Elderly Text-Entry Performance on Touchscreens

Hugo Nicolau Joaquim Jorge
IST / Technical University of Lisbon / INESC-ID
{hman, jaj}@vimmi.inesc-id.pt

ABSTRACT

Touchscreen devices have become increasingly popular. Yet they lack of tactile feedback and motor stability, making it difficult effectively typing on virtual keyboards. This is even worse for elderly users and their declining motor abilities, particularly hand tremor. In this paper we examine text-entry performance and typing patterns of elderly users on touch-based devices. Moreover, we analyze users' hand tremor profile and its relationship to typing behavior. Our main goal is to inform future designs of touchscreen keyboards for elderly people. To this end, we asked 15 users to enter text under two device conditions (mobile and tablet) and measured their performance, both speed- and accuracy-wise. Additionally, we thoroughly analyze different types of errors (insertions, substitutions, and omissions) looking at touch input features and their main causes. Results show that omissions are the most common error type, mainly due to cognitive errors, followed by substitutions and insertions. While tablet devices can compensate for about 9% of typing errors, omissions are similar across conditions. Measured hand tremor largely correlates with text-entry errors, suggesting that it should be approached to improve input accuracy. Finally, we assess the effect of simple touch models and provide implications to design.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – Input devices and strategies

General Terms

Design, Experimentation, Human Factors.

Keywords

Elderly, Touchscreen, Text-Entry, Tremor, Mobile, Tablet.

1. INTRODUCTION

There was a time where touchscreen technology was affordable to a few. Nowadays, this technology is widely spread among different devices, applications and environments, such as ATM machines, information kiosks, ticket machines, health control devices, etc. Most of us use touchscreens on a daily basis due to its enormous success in mobile devices. Indeed, these are increasingly replacing keypad-based applications.

The ability to directly touch and manipulate data on the screen without intermediate devices has a strong appeal, since it provides for a more natural and engaging experience. Moreover, touchscreens offer high flexibility, making it possible to display different interfaces on the same surface or to adapt to the users' needs and/or preferences [5]. For all their advantages, touch interfaces present similar challenges: they lack both the physical stability and tactile feedback ensured by keypads, making it harder for people to accurately select targets. This becomes especially pertinent to elderly people who suffer from increased hand tremor [16]. This effect becomes worse for interfaces that feature small targets and spacing [9], such as virtual keyboards.

Figure 1. Participant typing on a touchscreen device.

Indeed, mobile text-input is a major challenge for elderly users. Since text-entry is a task transversal to many applications, such as basic communications, managing contacts, editing documents, web browsing, etc., these users are excluded from the innumerous opportunities brought by touch devices to different domains: social, professional, leisure, entertainment, shopping, communication, or healthcare. Still, touch interfaces have the potential to reduce this "technology gap", due to their high customizability, which makes them appropriate to custom-tailored or adaptive solutions that can fit the needs of different users. This highlights the need to understand how elderly people input text on current touchscreen devices. Because there is little or no quantified knowledge on the problems that these users experience with standard virtual keyboards, it is difficult to improve them. Furthermore, since touch interfaces are highly customizable, empirical data can be used to automate and provide user-dependent solutions.

Our goal with this work was to provide the knowledge needed to design both effective and efficient text-entry solutions for elderly people. We performed evaluations with 15 users (Figure 1) and two touch-based devices (mobile and tablet), analyzing the effect of hand tremor on text-entry performance. Also, we thoroughly analyze the users' typing behaviors and performance errors, as well as their comments. We were interested in answering questions such as: What will be the most common input errors and their causes? Will hand tremor be correlated with input performance? Will tablet devices compensate mobile difficulties? How can we enhance text-entry accuracy?

Our main contribution is a thorough understanding of text-entry performance in touch-based devices by elderly users. We provide an empirical body of knowledge to leverage future development of virtual keyboards and a better understanding on how text-input performance correlates to hand tremor. We also demonstrate the potential and virtues of simple touch models and provide design implications that should motivate researchers to develop more effective solutions.

2. RELATED WORK

We discuss related work in two areas: first, we look into previous research that attempts to better understand tremor and how it affects elderly users. Second, we discuss HCI research aimed at creating new touch-based solutions for older adults.

2.1 Elderly and Tremor

Generally, tremor is defined as any involuntary, approximately rhythmic, and roughly sinusoidal motion around a joint. Tremor is present in all individuals and is the most common form of movement disorder with an increased prevalence among elderly individuals [16]. There are two classification systems used in evaluating tremor: type of movement and cause. The first distinguishes whether tremor occurs at rest (resting tremor) or is caused by action. Tremors associated to movement (action) include postural tremor, which occurs with maintained posture; kinetic or intention tremor, which occurs with movement from point to point; and task-specific tremor, occurring only when doing highly skilled activity. Postural tremor is usually detected by having a patient holding the arms stretched out in front, while kinetic or intention tremor can be tested by using the finger-to-nose maneuver. The second tremor classification is by cause. Tremor can be due to a variety of conditions both physiologic and pathologic. Physiological tremor in healthy individuals is characterized as a low amplitude postural tremor with a modal frequency of 8–12 Hz [4] in the hands. Pathological tremor is the most extensive movement disorder and can be observed in several pathologies, such as: Essential Tremor, Parkinson's disease, dystonic disorders, cerebellar disease or head trauma. Most of these pathologies are more prevalent among elderly individuals.

Currently accepted standards for evaluating motor performance include subjective measures such as self-reporting and clinical rating scales. The Unified Parkinson's Disease Rating Scale (UPDRS) rates motor manifestations from 0 to 4, where higher scores denote greater severity [7]. However, there are certain limitations in the utility of this rating instrument, because scores are subjective and imprecise. On the other hand, objective motor assessment is an open challenge for movement disorder specialists. Handwriting and drawing samples have long been used to quantify tremor during movement due to their simplicity [1]; however, these tests are not suitable for measuring resting or postural tremor. Accelerometers are currently one of the most commonly used instrument in tremor studies, since they are capable of providing reliable and objective indices by measuring linear acceleration. Many tremor quantification algorithms use power spectral analysis in the frequency domain [14] and define tremor amplitude as the amplitude of a peak in the power spectrum between 3 and 7 Hz. Analyzing the peak amplitude in the 7-12 Hz spectrum may also prove worthwhile to measure physiological tremor. Overall, objective measures of tremor disorders motivated much research by clinicians in the last decades and it will be of significant relevance to the HCI community as well. With the global increase of the senior population, understanding, modeling and dealing with tremor will be a significant concern in designing future assistive technologies.

2.2 Elderly, Touch, and Text-Entry

There is a large body of work that tries to understand and maximize performance of users when interacting with touch interfaces. Past research has investigated optimal target size, spacing and position [9] to derive recommendations and general guidelines for older adults when these interfaces. Still, most approaches do not consider the particular challenges of text-entry: large number of targets, small key size and spacing. Solutions for able-bodied users have been proposed in order to deal with incorrect characters. Gunawardana et al. [8] presented a method to expand or contract key areas for each press using language models, while others have proposed using touch models to adapt to individual typing patterns [5] and improve overall input accuracy. While text-prediction features have also been explored, older adults usually dislike them [11]. Nonetheless, few researchers have explored the specific needs of elderly users in touch typing tasks.

Chung et al. [3] showed that both younger and older users preferred a touchscreen keypad for numeric entry tasks, since it did not force them to divide their attention between the input device and screen content. Wobbrock et al. [18] proposed a stylus-based approach that uses edges and corners of a reduced touch screen to enable text-entry tasks, showing an increase of accuracy and motion stability for users with motor impairments. Similarly, Barrier Pointing [6] uses screen edges and corners to improve pointing accuracy. By stroking towards the screen barriers and allowing the stylus to press against them, users can select targets with greater physical stability. Wacharamanotham [17] takes a similar approach by proposing a technique that uses swipe gestures towards the screen edges in order to select targets. Although these works insightfully explore the device physical properties to aid people interacting with touchscreens, there is little empirical knowledge about elderly users performing text-entry tasks with traditional virtual keyboards. Previous research does not take into consideration elder challenges (such as tremor) that might affect their use of virtual keyboards. The study reported in this paper bridges this gap by analyzing their performance when typing with touchscreen devices, enabling designers to take advantage of this knowledge to build future solutions.

3. USER STUDY

Touch screen devices are increasingly replacing their button-based counterparts. The physical stability and haptic feedback once provided by buttons are being lost, which makes it harder to accurately select targets. This is especially relevant in text-entry tasks due to both small target size and spacing. In this user study we evaluate two different types of touch devices – mobile phone and tablet – and thoroughly analyze how elderly users enter text.

3.1 Research Questions

This user study aims to answer four main research questions:

1. *How do elderly users perform speed and accuracy wise in touch-based devices?*
2. *What are the most common types of errors and causes?*
3. *Do tablet devices compensate the difficulties of elderly users when using mobile phones?*
4. *Does tremor affect text-entry performance? If yes, how does user performance correlate with hand tremor?*

3.2 Participants

Fifteen participants, eleven females and four males, took part in our user study. Their age ranged from 67 to 89 with a mean of 79 (*sd*=7.3) years old. All participants were right-handed. They were recruited from a local social institution and no pre-screening to recruit participants with or at risk of developing tremor disorders was performed. None of the participants had severe visual impairments and all were able to see screen content. Twelve of the participants owned a mobile phone, however they were only able to receive and make calls. Only one participant had used touchscreen technology before, but had never entered text. Regarding QWERTY familiarity, six participants had used this type of keyboard whether in typing machines (four participants) or personal computers (two participants).

3.3 Procedure

This user study had two main phases: familiarization and evaluation. At the beginning of the first phase, participants were told that the overall purpose of the study was to investigate how text-entry performance is affected by the type of device. Following this, participants filled in a pre-questionnaire about demographics and mobile phone usage. We then explained and exemplified to them how to use a virtual keyboard. Although most participants were reluctant to interact with the devices at the beginning, they seamlessly coped with the "touch-to-select" metaphor and easily understood how to write. Nevertheless, because most of them were not familiar with touch devices and QWERTY keyboards, we asked participants to perform two familiarization tasks using each device. The first consisted in entering single letters. They had to copy a letter, displayed at the top of the screen, to a text box. Participants performed this task for 10 minutes (to guarantee an equal amount of training across individuals). The second task consisted in copying sentences. Error correction (delete) was not available. The sentences had a maximum of five words, similar to those presented on the evaluation phase. Participants performed this task for 20 minutes.

In the evaluation phase, we started by assessing the users capabilities regarding tremor (postural and action tremor) applying two different methods. We first asked participants to draw an Archimedes spiral with each hand without leaning hand or arm on table [1]; we then asked participants to hold the mobile device at the arm's length for 30 seconds with each hand and remain still, while we captured data from the accelerometer sensor [15]. Subjects were then informed about the experiment and how to use our evaluation application. We evaluated the participants' performance with two devices: mobile phone and tablet.

Before each condition participants had a five minute practice trial to get used to the virtual keyboard. We did not force participants to interact with a specific finger, thus they were allowed to choose the most comfortable typing strategy, as long as it was consistent during that condition. For the mobile phone condition, participants had to hold it in their hand, since it is a handheld device (Figure 1); for the tablet device condition, it was placed on the table in front of them. For each evaluation condition, participants copied five different sentences (first sentence was a practice trial), displayed one at a time, at the top of the screen (Figure 2). Copy typing was used to reduce the opportunity for spelling and language errors, and to make error identification easier. Participants were instructed to type phrases as quickly and accurately as possible. Both required and transcribed sentences were always visible. Error correction (delete key) was not

Figure 2. Screen shot of evaluation application. Participants were not able to correct errors. The button 'Avançar' allowed them to continue to the next

available, since we wanted to capture typing performance regardless of correction strategies. Participants were told that they could not correct errors and were instructed to continue typing if an error occurred. Once participants had finished entering each sentence, they pressed the 'next' button. After the five sentences were entered, we asked them to perform the same tasks with a different device. The order of conditions was counter balanced to avoid bias associated with experience. The evaluation procedure took approximately 40 minutes per participant. Each subject entered a total of 10 different sentences. These sentences were extracted from a written language *corpus*, and each one had five words with an average size of 4.48 characters and a minimum correlation with the language of 0.97. Sentences were chosen randomly such that no sentence was written twice per participant.

3.4 Apparatus

An HTC Desire and ASUS Transformer TF101 Tablet were used during the user study. A QWERTY virtual keyboard, similar to android's SDK keyboard, was used in both devices (Figure 2); for the HTC Desire each key was 10x10mm on landscape mode, while for the ASUS tablet each key was 20x10mm. Letters were entered when the user lift his finger from keys. Neither word prediction nor correction was used. All participants' actions were logged through our evaluation application and the user study was filmed to observe the participants' behaviors.

3.5 Dependent Measures

The performance during the text-entry task was measured using different quantitative variables [12]: *words per minute* (*WPM*), *minimum string distance* (*MSD*) error rate, and character-level errors (*substitutions* – incorrect characters, *insertions* – added characters, and omissions – omitted characters). Qualitative measures were also gathered at the end of the experiment by debriefing each participant. We also gathered tremor-related measures of each participant before text-entry tasks in order to characterize their level of impairment.

3.6 Design and Analysis

We used a within subjects design where each participant tested all conditions. For each device condition each participant entered 5 sentences (1 practice + 4 test), resulting in a total of 20 sentences per participant. In summary the study design was: 15 participants x 5 sentences x 2 devices. We performed Shapiro-Wilkinson tests of the observed values for *WPM*, *KSPC*, *MSD error rate,* types of errors and tremor measures. If dependent variables were normally distributed we applied parametric statistical tests, such as repeated measures ANOVA, t-test, and Pearson correlations. On the other hand, if measures were not normally distributed, we used non-parametric tests: Friedman, Wilcoxon, and Spearman correlations. Bonferroni corrections were used for post-hoc tests.

4. RESULTS

Our goal was to understand how elderly people input text with traditional touchscreen devices. We describe and characterize each user‘s tremor profile and relate it with text-entry performance. Moreover, we analyze input speed and accuracy for both device conditions, focusing on type of errors.

4.1 Tremor Profile

Task-specific tremor, which is a type of action tremor, was measured in both hands, using the Archimedes spiral test. The drawings were classified by a clinical professional as Absent, Slight, Moderate, Severe or Marked. For the dominant-hand drawings, 7 participants (46.7%) showed no tremor, 4 (26.7%) showed slight tremor, 1 participants (6.7%) demonstrated moderate tremor, 2 (13.3%) showed severe tremor, and 1 participant (6.7%) demonstrated marked tremor. Regarding the non-dominant hand drawings, 5 participants (33.3%) showed absence of tremor, 6 participants (40%) showed slight tremor, 1 participant (6.7%) demonstrated moderate tremor, 1 (6.7%) showed severe tremor, and 2 participants (13.3%) demonstrated marked tremor. Figure 3 illustrates some examples of drawings.

Table 1. Postural hand oscillation for all axes (m/s^2).

mean (*sd*)	**X**	**Y**	**Z**	**XYZ**
Dominant	0.19 (*.07*)	0.15 (*.06*)	0.3 (*0.13*)	0.14 (*.04*)
Non-Dominant	0.17 (*.07*)	0.12 (*.02*)	0.3 (*.15*)	0.1 (*.03*)

In addition to subjective measures, we also measured tremor through the device’s accelerometer. Particularly, we measured the postural – a type of action – tremor. From the captured data we analyzed three main values: acceleration standard deviations, which correspond to hand oscillations [2]; the peak amplitude in the power spectrum of 3 to 7 Hz, and 7 to 12 Hz. We report the peak amplitude in different frequency ranges since physiological and pathologic tremors are usually distinguishable and may affect users’ performance differently. Results for hand oscillation (Table 1) showed a mean magnitude of 0.186 m/s^2 (*sd*=.074), 0.15 m/s^2 (*sd*=.06), 0.3 m/s^2 (*sd*=.13), and 0.137 m/s^2 (*sd*=.044) for X, Y, Z, and XYZ axis, respectively. Regarding the non-dominant hand, due to a logging issue we were only able to record 9 of the 15 participants’ accelerometer data. Mean oscillation was 0.174 m/s^2 (*sd*=.07), 0.115 m/s^2 (*sd*=.024), 0.3 m/s^2 (*sd*=.149), and 0.101 m/s^2 (*sd*=.03), for X, Y, Z, XYZ axis, respectively. Regarding the frequency analysis, results showed a mean peak magnitude of 0.362 m/s^2 (*sd*=0.429), and 0.17 m/s^2 (*sd*=0.17), for the 3 to 7 Hz, and 7 to 12 Hz, respectively. Concerning the non-dominant hand, results showed a mean peak magnitude of 0.17 m/s^2 (*sd*=0.162), and 0.105 m/s^2 (*sd*=0.175) for the 3 to 7 Hz, and 7 to 12 Hz, respectively. It is worth noticing that the results for each of the frequency ranges show high standard deviation, suggesting that tremor severity varies widely among participants.

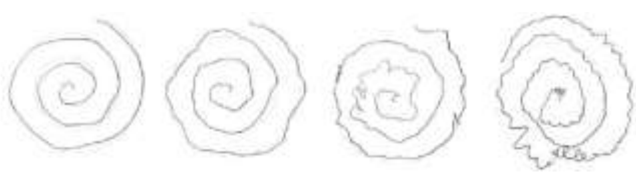

Figure 3. Archimedes spiral drawings. From left to right: absent, slight, severe, marked.

4.2 Text-Entry Performance

In this section we thoroughly analyze input performance regarding speed and accuracy for both device conditions. During text-entry tasks, all participants consistently used their non-dominant hand to hold the mobile device and dominant index finger to select intended keys. However, one of the participants was unable to use the mobile device, due to visual impairments. Although text font was large (Figure 2), participant #9 was not able to read keyboard characters, and therefore did not complete the *Mobile* condition.

4.2.1 Input Speed

To assess speed, we used the words per minute (WPM) text input measure calculated as *(transcribed text – 1) * (60 seconds / time in seconds) / (5 characters per word)*.

Tablets allow higher input rates. Participants typed an average of 4.73 *WPM* (*sd*=3.06) in *Mobile* and 5.07 *WPM* (*sd*=2.93) in *Tablet* conditions. A paired-samples t-test was conducted to evaluate the effect of device on text-entry speed. A statistically significant increase in *WPM* [$t(13) = -2.752$, $p<.05$] was found, suggesting that participants can achieve higher input rates with tablet devices.

Experience makes the difference. Overall, input rate was strongly correlated with QWERTY keyboard experience, which explains 46% [Spearman rho=.648, n=14, $p<.05$] and 29% [Spearman rho=.534, n=15, $p<.05$] of shared variance for *Mobile* and *Tablet* conditions, respectively; that is, participants that used a (non touch-based) QWERTY keyboard in the past inputted text faster.

4.2.2 Input Accuracy

We measured the quality of the transcribed sentences using the *Minimum String Distance* (MSD) *Error Rate*, calculated as MSD(*required sentence*, *transcribed sentence*) / *mean size of alignments* x *100*. Figure 4 illustrates participants’ *MSD Error Rate* for both *Mobile* and *Tablet* conditions.

Experience is not enough. As opposed to the results obtained in input speed, there was a weak correlation between quality of transcribed sentences and QWERTY experience for *Mobile* [Pearson r=.145, n=14, p=.621] and *Tablet* [Pearson r=.155, n=15, p=.58] conditions. This result suggests that previous experience is not enough to compensate for typing errors.

Tablets compensate difficulties. Participants achieved an average *MSD error rate* of 25.97% (*sd*=19.72%) and 16.55% (*sd*=11.9%) in *Mobile* and *Tablet* conditions, respectively. Results show a statistically significant decrease of 9.42%, which suggests that elderly users indeed benefit from tablet devices, either due to key size or its static position (on the table).

Hand tremor explains (mobile) error rates. In *Mobile* condition, *Hand Oscillation* of the non-dominant hand in the *Y* [Pearson

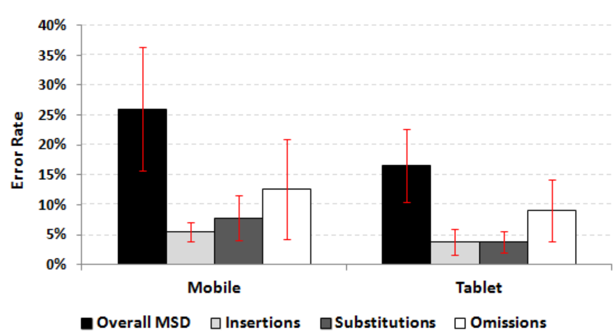

Figure 4. Overall MSD, Insertion, Substitution, and Omission error rate for each device condition. Error bars denote 95% confidence intervals

r=.751, n=9, p<.05] and Z [Pearson r=.613, n=9, p=.079] *axis* were strongly correlated with *MSD error rate*. Since the mobile device was held in the non-dominant hand during text-entry tasks, these results suggest that hand oscillations can explain as far as 56.4% of shared variance. As for *Tablet* condition, we found no strong correlations between tremor measures and *MSD error rate*.

4.2.3 Typing Errors

This section presents a fine grained analysis by categorizing types of input errors: *insertions*, *substitutions*, and *omissions* [12]. Figure 4 shows the type of errors in *Mobile* and *Tablet* conditions.

Omissions are the most common error type. Results show that, on average, omission errors are the most common type in both *Mobile* (m=12.65%, sd=16%) and *Tablet* (m=9%, sd=10%) conditions. *Omissions* are often described as cognitive errors, since they do not depend on motor abilities [10]. Instead, users usually forget to type the intended characters or misunderstand the required sentence. Since we did not account for cognitive differences, we cannot confirm this hypothesis. Nevertheless, it would be expected that cognitive errors across device conditions remained unchanged. In fact, no statistical significant differences on *omission error rate* were found between *Mobile* and *Tablet* conditions [Z=-.722, p>.4], suggesting that cognitive demand is constant and may be playing an important role. Further discussion on this topic is available in Section 4.2.6.

Hand tremor can be used to reduce substitutions. *Substitutions* were the second most common error type. Participants obtained a mean 7.8% (sd=7%) error rate in *Mobile* and 3.75% (sd=3.61%) in *Tablet* condition. For both conditions, we found large positive correlations between *substitution* error rate and *task-specific tremor*; that is, participants with higher hand tremor had higher *substitution* error rates. In *Mobile* both *dominant* [Spearman rho=.624, n=15, p<.05] and *non-dominant hand* [Spearman rho=.541, n=9, p<.05] *task-specific tremor* accounted for 39% and 29% of shared variance, respectively. In *Tablet* condition, *dominant hand task-specific tremor* [Spearman rho=.539, n=15, p=.038] explained 29% of shared variance.

Insertions are not predicted by tremor. Overall, *Insertions* were the least common error type (although no significant differences were found) in both conditions. Moreover, we did not find strong correlations with tremor measures, suggesting that there is a weak relationship between *insertion error rate* and *hand tremor*.

Overall, magnitude of errors is lower in Tablet condition, with one exception. We found significant differences between device conditions for *insertion* [Z=-2.103, p<0.05] and *substitution* [Z=-2.731, p<.01] *error rates*. On the other hand, no significant differences were found for *omission errors*, suggesting that these errors do not depend on participants' physical abilities.

4.2.4 Insertion Errors

Insertion errors had two main causes: 1) *accidental* touches, for instance when users were scanning the keyboard for the intended key and accidental touched other key; and 2) *bounce* errors, which occurred when a key was unintentionally pressed more than once, producing unwanted characters. In this section we analyze in detail these two types of errors for both device conditions. Knowing how to identify these errors whilst users type can be of great value to prevent incorrect characters from being entered. Error classification was done through visual inspection of both transcribed sentences and video recordings in order to guarantee a high level of accuracy. Error rates were calculated as *number of errors* / *number of keystrokes*.

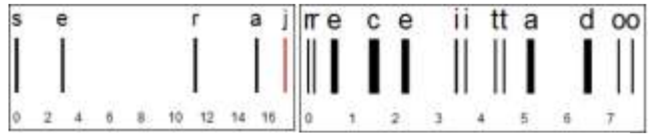

Figure 5. Accidental touch (left) and bouncing errors (right). Time in seconds is represented in *x*-axis.

Accidental touches are less common in Tablet condition. Overall, *bouncing* errors and *accidental* touches account for the majority of insertion errors. Concerning *bounce error rate*, participants obtained a mean of 1.55% (sd=1.7%) in *Mobile* and 2.25% (sd=3.5%) in *Tablet* condition. Regarding *accidental touches*, participants achieved error rates of 3.28% (sd=3.9%) and 1.05% (sd=1.22%), respectively, in *Mobile* and *Tablet* conditions. We found a significant decrease of *accidental* touches in the *Tablet* condition [Z=-2.292, p<0.05]. Conversely, *bouncing* errors were not statistically different between device conditions [Z=-.314, p=0.754], although there was an increase of *bounce* errors with the tablet device.

Mobile bounces and accidental touches are related with hand tremor. We found strong positive correlations between mobile *bounce error rate* and tremor measures: *dominant hand Oscillation on the X axis* [Spearman rho=.596, n=14, p=.025] and *non-dominant hand peak magnitude acceleration between 7 and 12 Hz* [Spearman rho=.532, n=9, p=.14]. Regarding *accidental touches*, large correlations were also found, particularly with non-dominant hand tremor: *Oscillation Y axis* [Spearman rho=.762, n=9, p=.017], *Oscillation Z axis* [Spearman rho=.536, n=9, p=.162], and *peak magnitude acceleration between 3 and 7 Hz* [Spearman rho=.508, n=9, p=.162].

Classifying insertions through key press duration and inter-key interval. From illustrations in Figure 5, both *bouncing errors* and *accidental touches* are easily identified due to reduced press duration and inter-key interval. We believe that a significant percentage of these errors can be automatically classified and filtered by analyzing these typing features. Hand tremor features should also be used to improve filtering solutions. Our data show that individuals are consistent in their input behaviors; however, typing patterns may be both user- and device-dependent.

4.2.5 Substitution Errors

In this section we will analyze common substitution patterns and keyboard layouts that emerged from participants' key presses.

Similar difficulties across all keys. In general, participants had similar difficulties across all keys. No row, column or side patterns emerge from the data for both device conditions.

Right-bottom substitution pattern. To analyze the most common substitutions, we created confusion matrices. Some of the most frequent errors in *Mobile* condition were: C→SPACE (6.83%), C→V (3.17%), O→P (4%), T→Y (3.96%), S→Z (4.34%). As we can see there is a clear predominance of right and bottom key substitutions in the data, which suggests that participants found it easier to hit keys in the right-bottom (southeast) direction. These findings may be related to hand dominance, but further investigation is needed to confirm this hypothesis. Additionally, errors are at a distance of one key. Indeed, this pattern can be seen in Figure 6, which illustrates all lift points of *Mobile* condition. The pattern remains unaltered in the *Tablet* condition, with

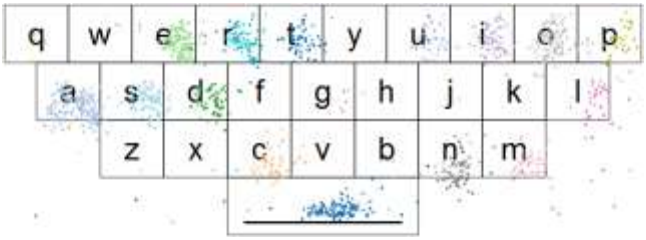

Figure 6. Touch (lift) points for all participants in *Mobile* condition.

common *substitutions* being: C→SPACE (6.8%), R→T (4.2%), S→Z (4.3%), S→D (3%), U→J (1.9%).

Similar and symmetrical letters result in cognitive errors. A common error that cannot be explained by previous substitution pattern is P→Q. We believe this to be a cognitive error, instead of motor error, since it commonly occurred in both conditions. Participants had an improper model of the letter and have confused it with a very similar one (symmetrical: p→q). Indeed, similar problems occurred with the letter 'i', which was frequently replaced by the letter 'l' (3.1%). Participant #11 consistently confused symmetrical letters: m→w (66.7%), n→u (52.85%). These results suggest that some *substitutions* are not due to motor errors alone; cognitive errors play an important role in text input for elderly users.

Most errors are due to poor aiming. In this user study, we were also interested in finding why substitution error occurred; was it due to poor aiming or finger slips? We classified a finger slip as a correct land-on (i.e. land on the correct key) and incorrect lift (i.e. lift on nearby key – substitution). Poor aiming errors consist in landing on and lifting of an incorrect key. Most *substitution* errors were due to incorrect land-on (i.e. *poor aiming*), with an average of 6.71% in *Mobile* and 3.5% in *Tablet* condition. On the other hand, *slips* accounted for an average of 1.1% and 0.24% of typing errors in *Mobile* and *Tablet* conditions, respectively. In fact, slip errors were significantly lower than poor aiming errors in both device conditions: *Mobile* [Z=-3.107, p<.01], *Tablet* [Z=-2.944, p<.01]. Similar results have been reported in [13] for situationally impaired users.

Slip errors are related with hand oscillation, while poor aiming errors are related with task-specific tremor. Different features of hand tremor correlated with *poor aiming* and *slip* errors. While *poor aiming* was strongly correlated with *task-specific tremor* in both device conditions: *Mobile* - [Spearman rho=.541, n=14, p=.046], and *Tablet* - [Spearman rho=.563, n=14, p=.029]; *slip errors* were strongly correlated with non-dominant hand oscillation on XYZ axis only in *Mobile* condition [Spearman rho=.714, n=9, p=.031].

Novel layouts should give more emphasis to key width. Last, we were interested in the overall virtual keyboard layout that would emerge from elderly users touch points. For this analysis, we calculated key centroids for each key across all participants. We removed outlying points that were more than one key distance away from the center of each key in either x or y direction, to account for transposition or cognitive errors. Additionally, we calculated the standard deviation of finger-lift points for each key in x and y directions. We then grouped the 26 keys by row and side. Right keyboard side contained the keys P, O, I, U, Y, L, K, J, H, M, N, B, and the left side contained the remaining letters. The keyboards that emerged from this analysis were shifted to the bottom-right in comparison to the traditional QWERTY keyboard, which was expected from previous findings. Also, we found no significant effect of *row* on deviations for both x- and y-directions in either *Mobile* or *Tablet* conditions; that is precision is equal across all rows. However, we found a significant decrease of x-direction deviations from left to right side of *Mobile* keyboard [t(13)=-3.043, p<.01]. This result suggests that keys on the left side of the *Mobile* keyboards should be slightly wider, when possible. In the *Tablet* condition, we also found a statistically significant increase of x-axis *dispersion* relatively to y-direction [t(14)=4.039, p<.001]. Again, these findings may be related with hand dominance and demonstrate that elderly users are more susceptible to x-direction deviations from their touch centroids.

4.2.6 Omission Errors

Omissions were the most common error type in this user study and are usually associated to cognitive errors. Understanding omission errors is particularly difficult since it is hard to understand the reason why participants failed to enter the intended character/word. Was it because of they forgot it or because the device was unable to recognize the users' touch? In order to answer these questions we resorted to video recordings.

Blank space was the most problematic key. During our user study with elderly participants, we found that forgetting to enter a blank space between words was a common issue. Although participants were instructed before the evaluation session, the concept of a blank character was sometimes difficult to understand. In fact, this key achieved the highest error rate across all keys (25-30%) and *omissions* were the main cause.

Forgetfulness and coordination are real issues. Some participants forgot to transcribe some letters or words during text-entry tasks. For instance, participant #13 usually forgot to transcribe words at the middle and end of sentences. Still, her performance was consistent across device conditions, which suggest that this was a cognitive error. For participant #8, the copy task seemed to be overwhelming as she could not manage and coordinate what she has transcribed and what was yet to be transcribed. She frequently asked things like: "where was I?", "have a written this?", thus resulting in omitted letters and words. These results illustrate some of the challenges in evaluating text-entry performance with elderly people.

Unintentional touches prevented key presses. While *omission* errors may be related with cognitive errors, there were also some issues (although less severe) regarding touch interaction. Particularly, unintentional touches occurred when participants were holding or resting their non-dominant hand on the device. These behaviors resulted in unrecognized key presses since the keyboard only handled a single input point.

4.3 Participants' Comments and Preference

At the end of the user study participants were debriefed and asked about their preferred device. Additionally, we also gathered general comments about their input performance.

When asked about each device ease of use (using a 5-point Likert scale), the median [IQR (Interquartile Range)] attributed by participants was 4.5 [1.75] and 5 [0.5] for *Mobile* and *Tablet* devices, respectively, showing a preference for the tablet device. Participants' classifications were generally high, which may be misleading when considering their difficulties. Still, when directly asked about their preferred device results are clear: thirteen (86.7%) participants chose the *Tablet*: with a 95% adjusted-Wald binomial confidence interval ranging from 60.9% to 97.5%, a lower limited well above the two-choice change expectation of 50%. The main reasons for their choice were the larger key size and spacing. Moreover, some participants also stated that letters (i.e. visual feedback) were easier to see. When asked about their main difficulties, participants referred diverse issues about: 1) key acquisition, particularly in the *Mobile* condition (*"I am always hitting neighbor keys"; "The hardest thing is trying not to tremble while texting"*), and 2) keyboard layout, mainly for those with no QWERTY experience (*"The main difficult for me is in knowing where the letters are. I am not used to it"*).

5. TOWARDS INCLUSIVE KEYBOARDS

The analysis presented above provide insight about elderly users' typing patterns and how keyboard features may be improved to better support text input on touch-based devices. In this section, we access the reliability of simple touch models and perform a user-dependent and user-independent analysis. The goal is to demonstrate the potential of such solutions, acknowledging that more efficient models can be found resorting to more sophisticated measures (e.g. tremor features) and algorithms.

5.1 Deal with Insertions

To deal with *insertion* errors, we calculated the optimal inter-key threshold; that is, the value that allowed reducing *insertions* without negatively affecting *MSD error rate*. In this analysis we used 40 values: the number of 25ms intervals from 0ms to 1000ms. Key presses that had an inter-key interval lower than the threshold being tested were considered insertions and were therefore discarded. We then computed *MSD error rate* from all resulting sentences and compared it against the baseline condition.

For the user-dependent analysis, we calculated the optimal inter-key threshold for each participant, based on their typing behavior. Results show that *MSD error rate* dropped, on average, 6.8% in *Mobile*, and 1.8% in the *Tablet* condition. Optimal threshold values varied from 25ms to 1000ms in the *Mobile* condition and from 50ms to 675ms in the *Tablet* condition. For the user-independent analysis, we calculated the mean i*nsertion* and *MSD error rate* of all participants and choose the inter-key threshold that would allow a higher performance gain (on average). In the *Mobile* condition the threshold was 100 ms and resulted in a reduction of 0.8% of *MSD error rate*. In the *Tablet* condition, the threshold was slight higher, 150ms and resulted in a decrease of 1.1% of *MSD error rate*. It is noteworthy the decrease in performance of the user-independent classifier, especially in the *Mobile* condition. This result suggests that filtering solutions should take into account each user typing behaviors. Moreover, this simple approach removed nearly 30% and 50% of *insertion* errors in the *Mobile* and *Tablet* conditions, respectively.

5.2 Deal with Substitutions

To deal with *substitution* errors, we performed a simple key classification based on the Euclidean distance between two points. Key centroids were calculated for each key and all key presses were re-classified according to the closest centroid. Regarding the user-dependent classification, we used a 10-fold cross-validation to calculate the mean centroid of each key for each training subset of data, and classified the remaining key presses. *MSD error rate* dropped, on average, 11.5%, and 1.2% in the *Mobile* and *Tablet* conditions, respectively. Overall, participants were consistent within themselves, repeatedly hitting the same places for the same keys. For the user-independent classification, we calculated the average of all key centroids for all participants and classified each participant's key presses based on the closest centroid. This approach also reduced *MSD error rates*, on average, by 9.8% for the *Mobile* condition, and 0.6% for the *Tablet* condition. Results show that the *Mobile* gain is higher in both classification approaches. However, the user-independent classification performed worst, suggesting that personalization should be taken into account when designing touch-based solutions for the elderly.

6. DISCUSSION

After analyzing all data, we are now able to answer the research questions proposed at the beginning of this user study.

1. How do elderly users perform speed and accuracy wise in touch-based devices? Elderly users achieved a maximum of 11.5 WPM using the tablet device (mean of 4.7 and 5 WPM for *Mobile* and *Tablet* conditions, respectively). Also, input speed was not correlated with tremor, instead it was strongly correlated with previous QWERTY experience. On the other hand, accuracy was mainly explained by *task-specific tremor* and *hand oscillation*, especially in *Mobile* conditions. Users obtained a minimum *MSD error rate* of 2.5% (mean of 26% and 17% for *Mobile* and *Tablet* conditions, respectively). Curiously, *Error Rate* was not correlated with previous QWERTY experience, suggesting that having some practice with keyboards is not sufficient to compensate the challenges that are imposed by touch interfaces.

2. What are the most common types of errors and causes? The most common error type among elderly people was *omission* errors (9-12.6%). This pattern occurred across device conditions, suggesting that it was due to cognitive errors. Nonetheless, the novelty of the task can also be playing an important role, thus it would be interesting to observe users' performance on a longitudinal study. Following *omission* errors were *substitution* (3.75-7.8%) and *insertion* (3.8-5.5%) errors. *Insertions* were mainly due to bounces and accidental touches; while *substitutions* were mostly due to poor aiming.

3. Do tablet devices compensate the difficulties of elderly users when using mobile phones? Overall, we found a decrease of 9% in MSD *error rate* from *Mobile* to *Tablet* devices. This finding suggests that tablet devices compensate some of the challenges imposed by mobile devices, either due to larger key sizes and/or static positioning. Indeed, users' comments and preference reinforced this result. Regarding types of error, there was a significant decrease of both *insertions* (1.7%) and *substitutions* (4%). No significant differences were found on *omission* errors, suggesting that they are device-independent.

4. Does tremor affect text-entry performance? If yes, how does user performance correlate with hand tremor? Although input speed was mainly related with QWERTY experience, errors were strongly correlated with participants' tremor profile. However, each error type was correlated with different measures of tremor. *Substitutions* were largely explained by a subjective measure - *task-specific tremor*, while *insertion* errors, particularly bounces and accidental touches were strongly correlated with O*scillation in the X axis* (dominant hand). The non-dominant hand also played an important role in *Mobile* errors: *Hand Oscillation* was strongly correlated with overall *MSD error rate*, *accidental touches*, and *slips*. These findings suggest that future mobile interfaces should take into account users' tremor profile in order to provide more suitable text-entry designs. Still, designers should consider different features of tremor.

7. IMPLICATIONS FOR DESIGN

We derive the following implications from our results:

Shift keyboard layout. Elderly participants theoretically benefit from a layout shift in the bottom-right direction as most substitution errors occur in this direction. This finding may be related to hand dominance, thus further research should explore this hypothesis. Future work should also explore whether this change should be visible to the user, similarly to [5].

Width rather than height. Whenever possible keys should be wider instead of taller. For both devices we found higher x-axis touch dispersion, suggesting that users are more favorable to wider keys. In fact, even though most 12-key physical keyboards respect this layout, it was lost in touch interfaces.

Narrower spacebar. Results of touch deviations suggest that spacebar should be narrower. Reducing its size has the potential to diminish substitution errors. We recommend a spacebar extending from middle of C to middle of B for both devices.

Avoid errors by understanding typing behaviors. Future designs should focus on model users typing patterns by analyzing touch features (e.g. x and y touch position, distance traveled during touch, key press duration, between keys duration etc.) and therefore increase typing accuracy.

Allow personalization. We observed several individual differences regarding typing behaviors, particularly hit point locations, and inter-key interval. Future research should tackle these issues by providing user-dependent solutions.

Deal with poor aiming rather than finger slips. Keyboard designers should deal with poor aiming errors. Although finger slips may occur they only account for a minority of substitution errors, particularly when typing on tablet devices.

Use language-based correctors. Cognitive errors were quite common among elderly users. Simple language-based solutions can provide a suitable answer to these types of errors. For example, to deal with blank space omissions or substitution of similar letters (e.g. p→q, m→w).

Compensate hand tremor. Future keyboards should adapt to users' hand tremor characteristics. Results showed large correlations between tremor measures and input accuracy, namely when considering substitution errors. Taking advantage of current mobile sensing capabilities, future solutions should trace users' tremor profile to compensate typing errors.

8. CONCLUSION

We have investigated text-entry performance of 15 elderly users on touch-based devices. Results showed that error rates are still relatively high compared to younger users' performance [13]. Hand tremor was strongly correlated with input errors, indicating that this information can be used to enhance text-entry accuracy. Most common types of error were *omissions* (10.8%), followed by *substitutions* (5.8%), and *insertions* (4.6%). From results emerged error patterns and design implications that should improve typing accuracy and persuade researchers to create more effective solutions for the elderly. Future work should improve proposed solutions and focus in coping with each user's abilities, enabling them to effectively input text on touchscreen devices.

9. ACKNOWLEDGMENTS

We thank all participants and people from *Centro de Dia e Lar de Algueirão Mem Martins*, particularly Luisa Ambrósio and Ana Rodrigues for their insights. This work was supported by FCT: individual grant SFRH/BD/46748/2008; project PEst-OE/EEI/LA0021/201; and project PAELife AAL/0014/2009.

10. REFERENCES

[1] Bain et al. Assessing the impact of essential tremor on upper limb function. *Journal of neurology*, 241(1):54–61, 1993.

[2] Bergstrom-Lehtovirta et al. The effects of walking speed on target acquisition on a touchscreen interface. *In Proc. of MHCI'11*, 143–146, 2011.

[3] Chung et al. Usability evaluation of numeric entry tasks on keypad type and age. *International Journal of Industrial Ergonomics*, 40(1):97–105, 2010.

[4] Elble and Koller. *Tremor*. Johns Hopkins University Press Baltimore, 1990.

[5] Findlater and Wobbrock. Personalized input: Improving ten-finger touchscreen typing through automatic adaptation. *In Proc. of CHI'12*, 815-824, 2012.

[6] Froehlich et al. Barrier pointing: using physical edges to assist target acquisition on mobile device touch screens. *In Proc. of ASSETS'07*, 19–26, 2007.

[7] Goetz et al. Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): Scale presentation and clinimetric testing results. *Movement disorders*, 23(15):2129–2170, 2008.

[8] Gunawardana et al. Usability guided key-target resizing for soft keyboards. In *Proc. of IUI'10*, 111–118, 2010.

[9] Jin et al. Touch screen user interfaces for older adults: button size and spacing. *Universal Acess in Human Computer Interaction. Coping with Diversity*, 933–941, 2007.

[10] Kristensson. Five challenges for intelligent text entry methods. *AI Magazine*, 30(4):85, 2009.

[11] Kurniawan. Older people and mobile phones: A multi-method investigation. *International Journal of Human-Computer Studies*, 66(12):889–901, 2008.

[12] MacKenzie and Soukoreff. Text entry for mobile computing: Models and methods, theory and practice. *Human–Computer Interaction*, 17(2):147–198, 2002.

[13] Nicolau and Jorge. Touch typing using thumbs: understanding the effect of mobility and hand posture. *In Proc. of CHI '12*, 2683–2686, 2012.

[14] Salarian et al. Quantification of tremor and bradykinesia in parkinson's disease using a novel ambulatory monitoring system. *Biomedical Engineering, IEEE Transactions on*, 54(2):313–322, 2007.

[15] Selker et al. Psychosocial indicators via hand tremor. *In Proc. of INTERACT'11*, 596–599, 2011.

[16] Strickland and Bertoni. Parkinson's prevalence estimated by a state registry. *Movement disorders*, 19(3):318–323, 2004.

[17] Wacharamanotham et al. Evaluating swabbing: a touchscreen input method for elderly users with tremor. *In Proc. of CHI'11*, 623–626, 2011.

[18] Wobbrock et al. EdgeWrite: a stylus-based text entry method designed for high accuracy and stability of motion. *In Proc. of UIST'03*, 70, 2003.