between devices and urgency was not found on reaction time. To improve usability of touch screens, designers should focus more on reducing response complexity and be cautious about the use of visual feedback.

Practitioner Summary: The study revealed that slower responses on touch screens involved more complex human cognition to formulate motor responses. Attention should be given to designing precise visual feedback appropriately so that distractions or visual resource competitions can be avoided to improve human performance on touch screens.

Keywords: touch screen; numerical typing; human errors; response complexity; visual feedback

1. Introduction

Touch screens are popular nowadays as alternative input devices to computer mice or physical keyboards. They have obvious advantages of being highly integrable with displays, customisable for the user's preference (Sears 1991; Huang and Lai 2008) and user-friendly through direct manipulation of controlled elements (Shin and Zhu 2011). Therefore, they are widely adopted in various environments including civil applications, such as point-of-sale (POS), medical devices, industrial control panels, public kiosks, automated teller machines (ATMs), personal electronic devices (Huang and Lai 2008), and military usage (Goode, Lenné, and Salmon 2012). When touch screens are used for selecting targets, they can be as fast and accurate as computer mice (Sears and Shneiderman 1991) or even superior (Chung et al. 2009). However, previous studies have shown that using touch screen keyboards to type is slower and subject to human errors, especially in typing alphabetical content (Sears 1991; Thomas and Mcclelland 1996). Studies by Sears and his colleagues (Sears 1991, Sears et al. 1993) investigated alphabetical typing on touch screens and sought to improve typing performance by optimising touch screen keyboard design. The design of touch screen keyboards included optimal screen declination angle, enlarged key size (2.27 cm), adaptive reception area to touch biases and provision of both visual feedback (inverse images for keys) and auditory feedback (a short tone) for selection and activation of touch keys. Despite their optimal design, the typing speed of frequent computer users on a touch screen was still inferior to that on a physical QWERTY keyboard (25 vs. 58 words per minute [WPM] for touch screen and physical keyboard, respectively). Their further study showed that an experienced typist could reach 32.5 WPM, still inferior to the performance on physical keyboards by 44% (Sears et al. 1993). In discussion, several possible reasons accounting for inferior performance were raised, including inability to use multi-finger typing, slow response time of touch screens and non-optimal placement of the touch keyboard. Based on these findings, the authors suggested future studies to be done in comparing mechanical keyboards with touch screen keyboards with equivalent size and repeating similar studies with new technologies that may reduce errors and improve typing rates.

1.1. Comparing touch screens with physical keypads for numerical typing

Among all tasks that can be possibly done on touch screens, numerical typing is one of the most frequent tasks that users perform and yet receive relatively little attention in research. Dialling numbers on mobile phones, typing amounts to withdraw on ATM machines, entering addresses on GPS system, to name a few, are commonly seen numerical typing tasks

---

*Corresponding author. Email: seanwu@buffalo.edu

that can be performed on touch screens. Numerical typing tasks, however, are subject to human errors. Numerical typing errors can result in serious loss in finance (Bohm et al. 2008) or fatal accidents in critical systems, such as medical facilities, nuclear power plants and military combat systems (Arndt et al. 1994; O'Hara, Higgins, and Brown 2008; Thimbleby and Cairns 2010). Chung et al. (2009) have compared touch screens with physical keyboards and found that older and young adults performed differently in numerical typing: older adults in their 60s and 70s preferred the touch screen keypads and could operate more quickly with more errors committed, while young adults in their 20s and 30s demonstrated no difference between using touch screens and physical keyboards in terms of typing speed or accuracy. Their results, however, are subject to several confounding factors. First, the touch screen was installed in front of the user with a slightly backward declination while the physical keypad was placed on the right-hand side of the user on a flat surface. The difference in performance therefore might come from biomechanical factors rather than devices, i.e. older adults might have committed fewer errors if the two keyboards had been placed at the same location. Second, keys on the touch screens were bigger than those on physical keypads in their study, so faster reactions and preference of older adults for touch screens may just reflect low difficulty in aiming (Fitts 1954). Finally, all young adults in the study were college students who tended to have more experience in using both physical keyboards and touch screens than older adults. The discrepancy between performance on physical keyboards and on touch screens for older adults may be related to their lack of experience in using touch screens. Although Chung et al. (2009) stated that older adults in the study had experience in using the Internet and cellular phones, their expertise in using physical keyboards and touch screens were not reported.

Thomas and Mcclelland (1996) conducted another comparison study about the use of touch screens in six control rooms of police forces (Thomas and Mcclelland 1996). In five out of the six rooms, the touch screen keypads installed in communication terminals drew complaints from police officers as slow and error prone while they were typing numerical data. One touch screen keypad was replaced with physical keypad following the complaint, and improvement of performance was reported. A small experiment was also conducted by Thomas and Mcclelland (1996) for number dialling tasks, and superior performance in speed and accuracy was found for physical keyboards. Although both Chung et al. (2009) and Thomas and Mcclelland (1996) compared numerical typing performance of young, experienced adults, the results are quite different. Urgency may play an important role in causing discrepancy in findings of the two studies. The urgency can be regarded as an intrinsic motivation to finish the work quickly with a certain level of accuracy so that a tangible reward such as money, time or promotion can be obtained. A recent study showed that numerical typing performance on a standardised numerical keyboard could be influenced by pacing, response complexity and urgency of the task (Lin and Wu 2011). It is not clear, however, if the results can be applied to other non-standard alternative typing devices. While working in the control rooms, police officers were likely under time pressure, and their internalised urgency may worsen the usability issues of touch screens, i.e. they were no longer able to cope with more complex motor reactions on touch screens and maintain behavioural performance under time pressure compared to young adults in a relatively casual situation (self-paced typing). The confounding factors in the studies require a controlled experimental study to clarify innate differences between the two devices, and the discrepancy of their findings elicits curiosity about the appropriateness of using touch screens under urgency.

1.2. *Pre-motor reaction time and innate difference*

Pre-motor reaction time which is derived from electromyographic (EMG) data has been used to measure the delays related to central programming activity (Ketelaars, Garry, and Franks 1997; Franks et al. 1998; Coombes, Cauraugh, and Janelle 2007) in addition to traditional reaction time because it can quantify response complexity associated with central programming difficulty in using a device and is relatively independent of non-programming factors and mechanical factors that are not necessarily innate to the device itself. In theory, reaction time, traditionally defined as the time lapse between the presentation of imperative stimuli and the ending of respondent movements, can be used to compare the complexity of perceptual-motor tasks when responses are made on the same device. For comparing touch screens with physical keyboards, however, variations in mechanical response time on two different devices can systematically bias the results (Li et al. 2010). That is, slower reaction time on one device may be caused by its slower mechanical response time and is not necessarily an innate disadvantage of performing the task on the device since its response time may improve with technology. Furthermore, users may strategically choose rapid execution on one device to compensate its slower mechanical response time and thus make more errors. The speed–accuracy trade-off associated with strategic execution is not the innate difference between the devices either, but, rather, users' choice, i.e. different users do not necessarily implement the same strategy on the chosen device for responses.

The innate difference between devices should be associated with the change during central motor programming merely due to the change of response factors (e.g. different typing devices) when the schema of the task and the device is realised by the user. The temporal length of central motor programming and the response complexity are related because the length is

determined by the parametric features of a particular response that are mapped onto the schema during the preparatory phase of the response (Glencross 1980). The predetermined response parameters contain information of invariant response units relevant to motor control system, such as numbers of synergies, phases and sequences for the specified movement. The demand levels to 'code' those parameters also vary according to accuracy requirements, amount of uncertainty and timing constraints (Glencross 1980; Spijkers and Walter 1985). In consequence, complexity of the central motor programming associated with the device would be reflected by the length of programming before the onset of the neural drive to peripheral motor units measured in EMG. Therefore, pre-motor reaction time which is defined as the lapse between the presentation of imperative stimuli and the beginning of myoelectric burst of muscle activity can quantify response complexity associated with innate difficulty of using a device, i.e. the response complexity that is related to the preparation of the response itself.

On touch screens, preparation of motor reactions (motor planning) would be more complex due to lack of tactile information. Neurological studies have found that tactile information is important to motor programming. The sensitive fingertips were able to detect small edges and curves that indicate initial fingertip location on physical keyboards, and the information about the start position is essential to executing rapid aiming accurately (Rabin and Gordon 2004). The contact force at fingertip on physical keyboards also provided information regarding the orientation of the finger and forearm in space (Rao and Gordon 2001) and may be used to replicate hand posture in space. When touch screens are used, no tactile information about start position or arm orientation can be obtained, and this lack of information expectantly increase the complexity of motor programming because start position and finger orientation need to be estimated from visual information and the estimation adds uncertainty to the motor programming, which in turn raises the demand level to code motor parameters.

1.3. Precise visual feedback

Since pre-motor reaction time is capable of measuring response complexity, whether the provision of visual feedback can help reduce the complexity in using touch screens is also of research interests. Compared to standard physical keyboards, lack of tactile feedback when a specific key was touched during touch screen operations was frequently mentioned in usability literature as a cause of frustration and inaccuracy (Martin 1988; Maguire 1999; Benko and Wigdor 2010). First, the smooth surface of the touch screen keyboard causes the fingertip to lose valuable tactile information for movement control. Fingertip anaesthesia studies showed that loss of sensory afferents from the fingertip could degrade typing accuracy dramatically (Gordon and Soechting 1995; Rabin and Gordon 2004) because tactile feedback provided both precise location of the finger based on edge pressure and orientation of the finger based on reaction force from the surface (Rao and Gordon 2001; Rabin and Gordon 2004). However, disruption of tactile feedback did not change fingertip trajectories of movements or typing speed but variability of movements in consecutive typing. It was thus concluded that tactile feedback was essential for planning of the movement, based on either start positions or error recognition at end positions, but not for compensatory strategies during the motion (Gordon and Soechting 1995; Rabin and Gordon 2004).

Second, imprecise visual feedback (IVF) that provides only discrete states of keys on touch screens creates inconsistency and does not help motor programming in finger movements. The activation of the feedback can happen either before or after the touch, depending on touch technologies used (Benko and Wigdor 2010). For example, infrared technology detects the finger before it touches the screen surface, while capacitive and resistive technologies sense the finger after it touches the surface. Therefore, for infrared touch screens there is no physical sensation connected with the change of feedback status, and it is frustrating for the user to see that she/he 'engages' to the interactive key before she/he really touches it. As an alternative, innovative visual feedback that provides information of touch positions (Benko and Wigdor 2010) instead of activation status, vibrotactile feedback that substitutes touch senses (Poupyrev, Maruyama, and Rekimoto 2002) and even a redundant use of both (Lee, Poliakoff, and Spence 2009) were suggested. While the provision of precise positional information through visual channels can be low cost and intuitive, it raises concerns about resource competition because too much information needs to be processed through visual modality, and the competition could be worse under urgency. Moreover, whether precise visual feedback (PVF) is helpful for touch screen use under urgent situations has not been studied in the literature.

1.4. Goal and research questions

The goal of this study is to answer the following research questions through an experimentation in which confounding factors in biomechanical and physical design aspects are eliminated. First, do users perform differently on touch screens and physical keyboards in the context of numerical typing? Second, does the use of touch screen cause higher response complexity in terms of motor programming and imply innate difficulty of use? Third, is urgency level an influential factor to consider when it comes to choosing touch screens or physical keyboards? Fourth, is the provision of precise positional

information through visual feedback helpful, regardless of urgency level, or not? The hypotheses to these research questions are: first, users do not perform differently on touch screens and physical keyboards if biomechanical and physical design factors are eliminated; second, touch screens are indeed more complex in terms of motor programming, but, somehow, users can manage to maintain the same level of performance; third, urgency is an influential factor to consider as the flexibility of users to maintain performance will deteriorate under urgency; finally, providing precise positional information through visual feedback is helpful under non-urgent condition but not under urgency.

2. Method

2.1. Experimental variables

In this experiment, two independent variables were the types of keyboards used and urgency realised in the numerical typing task. Three different types of keyboards were used: a standard numerical keyboard (Micro Innovations® KP-17B), a software keyboard with IVF and an optional version of the software keyboard with PVF. The software keyboard was implemented by a Microsoft Visual Basic® (Microsoft Corporation, Redmond, WA) program and displayed on a 12-inch capacitive touch screen (Elo Touchsystems® ET1215L). The PVF was a squared red dot (10 × 10 pixels) that appeared after each touch operation to show the participant the exact position of the touch (Figure 1). The function of this PVF was to provide the exact location of the centre of the touch detected by the touch screen in addition to the activation status of the simulated keys. This option was turned on only in trials of the Touch screen with PVF condition. When the feedback option was turned off, the image of the key was only inverted to simulate the visual effect of the depression of the key (IVF). Although users would know whether the key was activated thorough this visual effect, the inverted image of the key meant only that the key was touch within its effective sensing area. The user was unable to know the exact location of the touch through this visual effect due to lack of tactile feedback, i.e. the edge pressure on the fingertip to indicate the relative position of the finger to the key.

The size and spacing of the software keyboard was calculated based on the physical size and spacing of the standard keyboard so that its visual presentation was the same as the physical keyboard. That is,

$$SK_{pixel} = \frac{DS_{pixel}}{DS_{mm}} PK_{mm}, \tag{1}$$

where SK_{pixel} is the size of the software keyboard in pixels and PK_{mm} is the size of the physical keyboard in millimetres.
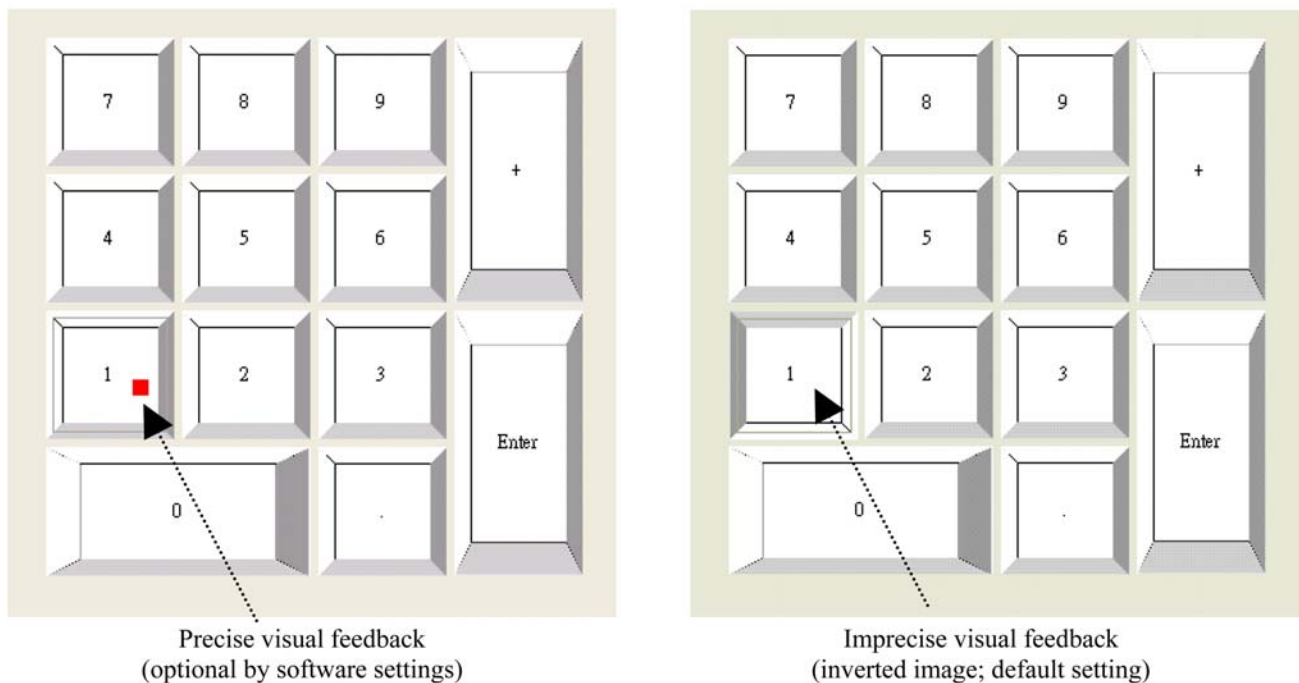

Figure 1. Software keyboard (left: PVF version; right: IVF version).

DS_{pixel} and DS_{mm} are the resolution (800 × 600 pixels) and the viewable area (246 × 184 mm) of the touch screen used. Size and spacing of the software keyboard was calculated based on the physical size (12.7 mm) and spacing (6.4 mm) of the standard keyboard, so its visual presentation was the same as the mechanical keyboard. Based on Equation (1), the software key size was set at 40 pixels and the gap is set at 20 pixels.

For the urgency, monetary rewards contingent to the quality of the work could increase the intrinsic motivation (Eisenberger and Cameron 1996) and thus were used to increase participants' internalised urgency, an intrinsic motivation to finish the work quickly with a certain level of accuracy. A recent study in numerical typing also successfully used different monetary rewards to manipulate levels of urgency realised by users (Lin and Wu 2011). Therefore, two different rewarding systems were applied in the experiment to elicit urgent feelings. In three out of six experimental trials (three keyboards and two levels of urgency resulted in six combinations), the participants were told in advance that they could obtain performance bonus (up to \$20 per trial) if their correct responses could be made within 600 ms. In the other three trials participants were also encouraged to perform well but received only flat-rate time compensation (\$ US dollars 10 per hour) independent of their performance.

Other than the two independent variables, task variables associated with the task, such as the speech rate of verbally presented numbers (1 digit per second), the physical dimensions of the keyboards (i.e. touch screen keyboards emulate physical keyboards), the placement of keyboards and other environmental settings that may be influential to typing performance, were controlled. The speech rate of one digit per second was selected because one second was a safe interval to avoid overlaps between two consecutive typing actions, i.e. the subsequent stimulus was presented before the on-going typing action was finished. A recent study showed that unnecessary overlaps between two consecutive typing actions could delay reaction time (Lin and Wu 2011). In addition, the control of the keyboard placement was implemented by a modified $3M^{TM}$ KP200E adjustable keyboard tray (Figure 2). When the touch screen and the physical keyboard were installed, the modified keyboard tray secured the centre of both keyboards at the same spatial location. In addition, alignment markers on the desk and the keyboard allowed experimenter to visually check whether keyboards were properly installed.

Pre-motor reaction time, reaction time and frequency of errors were three major dependent variables measured. During each trial, EMG activity from finger flexors was continuously collected in each trial (please see details in Section 2.3). EMG data were further processed to determine pre-motor reaction time (please see details in Section 2.4). Moreover, the behavioural data (including what key was pressed and when the key was pressed) were recorded by the software program. The behavioural data log was compared to the data log of numerical stimuli to compute average reaction time and frequency of errors in each trial. Finally, subjective ratings about the numerical typing task after each trial were also collected by National Aeronautics and Space Administration Task Load Index (NASA-TLX) questionnaires as supplementary dependent variables to determine if the main effects revealed by objective measures were also supported by subjective ratings (NASA 2011). Subjective ratings in six dimensions of NASA-TLX questionnaires were collected immediately after
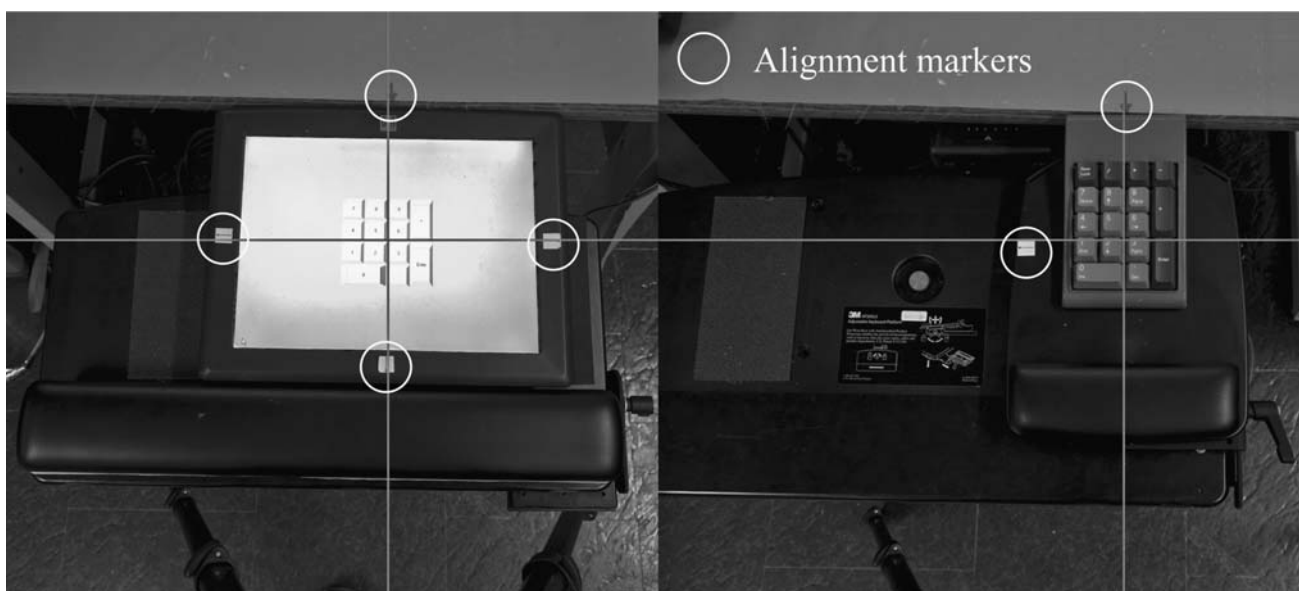

Figure 2. Adjustable keyboard tray (left: touch screen; right: physical keyboard).

participants finished each trial, and the results were also analysed statistically and compared with the results of objective measurements.

2.2. Participants

A total of 22 volunteers participated in this study including 12 males and 10 females (average age: 22 and 23 years, respectively; see Table 1). They were all young, native speakers of English without any hearing disability. The participant group was considered representative of a youthful working force (Anon 2010) with intact perceptual-motor abilities. A numerical typing pretest was required before participation to ensure their familiarity with numerical typing. They were instructed to type out 30, random, 9-digit numbers by their right index fingers on both the physical keyboard and the touch screen. To qualify, they had to finish the task in 200 seconds with at least 80% accuracy. This requirement is above the mean performance of participants in a previous numerical typing study (Marteniuk, Ivens, and Brown 1996) and comparable to representative number keying rates of skilled typists (Seibel 1977). All qualified participants signed informed consent before the experiment, and they were compensated for their time of participation (flat-rate compensation). Average typing speed in pretest for males and females was 96.49 keys per minute and 96.69 keys per minute, respectively. Average typing accuracy in pretest for males and females was 91.80% and 90.83%, respectively. A simple two-sample t-test showed no gender difference in terms of typing speed or accuracy in pretest ($t = 0.08$, $p = 0.937$).

2.3. Experimental task and procedure

The experiment was comprised of six trials of typical hear-and-type tasks that emulated daily typing work. In each trial, a computer program read out 30 random numbers of 9 digits at a speech rate of 1 digit per second. Participants were told to type out those numbers digit-by-digit immediately after each digit was heard on either type of three numerical keyboards. The experimenter clearly instructed the participant to *not* remember the entire number and type all digits together because doing this would degrade the reaction time and subject herself/himself to memory loss. Each random number was followed by a 2.5 second pause during which participants were reminded to press 'enter' and there was a small pause in between every three digits to imitate natural pause when consecutive numbers were read out (Lin and Wu 2011).

Before data collection started, the experimenter prepared participants for EMG measurement. First, the skin surface was cleaned by rubbing alcohol to remove excessive oil and dead skin cells (and shaved if necessary), and then the electrode was placed around the mid-point between medial epicondyle of humerus and styloid process of ulna (Figure 3, the exact position was determined by palpation) to capture EMG activity of finger flexors (flexor digitorum profundus and flexor digitorum superficialis) during finger movements (Basmajian and Blumenstein 1989).

At the beginning of the experiment, all participants were allowed to adjust the volume and other settings of the typing environment to their preference, including the height of an ergonomic office chair and the height of the keyboard tray. The settings were maintained the same throughout the whole experiment. Before formal experimental trials, participants were given four practice trials of experimental tasks on touch screens. A practice trial contained 30 numbers, and each number contained 10 key presses (nine digits plus an Enter key) if the participant responded correctly. Therefore, the four practice trials would yield about 1200 touch operations on the touch interface (4 trials × 30 numbers/trial × 10 key presses/number = 1200 key presses). According to Sutter (2007), aiming performance on touch devices would level off after 1000 touch operations for novice users; therefore, the amount of practice should be enough for saturating learning effects. During the practice the experimenter also asked participant to type both quickly and accurately. A short, five-minute break was provided to participants after finishing practice trials, and they continued with formal experimental trials.

Before each formal experimental trial, participants were told which keyboard would be used (physical keyboard, touch screen with IVF or touch screen with PVF) and the urgency level of the trial. Each formal experimental trial contained about 300 key presses if the participant reacted correctly (30 numbers × 10 key presses/number = 300 key presses), and there were three urgent trials and three non-urgent trials, respectively. In urgent trials, the experimenter also made it clear to the participant that a performance-based reward would apply and, therefore, they should try their best to type both accurately

Table 1. Pre-test performance by gender.

Gender	Male ($n = 12$)	Female: ($n = 10$)
Age	21.75 (± 2.06)	23.00 (± 4.64)
Speed (keys/min)	96.49 (± 12.22)	96.69 (± 15.83)
Accuracy (%)	91.80 (± 5.30)	90.83 (± 5.14)

and quickly to obtain as much bonus as possible. In contrast, participants were asked to maintain high accuracy in non-urgent trials, where quick responses were not necessary as long as they could catch up with the speech rate. The run order of the six trials was random, and a short, five-minute break was given after three trials so that the participants' performance would not be hampered by fatigue. If excessive fatigue was reported by the participant after any trial, the experimenter would ask the participant to take extra breaks until the fatigue is recovered, since fatigue was found to have adverse effects on typing performance (Liao and Drury 2000). After each trial, a NASA-TLX questionnaire (NASA 2011) was given to the participant to report their subjective workload.

2.4. EMG data collection and processing

EMG data were collected through an Ag–AgCl surface electrode with three leads (positive, negative and ground) attached to the ipsilateral side (right side, Figure 3) of the moving hand and connected to a data encoder (InfinityTM Flexcomp, Thought Technologies, Quebec, Canada). The gain was set to 1000 and the sampling rate was 2048 Hz. Raw EMG activity collected during trials was first band-pass filtered (10–500 Hz) by a Butterworth filter and rectified.

The EMG system was synchronised with the stimulus presentation system through a LabJack® interface (LabJack Corporation, CO). At the beginning of the sound playing procedure, the computer program would trigger the LabJack® interface to send a pulse to a separate channel of the EMG data encoder. The peak value of the oscillatory wave received by the EMG decoder would be identified as the onset of the auditory stimulus. The average duration of auditory stimuli played was 431 ms. A mean and a standard deviation were calculated from 50-ms EMG during the rest period before the onset of each auditory stimulus (Carlsen et al. 2003, Carlsen et al. 2004). The burst points were then identified as the first point when the smoothed EMG amplitude (moving average technique with 200-ms window) exceeds two standard deviations above the mean value from the corresponding period. The identification was automatically processed by a Matlab® procedure (The MathWorks, Inc, MA), and the pre-motor reaction time was calculated based on the lapse between the timing of the verbal stimuli recoded by the synchronisation signals and the identified burst point (Figure 4). Only the pre-motor reaction times from correct responses were included in the statistical analysis.

2.5. Statistical analysis

Behavioural data recorded in Microsoft Excel® files were manually compared with the data log of stimuli by the experimenter to determine the reaction time for each keystroke and to compute frequency of errors in each trial. The data
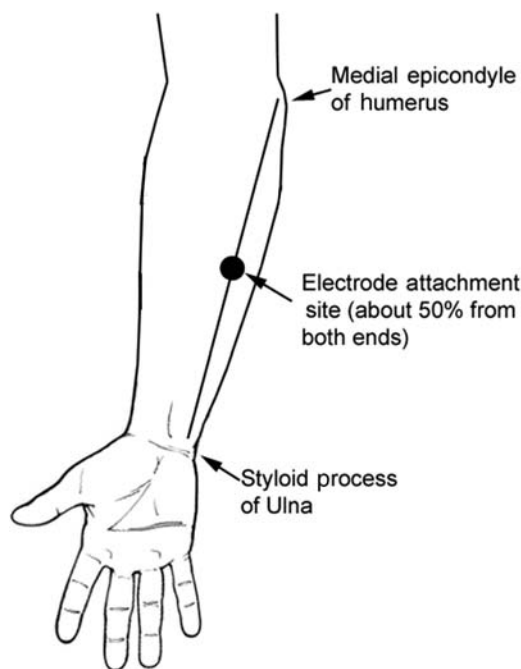
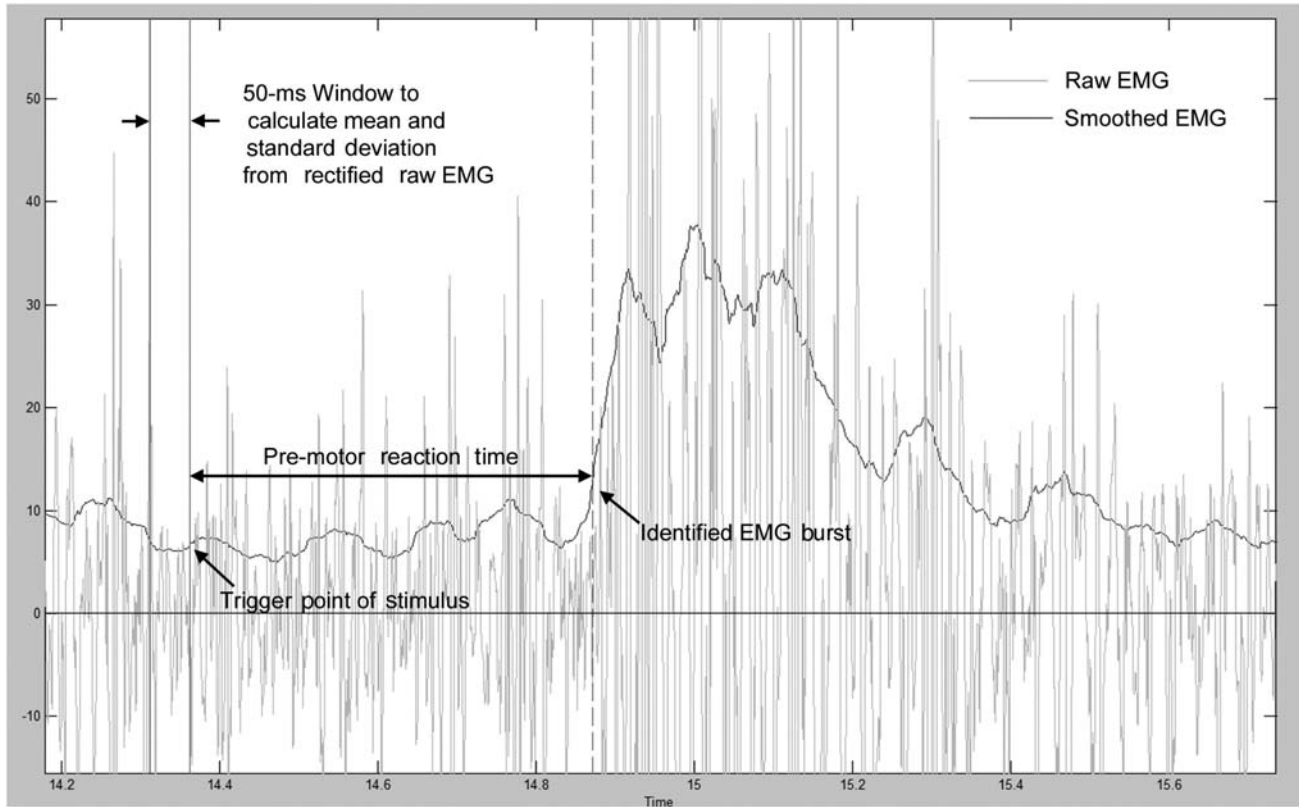

Figure 3. EMG electrode attachment.

Figure 4.   Identification of EMG burst points.

log contained timing of the onset of each auditory stimulus and the reception of the depression signal from each key press. The timing was registered in the unit of milliseconds and the lapse (reaction time) was calculated by subtracting the timing of the onset of the auditory stimulus from the timing of the reception of its corresponding key press. After reaction time for each correct keystroke was available, reaction time for all correct responses was averaged. Whenever an auditory stimulus was not responded correctly (either there was no key or a wrong key was pressed) or an unnecessary key was depressed, an error had occurred and would be registered. The frequency (number) of total errors committed was counted for each trial.

In statistical analysis, average pre-motor reaction time, average reaction time and frequency of errors were examined by analysis of covariance (ANCOVA) to confirm whether the expertise of typing shown in the pretest would significantly influence reaction time or accuracy in the formal test. A general linear model with two by three factorial, within-subject design was assumed for ANCOVA. Keyboards and urgency levels were two main factors considered in the initial statistical models, and individual difference was consider a random factor of the model. The significance (*p*-value) of experimental factors and their effects are reported in Section 3. Similar statistical analysis procedure was applied to subjective workload ratings except that pretest expertise was not considered in analysis of variance models.

3.   Results

3.1.   Pre-motor reaction time

For pre-motor reaction time, effects of keyboards and urgency were both significant (Figure 5). Urgency reduced pre-motor reaction time significantly ($F_{1,101} = 9.08$, $p = 0.003$), while the physical keyboard showed significant shorter pre-motor reaction time than the touch screen keyboards ($F_{2,101} = 4.78$, $p = 0.01$). Tukey's test showed no significant difference between IVF and PVF conditions ($t = 0.765$, $p = 0.7254$). Interaction between keyboards and urgency was not significant, but pretest speed was influential to pre-motor reaction time ($t = -2.16$, $p = 0.033$), i.e. the faster the pretest speed, the shorter the pre-motor reaction time. In summary, greater pre-motor reaction time was observed for touch screens, regardless of feedback conditions, when they are compared to the physical keyboard.
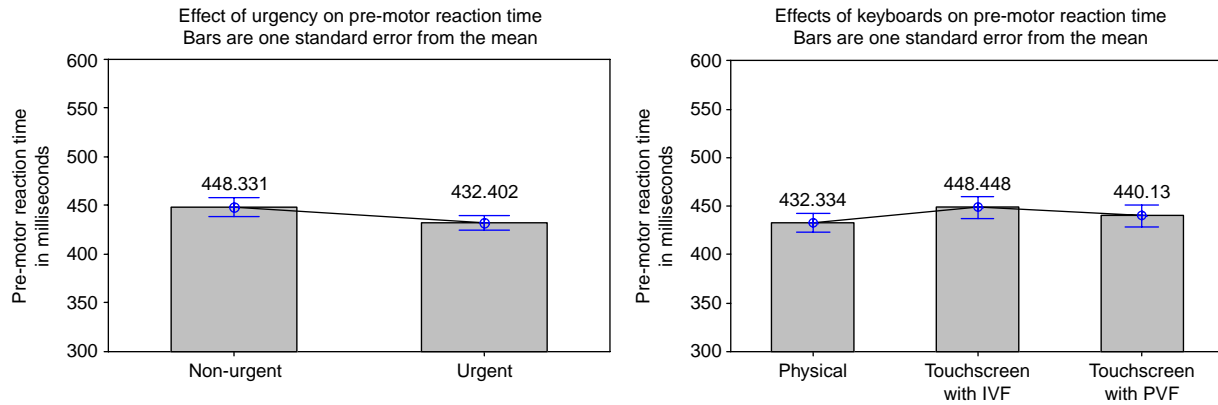
Figure 5. Effect plots for pre-motor reaction time.

3.2. Reaction time

For reaction time, interaction between keyboards and urgency was significant ($F_{1,99} = 4.05$, $p = 0.02$). From the interaction plot (Figure 6), reaction time was reduced significantly under urgency. Further stratified analysis (Table 2) revealed that reaction time on the physical keyboard was significantly shorter than that on the touch screen keyboards regardless of feedback conditions when the situation is urgent (Table 2, Urgent column). Tukey's method showed no significant difference between feedback conditions. In non-urgent situation, however, provision of PVF caused longer reaction time than touch screen keyboard with IVF, and the reaction time on the physical keyboard was the shortest (Tukey's method showed significant difference for all pairs; Table 2, Non-urgent column). In summary, users managed to shorten reaction time on the touch screen with PVF under urgency, resulting in no difference between feedback conditions (reaction time on the physical keyboard was still significantly shorter). When there was no urgency, numerical typing on the touch screen with PVF was significantly slower than that on the touch screen with IVF, and they are both significantly slower than numerical typing on the physical keyboard.

3.3. Typing errors

For typing errors, effects of keyboards and urgency were both significant (Figure 7). Urgency increased errors significantly ($F_{1,100} = 18.91$; $p < 0.001$), but only the touch screen keyboard with PVF elicited significantly more errors than the other
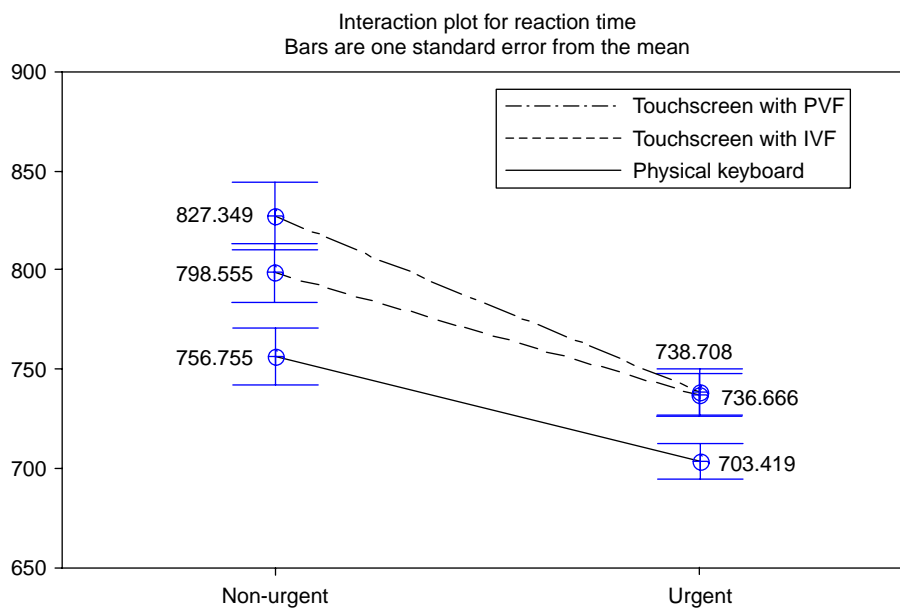

Figure 6. Interaction plot for reaction time.

Table 2. *Post hoc* p-values in stratified analysis.

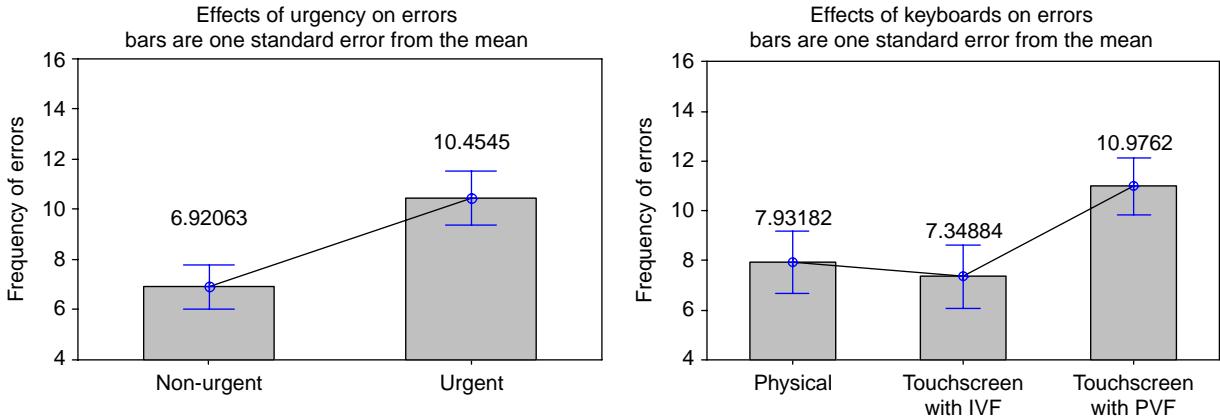| *Post hoc* pair	Urgent	Non-urgent
Physical keyboard versus touch screen with IVF	0.0003	0.0087
Physical keyboard versus touch screen with PVF	0.0007	<0.0001
Touch screen with IVF versus touch screen with PVF	0.9331	0.0111

Figure 7.   Effect plots for errors.

keyboards ($F_{2,100} = 13.57$, $p < 0.001$). Tukey's method showed no significant difference between the touch screen with IVF and the physical keyboard ($t = 0.2229$, $p = 0.9730$). Interaction between keyboards and urgency was not significant, and no significant covariate was found for typing errors. Surprisingly, PVF degraded user performance in accuracy on touch screens, while touch screens with IVF did not differ from physical keyboards in terms of accuracy.

3.4.   Subjective workload

Analysis of subjective ratings from NASA-TLX questionnaires (Figure 8) found significant effects of urgency on all six dimensions of workload (mental workload $F_{1,102} = 21.90$; physical workload $F_{1,101} = 16.37$; temporal workload $F_{1,103} = 77.39$; performance $F_{1,103} = 18.69$; effort $F_{1,103} = 57.95$; frustration $F_{1,101} = 40.06$; all $ps < 0.0001$); temporal and effort dimension showed largest difference between urgency levels (effect size $= 4.78$ and 4.42, respectively). Only temporal workload showed a significant difference between keyboards ($F_{1,103} = 3.95$, $p = 0.022$; Figure 9). Tukey's method revealed that the physical keyboard caused significantly lower temporal workload than touch screen keyboards (p-value of physical keyboard vs. touch screen with IVF $= 0.039$; p-value of physical keyboard vs. touch screen with PVF $= 0.046$). No significant difference was observed between feedback conditions ($t = 0.069$, $p = 0.9974$).

4.   Discussion

In this study, user performance and response complexity of numerical typing on a physical keyboard and a touch screen keyboard were investigated. Given the same physical placement and layout design, users' reaction time on the physical keyboard was significantly faster than that on the touch screen keyboards, no matter whether PVF was provided or not. However, typing accuracy in terms of errors was not significantly different on the two devices when there was no urgency and no PVF. Slower response time on touch screens was caused by slower pre-motor response time in preparing motor execution, implying higher response complexity to type on touch screens. This inference was also supported by higher subjective rating on touch screens in terms of temporal workload. In summary, touch screens appeared to be more complex for motor programming than physical keyboards, while the provision of precise positional feedback did not significantly reduce the complexity as signified by similar pre-motor reaction time.

   The effect of urgency on numerical typing with different keyboard designs was also studied. In general, urgency reduced both pre-motor reaction time and reaction time, regardless of keyboard types, but increased typing errors. The reduction in pre-motor reaction time revealed that, in addition to shortened execution, users might have also speeded up
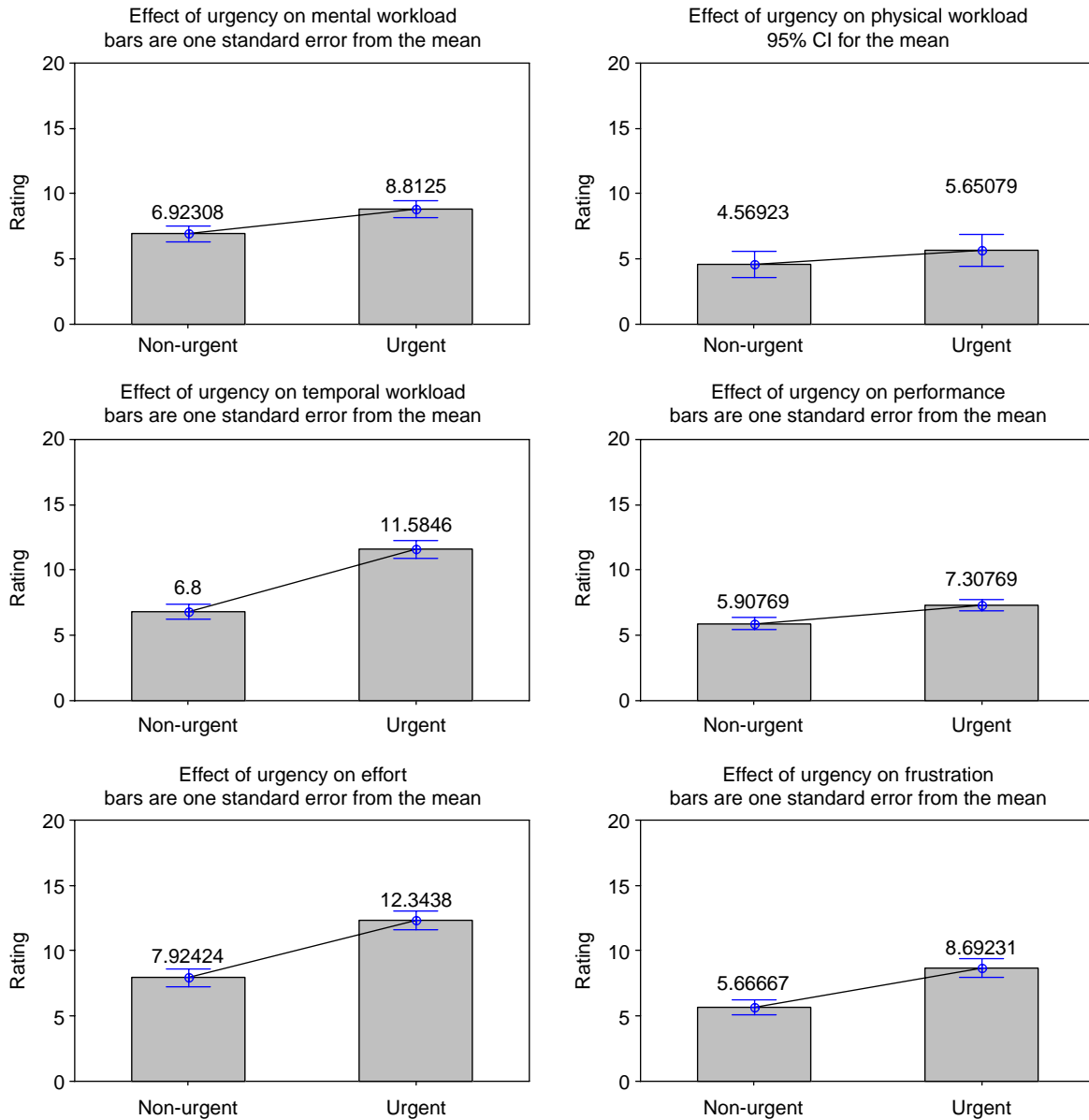
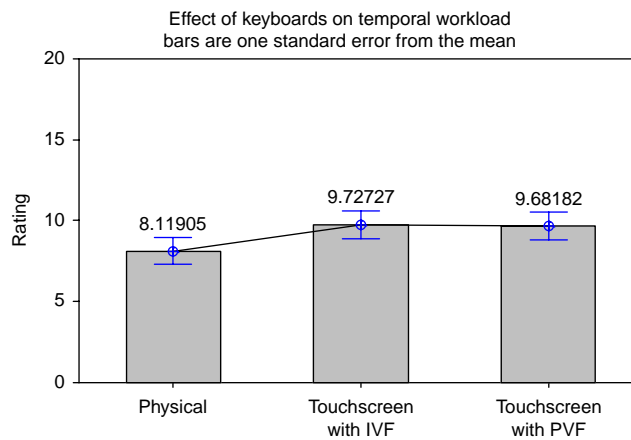Figure 8. Effect plots of urgency on NASA-TLX.

Figure 9. Effect plot of keyboards on temporal workload.

motor preparation under urgency. The increased processing demand on both cognitive and motor aspects of reaction may explain why users reported increased workload in all six dimensions including mental workload and effort. Moreover, urgency had a significant interaction with PVF on reaction time but not on pre-motor reaction time. Apparently, users managed to reduce motor execution time significantly on the touch screen with precise feedback under urgency so that reaction time did not differ between feedback conditions. The fact that a significant interaction was not found in pre-motor reaction time is consistent with the observation that pre-motor reaction time was relatively unaffected by peripheral execution factors such as movement speed, i.e. users were able to reduce the reaction time on the touch screen with PVF under urgency (possibly by moving faster) but were unable to do the same on the pre-motor reaction time.

Surprisingly, PVF on the touch screen used in this study did not help reduce reaction time and increased errors significantly. The effect of PVF was not significant on pre-motor reaction time (it reduced pre-motor reaction time insignificantly, Figure 5), but reaction time increased significantly in non-urgent situation with PVF, implying that longer motor execution was needed for using precise information provided by PVF. After the force-paced nature of the numerical hear-and-type task and the interaction between feedback and urgency were both considered, time pressure might have played an important role in causing more errors when PVF was provided. In non-urgent situation, users needed more time for motor execution to benefit from precise feedback. However, the average reaction time under this condition (about 827 ms) was close to the interval between two stimuli (one second). So, users felt higher time pressure (in non-urgent situation, temporal workload was 6.9 and 6.8 of the touch screen with and without feedback, respectively) and were likely to commit typing errors. In urgent situation, although users reduced the motor execution time probably by moving more rapidly, the rapid aiming may cause greater spatial variability of touch positions and thus more errors.

Comparison studies in the literature reported somewhat inconsistent results for user performance on touch screens and physical keyboards because several biomechanical and design factors were involved as well as different experimental task used. Sears (1991) compared touch screens with physical keyboards in alphabetical typing, but touch screens were optimally designed with enlarged key size and shifted touch-sensitive regions to accommodate touch biases (Sears 1991). If the touch screen keyboard used in Sears' study had not been optimally designed but just simulated the physical keyboard, the results of performance time and accuracy might have been different. Detterman (1993) and Thomas and McClelland (1996) both reported slower reaction time and lower accuracy in numerical typing using touch screens, while Chung et al. (2009) claimed that touch screens or physical numerical keyboards make no difference in numerical typing speed or accuracy for young adults. Different findings in those studies may be explained by technology advances: the latter study used faster, capacitive overlay technology compared to infrared or surface wave technology used in the former two studies. However, the statement that touch screens are not different from physical keyboards might also be questionable because the touch screen used in the study still benefitted from its larger buttons, generous spacing and closer position to the user. The current study is so far, to the authors' knowledge, one of few studies that compares two alternative typing devices under the same physical constraints. The findings are, given identical physical placement and layout design, touch screens indeed require longer reaction time for numerical typing due to not only slower response (possibly attributable to mechanical factors, e.g. slow response time or slippery surface on touch screens) but also higher response complexity that may involve cognitive factors. In order to improve user performance on touch screens, more understanding about the cause of high response complexity is needed in addition to amelioration of technology.

Furthermore, the study also suggested that visual precise feedback be used with caution. Lack of tactile feedback was considered a main reason for poor typing performance on touch screens because tactile sense provided position and orientation of fingers on the physical keyboard (Rao and Gordon 2001; Rabin and Gordon 2004). To compensate the sense of the engagement and visualise the touch, innovative visual feedback, such as ripple effects suggested in Benko and Wigdor (2010), has been proposed. Although PVF is supposed to provide extra information about the position of touch in addition to the engagement to the key provided by other visual effects (such as the inversed/highlighted image of the key), the current study found that this information may not benefit users in either accuracy or speed of typing if the presentation of precise positional information is not well-designed. Possible reasons include that the PVF used in this study become distractions and/or resource competitions in visual modality on which eye–hand coordination already heavily relied, and that users needed to move their fingers farther away to use the positional information and the additional movements interrupted their motor reactions. In conclusion, the PVF should be designed with caution to avoid distractions, visual competitions or motor interruptions. Future usability studies are needed to clarify the usefulness of providing precise information through visual channels systematically.

The current study has some limitations. First, lack of kinematic data makes confirmation of some inferences, e.g. users shortened motor execution time in urgency on the touch screen with visual feedback and thus generated greater spatial dispersion, impossible. If detailed kinematic data were collected, changes in motion characteristics such as velocity, acceleration and spatial variability would have been clearer. Second, the results of this study apply to young and healthy users only. Many studies have demonstrated that users in different age groups or with disabilities showed different

behaviours when they interacted with touch screens (Murata and Takahashi 2008; Chung et al. 2009; Duff et al. 2010; Sesto et al. 2012). Although healthy, young adults are the majority of user groups who interact with technology equipments, needs of older users, children and physically challenged people should not be ignored and require further study. Also, the participant group of this study were skilled typists who, on average, were able to produce more than 95 keystrokes per minute with 90% accuracy on both physical keyboards and touch screens. On one hand, the touch screen keyboard might be difficult to use for novice users, and their performance on the touch screen may be inferior to that on the physical keyboard in terms of both reaction time and accuracy due to lack of experience. On the other hand, professional typists, especially those who are able to do touch typing without seeing the keyboard, might be able to eliminate the difference caused by the complexity of motor programming on touch screens, i.e. the pre-motor reaction time might not be significantly different on the physical keyboard and the touch screen. This does not mean, however, that professional typist could achieve the same performance on touch screens and physical keyboards because the reaction time is still subject to other electromechanical and biomechanical factors. Further studies are needed to confirm these speculations. Third, although seating postures are common for office work, users are often standing when they use public kiosks. User performance may change due to standing posture, and further investigation is needed for comparing two alternative typing devices using standing posture.

Despite the above-mentioned limitations, this study provides a relatively useful comparison without confounding factors between touch screen numerical keyboards and physical numerical keypads. The study concluded that numerical typing on touch screens is not inherently different in accuracy but slower in speed, possibly due to slower pre-motor reaction time. Furthermore, the study provides information about touch screens being more complex for users to generate responses and cautions for designing PVFs. Given prevalent use of touch screens, future studies on reducing response complexity on touch screens through innovative feedback design are encouraged.

Note

[1]   Email: cl224@buffalo.edu

References

Anon. 2010. "Labor Force Statistics from the Current Population Survey – Demographics.", http://www.bls.gov/cps/demographics.htm#age

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.

Basmajian, J. V., and R. Blumenstein. 1989. "Electrode Placement in Electromyographic Biofeedback." In *Biofeedback: Principles and Practice for Clinicians*, edited by J. V. Basmajian, 363–377. Baltimore, MD: Williams and Wilkins.

Benko, H., and D. Wigdor. 2010. "Imprecision, Inaccuracy, and Frustration: The Tale of Touch Input." In *Tabletops – Horizontal Interactive Displays*, edited by J. Karat, and J. Vanderdonckt, 249–275. London: Springer.

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 R. J. Anderson, 313–364. Indianapolis, IN: Wiley.

Carlsen, A., R. Chua, J. T. Inglis, D. Sanderson, and I. Franks. 2003. "Startle Response is Dishabituated During a Reaction Time Task." *Experimental Brain Research* 154 (4): 510–518.

Carlsen, A. N., R. Chua, J. T. Inglis, D. J. Sanderson, and I. M. Franks. 2004. "Prepared Movements are Elicited Early by Startle." *Journal of Motor Behavior* 36 (3): 253–264.

Chung, M. K., D. Kim, S. Na, and D. Lee. 2009. "Usability Evaluation of Numeric Entry Tasks on Keypad Type and Age." *International Journal of Industrial Ergonomics* 40 (1): 97–105.

Coombes, S. A., J. H. Cauraugh, and C. M. Janelle. 2007. "Emotional State and Initiating Cue Alter Central and Peripheral Motor Processes." *Emotion* 7 (2): 275–284.

Detterman, D. K. 1993. *Response devices and cognitive tasks*. Cleveland, OH: Case Western Reserve University.

Duff, S. N., C. B. Irwin, J. L. Skye, M. E. Sesto, and D. A. Wiegmann. 2010. "The Effect of Disability and Approach on Touch Screen Performance During a Number Entry Task." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 54 (6): 566–570.

Eisenberger, R., and J. Cameron. 1996. "Detrimental Effects of Reward: Reality or Myth?" *American Psychologist* 51 (11): 1153–1166.

Fitts, P. M. 1954. "The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement." *Journal of Experimental Psychology* 47 (6): 381–391.

Franks, I. M., P. Nagelkerke, M. Ketelaars, and P. Van Donkelaar. 1998. "Response Preparation and Control of Movement Sequences." *Canadian Journal of Experimental Psychology* 52 (2): 93–101.

Glencross, D. J. 1980. "Response Planning and the Organization of Speed Movements." In *Attention and Performance*, edited by R. S. Nickerson, 107–125. Hillsdale, NJ: Lawrence Erlbaum.

Goode, N., M. G. Lenné, and P. Salmon. 2012. "The Impact of On-Road Motion on BMS Touch Screen Device Operation." *Ergonomics* 55 (9): 986–996.

Gordon, A. M., and J. F. Soechting. 1995. "Use of Tactile Afferent Information in Sequential Finger Movements." *Experimental Brain Research* 107 (2): 281–292.

Huang, H., and H.-H. Lai. 2008. "Factors Influencing the Usability of Icons in the LCD Touchscreen." *Displays* 29 (4): 339–344.

Ketelaars, M. A. C., M. I. Garry, and I. M. Franks. 1997. "On-line Programming of Simple Movement Sequences." *Human Movement Science* 16 (4): 461–483.

Lee, J.-H., E. Poliakoff, and C. Spence. 2009. "The Effect of Multimodal Feedback Presented via a Touch Screen on the Performance of Older Adults: Haptic and Audio Interaction Design." In *Computers Helping People with Special Needs*, edited by M. Altinsoy, U. Jekosch, and S. Brewster, 128–135. Berlin/Heidelberg: Springer.

Li, X., Z. Liang, M. Kleiner, and Z.-L. Lu. 2010. "RTbox: A Device for Highly Accurate Response Time Measurements." *Behavior Research Methods* 42 (1): 212–225.

Liao, M. H., and C. G. Drury. 2000. "Posture, Discomfort and Performance in a VDT Task." *Ergonomics* 43 (3): 345–359.

Lin, C.-J., and C. Wu. 2011. "Factors Affecting Numerical Typing Performance of Young Adults in a Hear-and-Type Task." *Ergonomics* 54 (12): 1159–1174.

Maguire, M. C. 1999. "A Review of User-Interface Design Guidelines for Public Information Kiosk Systems." *International Journal of Human-Computer Studies* 50 (3): 263–286.

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.

Martin, G. L. 1988. "Configuring a Numeric Keypad for a Touch Screen." *Ergonomics* 31 (6): 945–953.

Murata, A., and R. Takahashi. 2008. "Optimal Slope of Touch Panel- Comparison Between Young and Older Adults." *Fourth International Workshop on Computational Intelligence and Applications*, Hiroshima University, Japan.

NASA. 2011. "NASA TLX: Task load index.", http://humansystems.arc.nasa.gov/groups/TLX/paperpencil.html

O'Hara, J. M., J. C. Higgins, and W. S. Brown. 2008. *Human Factors Considerations with Respect to Emerging Technology in Nuclear Power Plants*. Upton, NY: Energy Sciences and Technology Department, Brookhaven National Laboratory.

Poupyrev, I., S. Maruyama, and J. Rekimoto. 2002. "Ambient Touch: Designing Tactile Interfaces for Handheld Devices." *Proceedings of the 15th annual ACM symposium on User interface software and technology*, 51–60. Paris, France: ACM.

Rabin, E., and A. Gordon. 2004. "Tactile Feedback Contributes to Consistency of Finger Movements During Typing." *Experimental Brain Research* 155 (3): 362–369.

Rao, A., and A. Gordon. 2001. "Contribution of Tactile Information to Accuracy in Pointing Movements." *Experimental Brain Research* 138 (4): 438–445.

Sears, A. 1991. "Improving Touchscreen Keyboards: Design Issues and a Comparison with Other Devices." *Interacting with Computers* 3 (3): 253–269.

Sears, A., D. Revis, J. Swatski, R. Crittenden, and B. Shneiderman. 1993. "Investigating Touchscreen Typing: The Effect of Keyboard Size on Typing Speed." *Behaviour and Information Technology* 12 (1): 17–22.

Sears, A., and B. Shneiderman. 1991. "High Precision Touchscreens: Design Strategies and Comparisons with a Mouse." *International Journal of Man-Machine Studies* 34 (4): 593–613.

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

Sesto, M. E., C. B. Irwin, K. B. Chen, A. O. Chourasia, and D. A. Wiegmann. 2012. "Effect of Touch Screen Button Size and Spacing on Touch Characteristics of Users With and Without Disabilities." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 54 (3): 425–436.

Shin, G., and X. Zhu. 2011. "User Discomfort, Work Posture and Muscle Activity While Using a Touchscreen in a Desktop PC Setting." *Ergonomics* 54 (8): 733–744.

Spijkers, W. A. C., and A. Walter. 1985. "Response Processing Stages in Choice Reactions." *Acta Psychologica* 58 (2): 191–204.

Sutter, C. 2007. "Sensumotor Transformation of Input Devices and the Impact on Practice and Task Difficulty." *Ergonomics* 50 (12): 1999–2016.

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.

Thomas, B., and I. Mcclelland. 1996. "The Development of a Touch Screen Based Communications Terminal." *International Journal of Industrial Ergonomics* 18 (1): 1–13.