UNIVERSIDADE TÉCNICA DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

INSTITUTO SUPERIOR TÉCNICO

Disabled 'R' All: Bridging the Gap between Health- and Situationally-Induced Impairments and Disabilities

HUGO MIGUEL ALEIXO NICOLAU

Supervisor: Doctor Joaquim Armando Pires Jorge

Thesis approved in public session to obtain the PhD Degree in

Information Systems and Computer Engineering

JURY FINAL CLASSIFICATION: **PASS WITH MERIT**

Jury

Chairperson: Chairman of the IST Scientific Board

Members of the Committee:

Doctor Vicki Lynne Hanson

Doctor Joaquim Armando Pires Jorge

Doctor Luís Manuel Pinto da Rocha Afonso Carriço

Doctor João Manuel Brisson Lopes

2013

UNIVERSIDADE TÉCNICA DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

INSTITUTO
SUPERIOR
TÉCNICO

Disabled 'R' All: Bridging the Gap between Health- and Situationally-Induced Impairments and Disabilities

HUGO MIGUEL ALEIXO NICOLAU

Supervisor: Doctor Joaquim Armando Pires Jorge

Thesis approved in public session to obtain the PhD Degree in

Information Systems and Computer Engineering

JURY FINAL CLASSIFICATION: **PASS WITH MERIT**

Jury

Chairperson: Chairman of the IST Scientific Board

Members of the Committee:

Doctor Vicki Lynne Hanson, Professora Catedrática, School of Computing, University of Dundee, Escócia

Doctor Joaquim Armando Pires Jorge, Professor Catedrático, Instituto Superior Técnico, Universidade Técnica de Lisboa

Doctor Luís Manuel Pinto da Rocha Afonso Carriço, Professor Associado, Faculdade de Ciências, Universidade de Lisboa

Doctor João Manuel Brisson Lopes, Professor Auxiliar, Instituto Superior Técnico, Universidade Técnica de Lisboa

Funding Institutions

Fundação para a Ciência e a Tecnologia (SFRH/BD/46748/2008)

2013

Abstract

Mobile devices are being increasingly used in many different situations, namely whilst walking. These new interaction contexts are blurring the frontiers between able-bodied and disabled users, as new challenges arise. Mobile users often face demands that compete for the same human resources needed to operate electronic devices, leading to situationally-induced impairments and disabilities (SIID). For instance, when on the move the whole body is prone to vibrations, inducing hand tremor that hampers interaction with mobile devices. This effect worsens when the interface features a touchscreen and both small targets and narrow spacing, such as virtual keyboards. In fact, these mobile users seem to share some problems with older adults, who experience increased physiological hand tremor, a health-induced impairment and disability (HIID).

We propose to bridge the gap between SIID and HIID, showing how designers can build mobile solutions that address a wide range of abilities. We thoroughly investigated each user group performance on text-entry tasks in order to exploit their main similarities and deal with their differences. We found that the increased demands of mobility seemingly induce a disability continuum, where the performance of both user groups was interleaved. Also, results showed overlapping problems, which strongly correlate with a unifying measure: hand motion.

We designed and evaluated multiple error-compensation models to improve typing accuracy of users with either SIID or HIID. Particularly, several motion-enhanced models that took advantage of devices' built-in accelerometers were developed. These solutions show significant improvements over models that only use touch information. We propose a Unifying Model that leverages accelerometer data and deals with two kinds of typing errors: insertions - added characters, and substitutions - incorrect characters. Results show that the model improves typing accuracy of users with either situationally- or health-induced impairments and disabilities. This confirms that indeed mobile interfaces can be designed to address a wide range of abilities, independently of their cause, thus bridging the gap between "mainstream" solutions and accessible computing.

The thesis proposed in this dissertation is thus:

> *We can improve typing performance of both health- and situational-impaired users by leveraging motion data on touch-based mobile devices.*

Resumo

Os dispositivos móveis são cada vez mais usados em múltiplas situações, nomeadamente em deslocamento (andar). Estes novos contextos de interacção têm vindo a esbater as diferenças entre utilizadores com e sem deficiências. Em ambientes móveis, os recursos humanos outrora somente usados para controlar o dispositivo são necessários para outras tarefas, originando deficiências situacionais (*Situationally-Induced Impairments and Disabilities* - SIID). Por exemplo, durante o andar, todo o corpo está em movimento, induzindo tremor, que por sua vez afecta a interacção com aparelhos móveis. Este efeito é ainda mais pronunciado aquando da utilização de uma interface táctil, com alvos e espaçamento de pequenas dimensões (ex. teclado virtual). De facto, estes utilizadores parecem partilhar alguns dos problemas encontrados noutras populações que devido a razões de saúde (*Health-Induced Impairments and Disabilities* - HIID) experienciam um aumento de tremor nas mãos. É o caso de utilizadores idosos que têm também dificuldades em selecionar alvos de pequenas dimensões.

Nesta dissertação propomo-nos a encontrar uma relação entre SIID e HIID e mostrar como construir soluções que abranjam um largo espectro de capacidades. Para atingir este objectivo, investigámos cada um destes domínios em separado, recorrendo a tarefas de entrada de texto. Esta análise permitiu-nos identificar as principais diferenças e semelhanças entre os dois tipos de deficiências. Entre outros resultados, a elevada exigência das condições de mobilidade introduziu um fenómeno de continuidade ("*disability continuum*"), onde o desempenho de utilizadores idosos se intercala com os restantes. Mais, foi encontrada uma sobreposição de problemas entre domínios, fortemente relacionada com uma medida unificadora: o tremor.

Em seguida, foram criados e avaliados vários modelos de compensação de erros, com o objectivo de melhorar o desempenho de utilizadores com SIID ou HIID em tarefas de entrada de texto. Em particular, desenvolvemos diversos classificadores que foram aumentados com informação de movimento (aceleração). Estes mostraram serem mais eficazes em relação a classificafores que continham somente informação acerca do toque. Por fim, é proposto um Modelo Unificador que tira partido da informação lida pelo acelerómetro tridimensional do dispositivo e lida com dois tipos de erros: inserções - caracteres involuntariamente adicionados; e substituições - caracteres errados. Os resultados mostram que o modelo é capaz de melhorar significativamente a qualidade final das frases transcritas pelos utilizadores com SIID ou HIID. Isto confirma que, de facto, interfaces móveis podem ser criadas para

um espectro alargado de capacidades, independentemente dos factores que as afectam, aproximando soluções *mainstream* daquelas criadas no domínio da acessibilidade.

A hipótese de investigação desta dissertação é:

> *Podemos melhorar o desempenho, em tarefas de entrada de texto, de utilizadores com deficiências situacionais ou por razões de saúde, aproveitando informação sobre o movimento do dispositivo.*

Keywords

Tremor, SIID, HIID, Older Adult, Walking, Abilities, Mobile Device, Touchscreen, Text-Entry, Motion

Palavras-Chave

Tremor, SIID, HIID, Idoso, Andar, Capacidades, Dispositivo Móvel, Ecrã Táctil, Entrada de Texto, Movimento

Acknowledgements

I would like to thank Professor Joaquim Jorge, for his faithful support, as an advisor, during many years. His immeasurable enthusiasm and insight have been a source of motivation since I was a master's student fresh to the day I finished my PhD. I truly appreciate the opportunity and autonomy that was given to me, so I could do the work I really love.

I would like to thank the other members of my thesis committee, Luís Carriço and João Brisson Lopes, for their time and energy in improving this work. Your valuable comments and hard questions helped me shape the direction of this work. I would also like to thank Vicki Hanson for promptly accepting the invitation to be my international contestant. I could not ask for a more inspiring researcher. Her efforts in the accessible computing field are invaluable, and hopefully some of that is reflected in my dissertation.

A number of students, faculty and others from the INESC-ID's Visualization and Intelligent Multimodal Interfaces group, have influenced this work as well: Ricardo Jota, Bruno Araújo, Tiago Guerreiro, João Guerreiro, Gabriel Barata, Ricardo Dias, Sandra Gama, Daniel Gonçalves, Manuel J. Fonseca, Alfredo Ferreira, Vasco Gervásio, Pedro Lopes, João Oliveira, David Lucas, João Martins, among others. Thank you all.

I would like to give special thanks to my fellow researchers, Bruno Araújo and João Guerreiro, for their availability (like you had any choice!) in listening, discussing, and brainstorming ideas. For their time in proofreading this document, thank you Abi Roper and Pedro Lopes, I owe you a couple of beers. Lastly, I thank Tiago Guerreiro, who shared an office with me and was always present through all my (and still short) research career. It was a very interesting partnership and we surely complemented each other. Hopefully, we will be able to continue collaborating in this new stage of our lives. More than a research colleague, I now have the pleasure of calling you a friend.

I would not have finished this work without the support and collaboration from several institutions. I thank Fundação Raquel e Martin Sain, Lar Militar da Cruz Vermelha Portuguesa, and Centro de dia Algueirão Mem-Martins. All these institutions, and the professionals that represent them, always showed great enthusiasm toward my ideas. Thank you for opening your doors and allow me to put my work to the test. I would like to highlight two persons, Luisa Ambrósio and Ana Rodrigues, for their time and availability in recruiting participants. My special thanks go to both of you. Along these lines, I also thank all participants for their availability without asking anything in return.

I would like to thank the institutions that allowed me to conclude my work. My host institution, INESC-ID Lisboa, where I found the perfect conditions to do my research; and FTC, for the 4-year funding with the individual grant SFRH/BD/46748/2008.

Less tangible, but no less important has been the support from all my friends. Especial thanks go to my lifelong friends; brothers from other mothers, David Nunes, Tiago do Paço, Bruno do Paço, and João Marques. Always there when needed, with a cold drink in their hands and a smile on their faces, just to cheer me up.

Most importantly, I would like to thank my family. People say families are the friends we did not get to choose. Well, I could not have chosen any better. To my grandmother, uncles, and aunts thank you for the unconditional support and lifelong lessons. To the love of my life, Matilde, who shared all the highs and lows throughout this process, I thank her for making me a happy man. I could not have done it without you. I would also like to thank my in-laws for all the great moments, most of them spent at the table, eating a hot tasty meal, accompanied by the most amazing wines.

Finally, I dedicate this thesis to my parents, Carlos and Amélia. I owe you what I am today and will be in the future. Without your encouragement, education, values, and principles, I would not be where I am now. Thank you for all your sacrifices and devotion.

To my loving parents,

Carlos and Amélia.

Contents

List of Figures

List of Tables

Statement of Collaboration

This thesis was conducted under the supervision of Professor Joaquim Jorge. I am the primary contributor to all aspects of this research, with the exception of the Investigation of Motor Demands in Chapter 3. In this user study, Tiago Guerreiro and I were equal contributors to the conception, research and data analysis. Daniel Gonçalves also played an important role on the conception stage. Master student João Martins developed the evaluation application and ran the user tests.

David Lucas contributed on the study design of Investigation of Visual Demands in Chapter 3 and ran the respective evaluations under my close guidance. João Oliveira contributed to the gathering of the evaluation sentences used in text-entry evaluations.

1 Introduction

Over the last two decades mobile devices have become ubiquous in our society. The rapid uptake of mobile technology is an incredible phenomenon that has forever changed our lives. Indeed, mobile subscribers have increased in number, 15% every two years, past the 5 billion mark, world-wide (Budde, 2010). In developed countries, mobile phones per capita already exceed one[1]. These numbers reflect how important mobile phones have become in our daily routines. Also, they rapidly gained new functions and applications, and became smaller, cheaper, and increasingly powerful. Indeed, mobile phones no longer stand for mere communication tools but span onto newer domains, such as: work, entertainment, shopping, communication, and transportation.

Furthermore, these devices are used in wide demographics, independently of social or economical status, age, preferences, values, or culture (Budde, 2009). Mobile devices are used worldwide in developed and third-world countries; by people with completely different physical and cognitive abilities, such as adolescents and older adults; they serve different purposes and goals, depending on its user and situation. The diversity of their target audience is enormous, and each individual has a very different set of requirements. However, current interfaces do not address these needs well.

Nonetheless, mobile devices spend more time near us than any other electronic device: at home, at work, in the car, while shopping, in public transports, whilst walking, etc. The variety of interaction contexts is enormous (Weiser, 1993). In contrast to the traditional desktop environment, there is a clear paradigm shift (Lumsden and Brewster, 2003). The

[1] http://data.worldbank.org

interaction context evolved from static, stationary, and controlled environments to highly dynamic and heterogeneous settings.

Users are often in motion whilst using their devices, which can result in the overload of their abilities (both physical and cognitive) and even put at risk their physical integrity (Richtel, 2010). These issues are collectively called Situationally-Induced Impairments and Disabilities (SIID) (Sears and Young, 2003). Namely, motor abilities are often challenged in SIID; for instance, whilst walking or in a bus, the vibration originated from changes in speed has a negative effect in our performance with mobile devices (Mizobuchi et al., 2005)(Lin et al., 2007)(Yatani and Truong, 2009)(Nicolau and Jorge, 2012c). The same effect can be observed in Health-Induced Impairments and Disabilities (HIID), where users with increased physiological tremor show irregular movements and lower accuracy (Nicolau and Jorge, 2012a). These challenges are particularly visible in current touchscreen interfaces.

1.1. The need for Accessible Touch Interfaces

There was a time where touchscreen technology was only affordable to a selected few. Nowadays, this technology has spread to many devices, applications and environments, such as ATM machines, information kiosks, ticket machines, health control devices, and so forth. Most of us use touchscreens on a daily basis due to its enormous success in mobile devices. In fact, since the first iPhone[2] was released in 2007, touchscreens have been increasingly replacing traditional mobile phone keypads (Figure 1.1).

Touch devices offer many advantages over their button-based counterparts. Particularly, they can display different interfaces in the same surface or adapt to the users' needs and preferences (Findlater and Wobbrock, 2012). Moreover, the ability to directly touch and manipulate data on the screen without any intermediary devices has a strong appeal, since it provides for a more natural, engaging, personal and even intimate experience (Powers and Brown, 2012).

Nevertheless, for all their advantages, touch interfaces present numerous challenges. They lack both the tactile feedback and physical stability guaranteed by keypads, making it harder for people to accurately select targets (Brewster et al., 2007). This becomes especially relevant to those experiencing hand tremor, be it either situationally- or health-induced. Touch interfaces can become similarly inaccessible to either user groups. For instance, whether inside a moving train or being an older adult, accurately typing on a virtual keyboard can prove to be an overwhelming task, due to increased hand oscillations.

Given the importance of mobile devices on our lives, the inability to fully control and

[2] http://www.apple.com/iphone/

Figure 1.1: Mobile devices evolution, from keypad-based to touchscreens.

operate these tools is likely to be a strong factor of exclusion. For users who are not able to compensate their disabilities, this effect defeats the original purpose of mobile devices: improving communication among people. Indeed, failing to design effective and accessible interfaces is likely to widen the "technology gap" between able-bodied and disabled users, excluding the later from many opportunities.

Still, touch devices have the potential to reduce this gap due to their high customizability, which makes them appropriate to custom tailored and adaptive solutions that can fit the needs of different users.

Why Text-Entry?

Most of our research focuses on text-entry tasks, particularly with QWERTY virtual keyboards. As described earlier, touch interfaces make target selection a much harder task in comparison to keypad devices, since there are not any physical buttons on the screen. This effect becomes even worse for interfaces that feature small targets and spacing (Mizobuchi et al., 2005)(Jin et al., 2007), such as soft keyboards.

Indeed, mobile text-input is a major challenge for both older adults and situationally-impaired users. Since this is such a transversal task to many applications, e.g. basic communications, managing contacts, editing documents, note taking, web browsing, tweeting, searching, and so forth, these users are excluded from several opportunities in a wide range of domains, ranging from social, professional, leisure, shopping, healthcare, to communication. In light of these challenges, we decided to focus our research on understanding and developing effective solutions to improve typing performance.

Hand Tremor as a Unifying Problem

Situationally-Induced Impairments and Disabilities can result from either the work environment or the activities the user is engaged in (Sears et al., 2003). These contextual factors may negatively affect users' performance with their mobile devices. For instance, depending on input technologies available, hands-busy situations may result in a physical impairment and device-oriented disability. The absence of a limb can result in similar disabilities. In the same way, low temperatures or inflamed fingers due to arthritis result in disabilities associated with touch-based interactions. Moreover, users exposed to sunlight's glare and blind people are both unable to see the screen content. Indeed, we believe that a variety of health conditions and contextual factors may result in similar challenges when using mobile devices; still, this is an open question (Wobbrock, 2006).

While we deliberately draw an analogy between health- and situationally-induced impairments and disabilities, we also explicitly acknowledge that these present some differences. SIID tend to be temporary and dynamic, which makes it more difficult to develop compensation strategies. On the other hand, health-related disabilities are usually permanent, allowing users to develop more elaborate mechanisms to overcome their disabilities. Still, there seem to be overlapping challenges posed by these two groups.

In this dissertation, we intentionally focus on tremor-induced impairments and disabilities. For example, whilst walking the whole body is prone to vibrations; hand oscillations can hamper typing performance on a virtual keyboard. At the same time, older adults (age 65+) who experience increased physiological tremor (Raethjen et al., 2000) may experience similar accuracy difficulties. We approach both these problems as one; whether due to mobile context or health-related issues, users in both groups experience a similar cause of impairment: hand tremor. In the context of this dissertation, we define tremor as any involuntary, approximately rhythmic, and roughly sinusoidal motion around a joint (Raethjen et al., 2000).

1.2. Goals and Hypothesis

The primary research goal of this work was to show that hand tremor, particularly motion data, can be used to enhance typing performance for users who experience either situationally- or health-induced tremor.

Thus, the thesis proposed in this dissertation is:

> *We can improve typing performance of both health- and situational-impaired users by leveraging motion data on touch-based mobile devices.*

In our early research, we focused on understanding the effect of both motor and vision abilities when using touch interfaces. Although not explicitly targeted at tremor-induced impairments, the initial user studies allowed us to refine and adjust our expectations. Indeed our first user study assessed motor-impaired users' (i.e. tetraplegic) performance within a set of touch-based interaction techniques. Our goal was to *identify the main differences and similarities between motor-impaired and able-bodied users when interacting with a mobile touch interface.* Results were very promising as interesting similarities emerged; however, we found that while tetraplegic users experience some form of tremor, they also have severe restrictions of movement and reach. Also, the magnitude of errors between the two user groups was significantly different. Thus, we began exploring mobility conditions in order to reduce this gap.

Our second user study focused on mobile usage whilst walking on visually demanding environments. Since graphical information is inadequate in these contexts, we could *apply assistive technologies that were especially designed for eyes-free interaction in mobile settings.* However, contrary to our expectations, assistive technologies for blind people proved to be ineffective for sighted users whilst mobile, at least when visual feedback was also available. Nevertheless, issues related to target selection accuracy emerged.

From this point on, we explored tremor-related impairments towards developing more inclusive solutions. Still, we had to observe how tremor-related impairments emerged whilst walking and their effect on user performance. Moreover, we needed to better understand the differences between situationally- and health-induced tremor. Thus, we investigated physiological tremor on older adults as well.

In order to characterize users' performance whilst mobile, we conducted a user study to *assess the effect of walking conditions, particularly hand tremor, on typing performance.* Also, we investigated whether different hand postures were able to compensate the challenges of mobility. Only by collecting data on users' performance could we improve current solutions.

Similarly, we needed to *understand how older adults input text to touch-based devices and its relationship to their tremor profiles.* Thus, we gathered data from older adults performing text-entry tasks using two devices.

The question that arose at this point was: *to what extend is there an overlap of problems between SIID and HIID user groups.* We performed a comparative analysis of input performance and tremor features to answer this question. Indeed, similarities emerged from the data, resulting in the "Disability Continuum" concept, where both situationally- and health-impaired users' performance interleaved. Also, we identified differences between groups that would need to be addressed in future solutions.

Nevertheless, we obtained very promising results, which encouraged us to design and evaluate models to leverage motion data and enhance typing performance. The goal was to *validate our approach and propose solutions for both situationally- and health-impaired users.*

The culmination of our research work consisted in *verifying, through a set of simulations, whether the newly proposed models improved text-entry performance*. Indeed, they did. We expect that the knowledge derived from this dissertation can persuade designers to improve mobile solutions by using inclusive and unifying approaches that can benefit a large number of users.

1.3. Research Approach and Overview

We took a user-centered approach to meet the goals of this thesis, starting with exploratory user studies, followed by two in-depth controlled laboratory studies. Using the data thus collected we designed, developed, and evaluated different models to improve user performance.

Our work draws inspiration from diverse approaches, including Universal Usability (Vanderheiden, 1998)(Shneiderman, 2000), Design for all (Stephanidis et al., 1998), User Interfaces for All (Stephanidis, 2000), Inclusive Design (Newell and Gregor, 2000), Capability-Demand (Persad et al., 2007), and Ability Design (Wobbrock et al., 2011). These considerably overlap each other towards the common goal of providing or enhancing access to technology. However, where and how each approach directs a researcher's focus, may vary.

Similarly to Inclusive Design (Newell and Gregor, 2000) and Ability Design (Wobbrock et al., 2011), we acknowledge that each user has a set of abilities. Interaction arises out of the relationship between technology, context, and users' abilities (Dey, 2001)(Sears et al., 2003). According to this conceptual framework, users' abilities may be affected by contextual factors (SIID) in ways similar to health-related impairments (Newell, 1995).

To demonstrate this effect, we faced major challenges in capturing, quantifying and comparing users' abilities. Others have approached these hurdles through clinical functional assessment (Oliveira et al., 2011a) and formal disability characterization (WHO, 2009); however, these methods require specific professional expertise (e.g. clinician), and yield a too high level characterization that lacks useful information to interface design. Our approach focuses on capturing user abilities by directly monitoring their performance with technology. This produces objective measures of speed and accuracy that can be used to adapt current interfaces either at design- or run-time. Types of errors are particularly important to model user abilities, since these offer information about the most relevant issues and challenges found when using a particular solution.

Our approach addresses a wide range of abilities, either affected by situationally- or health-induced impairments, in a unified manner, bridging the gap between user groups that are typically seen as distinct. This unified view of users' abilities offers several advantages to the research community. First, it *avoids the duplication of work* by raising awareness on overlapping issues. Second, in addition to avoid doing work already created by others, it *promotes reusing knowledge* between accessibility, mobile research and general HCI (Human-Computer Interaction) areas. Since there are common challenges, similar solutions can be applied to both groups. Third, it *leverages research* by allowing a larger number of researchers, with distinct backgrounds, to create broader and multidisciplinary solutions. Fourth, it *reduces costs and increases availability* by bringing accessibility to mainstream products. Last but not least, it *removes the negative connotations of the words accessibility and accessible computing*, which are typically seen as a research area only for those with very specific motor or cognitive disabilities (i.e. minorities).

1.3.1. Research Overview

In order to achieve our goals and understand users' abilities, the work conducted in this thesis comprises distinct user studies. We first describe two preliminary studies that allowed us to focus and align our research (Chapter 3). These were design to explore the effect of both motor and visual abilities on user performance with touch-based interaction techniques. Its results showed similarities between motor-impaired and able-bodied users regarding the most effective techniques. Furthermore, when we explored typing performance on mobility settings, we found several issues regarding target acquisition, which led us to focus our attention on tremor-induced impairments.

From this point on, we aimed at understanding tremor impairments, either due to situational or health factors. For this end, we conducted two in-depth user studies and hereby report a detailed analysis of the users' input performance with special regards to hand tremor. These results had the potential to produce data-driven solutions to compensate some of the challenges of both target population.

Regarding SIID (Chapter 4), our main goal was to analyze first, the negative effect of walking on text-input performance, particularly the users' main difficulties and error patterns; and second, how these effects could be compensated by two-hand interaction and increased target size. We explored three mobility conditions (seated, slow walking, normal walking) and three hand postures (one-hand portrait, two-hand portrait, and two-hand landscape). Results show that independently of hand posture, users become indeed situationally-impaired due to mobility. We found that increased hand oscillations result in a significantly decrease of input quality, leading to specific error patterns.

To investigate health-induced tremor impairments (Chapter 5), we focused on older adults due to the age-related increased physiological tremor (Van Den Eeden et al., 2003)(Strickland and Bertoni, 2004)(Benito-Leon et al., 2003). We examined text-entry performance and typing patterns, relating results to hand tremor profile. Moreover, we explored two different touchscreen devices: mobile phone and tablet. Results show that measurements of hand tremor largely correlate to typing performance, suggesting that this should be approached to improve input accuracy.

Following these two user studies, we analyzed the similarities and differences between SIID and HIID, regarding typing performance and tremor measures (Chapter 6). We found out that while older adults are more likely to commit errors that not related with hand tremor, both user groups experience similar error patterns. Indeed the increased demands of mobility and type of device seemingly induce a “Disability Continuum”, where both situationally- and health-impaired user performances become interleaved. Moreover, motion features show promising results to be used to compensate typing errors in both groups.

We designed, developed and evaluated different data-driven models, using data collected from previous user studies. These models address distinct errors and leverage motion information to improve typing performance. We analyzed the effectiveness of different machine learning techniques, prediction models, and user-dependent solutions to conclude that motion data can be used as a unified measure to leverage input accuracy. Moreover, similar classifiers can be used for both target population, as long as they are trained using data from the user group where they will be used. In Chapter 7, we propose a Unifying Model that can be applied to both target audiences.

Finally, we ran a series of simulations to determine whether motion-enhanced models could improve typing performance. Overall, the proposed model theoretically improved typing accuracy on average about 58% and 23.5% for SIID and HIID, respectively. While we focused our research on two user groups (walking and older adults), our findings suggest that the proposed approach and solution should be applicable to a broad range of tremor-impairments and contexts.

1.4. Research Contributions

The work described in this dissertation bridges the gap between HIID and SIID. This was achieved by understanding in depth each user group and assessing their main differences and similarities. We then leveraged motion data as a unifying measure to improve typing performance. The research that led to these results provides the following contributions:

- **Understanding the differences and similarities between motor-impaired and able-bodied users.** As part of our preliminary and exploratory user studies, we analyzed target selection performance using three interaction mechanisms (tapping, crossing, and directional gesturing) and parameterizations (size and position). We found that despite the disparities in error rate, there were clear resemblances between user groups (Guerreiro et al., 2010)(Nicolau et al., 2012b).

- **Effect of applying accessibility solutions in mobile settings.** We investigated accessibility solutions, especially designed for the blind, in visually demanding mobile settings. While this technology transfer approach was not particularly efficient, results of the user study shed light on the main difficulties of typing whilst mobile and allowed us to explore new research avenues (Nicolau et al., 2012a).

- **An in-depth characterization of users' abilities whilst walking and typing.** We studied how walking and hand posture affected typing performance. Moreover, we analyzed the relationship between hand oscillations and input accuracy (Nicolau and Jorge, 2012c). This analysis can be used as the basis for designing future mobile keyboards.

- **An in-depth analysis of older adults typing performance on touchscreens.** We examined text-entry performance and typing patterns of older adults on touch-based devices in order to characterize their abilities. Moreover, we analyzed users' hand tremor profiles, both subjectively and objectively, and their relationship to typing behaviors (Nicolau and Jorge, 2012a). Results can be used to inform future designs of soft keyboards for older adults.

- **A comparative analysis of situationally- and health-impaired users.** Based on previous user studies, we performed a comparative analysis of typing performance between user groups. Particularly, we modeled users' abilities through common types of errors. Also, we propose guidelines for designing unifying solutions (Nicolau and Jorge, 2012b).

- **Design and performance analysis of touch- and motion-based models.** We developed and evaluated error-compensation models that use both touchscreen and acceleration information. These models prevent the insertion of unintentional characters (insertion errors) and deal with uncorrected typed letters (substitution errors). Moreover, we assessed how machine learning techniques, models, personalization, and user group could improve classifier accuracy.

- **Effect of motion-enhanced models on typing performance.** We ran simulations, from participants' data, to assess how proposed models would affect typing performance. Results show an improvement on typing accuracy, implying that motion data should be used as a channel for inclusive, unified, and broader mobile designs.

1.5. Topics Outside Scope

Our research work could have been taken in many directions. This section describes some of these directions, which were not explored in this dissertation:

- **Alternative keyboard layouts and interaction techniques.** Over the years, multiple alternative keyboard layouts and input techniques have been proposed by both research and industrial communities. Within this overabundance of solutions (see (MacKenzie and Soukoreff, 2002b)), we chose the most popular layout: QWERTY. Still, we encourage researchers to explore new techniques and layouts using an approach similar to ours.

- **Language-based correction and prediction algorithms.** Automatic correction algorithms are currently used in most keyboards to improve typing performance. Additionally, letter and word prediction are also common solutions that take into account previously entered text. Nevertheless, all these approaches are language-based, which implies an additional effort of building an algorithm per language. Although this is something that we have not looked at, our models can be easily extended to incorporate language information.

- **Novel machine learning techniques.** Our error-compensation models, take advantage of several touch and motion features. Even though we built customized classifiers, we did not improve on current machine learning approaches. In fact, this is a well established research area and improving learning algorithms has been the focus of several research projects (Witten and Frank, 1999). Our models take advantage of this knowledge by using existing techniques.

- **Clinical measurement of tremor disorders.** Several efforts have been made to use mainstream devices and their built-in sensors to identify and quantify tremor disorders (Van Someren, 1997)(Burkhard et al., 1999)(Gerr et al., 2000)(Hoff et al., 2001)(Burkhard et al., 2002)(Salarian et al., 2007). The everyday use of mobile devices can provide useful information to assess and predict tremor disorders, such as Essential Tremor, Parkinson's disease, or other pathologies (Van Den Eeden et al., 2003)(Benito-Leon et al., 2003)(Strickland and Bertoni, 2004). Indeed, this is a very interesting research area; however, our work does not investigate the usefulness of motion data in clinical settings.

1.6. Publications

Since I enrolled on my doctoral studies, I have co-authored more than 30 peer reviewed papers, including both international journals and conference proceedings. Although all of them contributed to my formation and led to the research described in this dissertation, I only list the peer-reviewed publications that are strongly related to this document.

International Journal Papers

1. *Hugo Nicolau, Tiago Guerreiro, Joaquim Jorge, and Daniel Gonçalves.* **Mobile Text-Entry and Visual Demands: Reusing and Optimizing Current Solutions.** Universal Access in the Information Society (UAIS), Special Issue on Mobile Accessibility. To appear.

2. *Hugo Nicolau, Tiago Guerreiro, David Lucas, and Joaquim Jorge.* **Mobile Touch Screen User Interfaces: Bridging the Gap between Motor-Impaired and Able-Bodied Users.** Universal Access in the Information Society (UAIS), Special Issue on Mobile Accessibility. To appear

3. *Tiago Guerreiro, Hugo Nicolau, Paulo Lagoá, Daniel Gonçalves, and Joaquim Jorge.* **From Tapping to Touching: Making Touch screens Accessible to Blind Users.** IEEE Multimedia, Special Issue on Accessibility, 2008.

International Conference Papers

1. *Hugo Nicolau and Joaquim Jorge.* **The Disability Continuum: Investigating Health and Situational Induced Impairments and Disabilities.** In 14th International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI'12) Workshop on Mobile Accessibility. San Francisco, California, USA, September, 2012.

2. *Hugo Nicolau and Joaquim Jorge.* **Elderly Text-Entry Performance on Touchscreens.** In Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'12). Boulder, Colorado, USA, October, 2012.

3. *Tiago Guerreiro, Hugo Nicolau, Joaquim Jorge, and Daniel Gonçalves.* **Mobile Text-Entry: The Unattainable Ultimate Method.** In Pervasive 2012 Workshop on Frontiers in Accessibility for Pervasive Computing. New Castle, UK, 2012.

4. *Hugo Nicolau and Joaquim Jorge.* **Touch Typing using Thumbs: Understanding the Effect of Mobility and Hand Posture.** In Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '12). Texas, USA, 2012.

5. *Hugo Nicolau.* **Disabled 'R' All: Bridging the Gap between Health and Situational Induced Impairments and Disabilities.** In Doctoral Consortium of the 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'11). Dundee, Scotland, October, 2011.

6. *Hugo Nicolau, Tiago Guerreiro, Joaquim Jorge, and Daniel Gonçalves.* **Towards Mobile Touch Screen Inclusive User Interfaces: Differences and Resem-**

blances between Motor-Impaired and Able-Bodied Users. In Proceedings of the 13th IFIP TC13 Conference on Human-Computer Interaction (INTERACT'11), Mobile Accessibility Workshop. Lisbon, Portugal, September, 2011.

7. *David Lucas, Hugo Nicolau, Tiago Guerreiro, and Joaquim Jorge.* **Investigating the Effectiveness of Assistive Technologies on Situationally Impaired Users.** In Proceedings of the 13th IFIP TC13 Conference on Human Computer Interaction (INTERACT'11), Mobile Accessibility Workshop. Lisbon, Portugal, September, 2011.

8. *Tiago Guerreiro, Hugo Nicolau, Joaquim Jorge, and Daniel Gonçalves.* **Towards Accessible Touch Interfaces.** In Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'10). Orlando, Florida, USA, October, 2010.

9. *Tiago Guerreiro, Hugo Nicolau, Joaquim Jorge, and Daniel Gonçalves.* **Assessing Mobile Touch Interfaces for Tetraplegics.** In Proceedings of the 12th International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI'10). Lisboa, Portugal, September, 2010.

1.7. Dissertation Outline

The remaining of this dissertation is organized into eight chapters. Following this introduction, we present a literature review on design approaches and interfaces that aim to provide IT access to broader user populations. Particularly, we focus on adaptive and adaptable user interfaces.

Chapter 3 describes our preliminary user studies and discusses the challenges and opportunities in designing for both situationally- and health-induced impairments and disabilities. Moreover, we detail our approach, which served as a guide for all research decisions.

Following this exploratory research stage, we then sought to understand the abilities of both user groups. Chapter 4 presents a detailed analysis of situationally-impaired users performing text-entry tasks. We explored several mobility conditions and gathered typing and motion data. In Chapter 5, we investigated touch typing with older adults, who present increased physiological hand tremor. We made a thorough characterization of users' abilities by measuring tremor profile recurring to different techniques and related them to typing performance. Chapter 6 discusses the main differences and similarities between user groups and highlights research opportunities in using motion data to compensate input errors.

Chapter 7 details the design, development and evaluation of error-compensation models. We propose several touch and motion features, as well as multiple classifiers that can be used to compensate insertion and substitution errors. We end this chapter by proposing a unifying model that leverages accelerometer data to improve typing performance for both target populations. Moreover, we assessed text-entry performance by means of the newly proposed model within previously collected user data. The results thus obtained, should encourage designers to build solutions that take into account a wide range of abilities.

In Chapter 8, we conclude by presenting the major results and an overall discussion of the key findings of our research. Additionally, we describe the benefits and limitations of this thesis as well as promising directions for further work in the area.

2 Related Work

This chapter provides an overview of research on accessible computing. First, we define concepts that will be used throughout this dissertation, such as impairment, disability, handicap, and context. Next, we describe some of the most relevant design approaches that attempted to deal with a wide range of human abilities, and how they evolved over time within the accessible computing research field. Finally, we present and discuss projects that tried to apply that knowledge by designing solutions for broader user groups.

2.1. Terminology on Impairments, Disabilities, and Handicaps

Concepts such as impairment, disability, and handicap are typically confused and freely mixed with each other. We believe that a clear understanding of the differences between each term will be useful to better understand the remainder of this dissertation. Therefore, in this section we present a definition for each concept and draw relationships between them.

The most recent report of the World Health Organization[1] sheds some light on these issues. According to the International Classification of Functioning, Disability and Health (ICF) (WHO, 2009), *impairments* are problems in the functions or body structures, which

[1]http://www.who.int/classifications/icf/en/, last visited on 07/11/2012

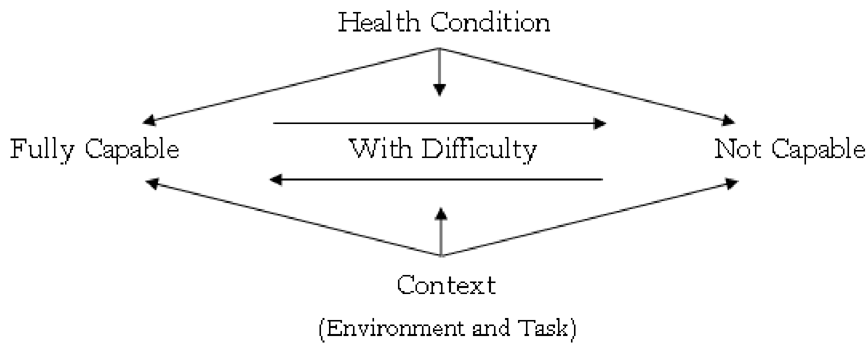


Figure 2.1: Relationship between health condition, context, and abilities. Image inspired in (Sears and Young, 2003).

can be revealed due to a specific health condition (e.g. nearsightedness). For example, people with arthritis (health condition) usually experience loss of control and strength in their movements - impairment. Still, contextual factors can also induce impairments. For instance, low temperatures can affect finger dexterity, the same way walking induces hand tremor.

A *disability* relates to the performance of a given task. Typically this refers to either performance challenges or inability to operate some device. Disabilities can be caused by impairments - loss of control of upper limbs (impairment) precludes people from using traditional input devices, such as a mouse and a keyboard. On the other hand, a long walk on a beach during the day does not result in an impairment - the users' visual abilities are not changed - but can induce a disability, since people may not be able to read the screen contents due to sun glare.

Finally, *handicaps* mostly refer to social effects of disabilities. People with communication disabilities may struggle to maintain social relationships with others; or children with learning disabilities may be excluded from educational activities in their school.

Overall, the ICF defines both concepts and relationships - health condition, impairment, disability, activity, handicap - that allow people to have a discussion on a common ground. Most importantly, this framework highlights that all individuals experience impairments and disabilities due to environmental factors:

> *"... every human being can experience a decrement in health and thereby experience some degree of disability. Disability is not something that only happens to a minority of humanity. The ICF thus 'mainstreams' the experience of disability and recognizes it as a universal human experience."*

By including environmental factors in this classification, the relationship between impairment and disability is no longer seen exclusively as either a "medical" or "biological" issue. Context plays an increasingly important role in users' (dis)abilities (see Figure 2.1).

In fact, context is not confined to location and environmental factors. There are innumer-

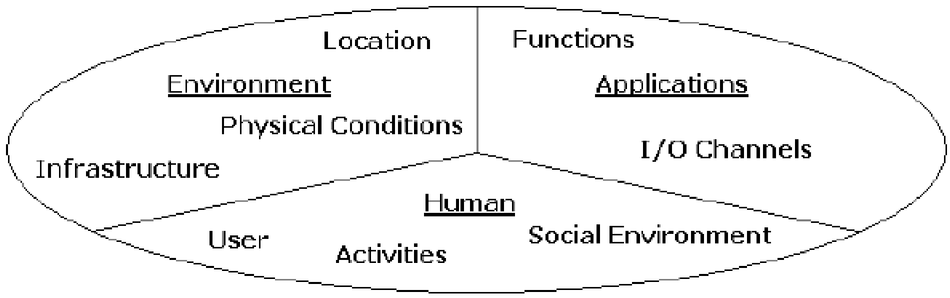


Figure 2.2: 3D context space (Sears et al., 2003).

ous situations or factors that can induce impairments or disabilities. Several authors have attempted to provide a definition of context (Pascoe, 1998)(Salber et al., 1999)(Abowd et al., 1999)(Dey, 2001). For instance, the general definition by Dey (Dey, 2001) says that:

> *"Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves."*

In this dissertation, we adopt Sears et al. (Sears and Young, 2003) tri-dimensional model that combines Schmidt et al. (Schmidt et al., 1999) 3D model with Dey et al. (Dey, 2001) definition. We believe that the model illustrated in Figure 2.2 provides a broad definition of context, enabling us to classify and comprise a large number of situations, while explicitly defining a well-known set of dimensions: Human, Environment and Applications.

Environmental factors, such as light, noise, vibration, and temperature may restrict the use of technological devices. Similarly, the activities (Human dimension) in which users are engaged may result in a cognitive, perceptual or motor overload of their abilities. The Application dimension relates to the device's characteristics, such as input and output modalities, which may prove inadequate to use when taking into account all other context dimensions. The impairments and disabilities, caused by context, are usually called Situationally-Induced Impairments and Disabilities (SIID) (Sears et al., 2003). On the other hand, impairments and disabilities caused by health conditions are called Health-Induced Impairments and Disabilities (HIID).

Although we deliberately draw a relationship between SIID - HIID and users' abilities, there are significant differences between these two groups. First, while HIID tend to be permanent, SIID are usually temporary. As a result SIID users generally do not have the time needed to develop compensation mechanisms (Sears and Young, 2003). Second, in addition to being temporary, SIID are highly dynamic: they can be progressive, regressive or static, depending on the users' context. Also, these impairments and disabilities can be experienced in different levels of severity (low, medium, high, etc.). On the other hand, HIID are often static and changes on the users' abilities are seen in a larger time-scale.

2.2. Accessible Computing Approaches

Several approaches to accessible computing have been proposed since the dawn of human-computer interaction (HCI) research area. Unavoidably, many of them share the same concepts and goals. In this section, we review some of the most relevant design approaches to accessible computing and how they evolved over time to fit the needs of our research field.

We start by the term '*User-Centered Design*' (UCD) that originated in Donald Norman's research laboratory at the University of California San Diego in 1980s and it became widespread with the publication of his book (Norman and Draper, 1986). This concept shifted the focus of interface design from systems to users. UCD is a term to describe design processes in which end users influence how a design takes shape. There are several methods to include users in this process, but the important aspect is that users are involved. For instance, during the requirements gathering phase and usability testing or even during the whole process as design partners. The goal of this approach is to make sure that end-users are able to use the system or product as intended with minimal effort. Moreover, Norman proposed seven principles of design, which were then refined by Ben Shniederman in 1987 (Shneiderman, 1987). Later, these were popularized by Nielsen (Nielsen, 1993) through his famous 10 heuristics for usability engineering. UCD is a broad design philosophy embedded within almost every modern design approach.

Directly related to accessible computing are *Assistive Technologies* (ATs). AT is an "umbrella term" that includes all software and hardware solutions for people with disabilities (Cook and Hussey, 2001). Its main goal is to promote greater independence for people by enabling them to perform tasks they were once unable to accomplish, changing the way users interact with devices. These technologies are usually seen as being useful only to a minority. Assistive technologies are associated with devices for "non-standard" users by means of an assistive component that bridge the gap between the user and the system. This often requires additional development and costs, since the component is not part of the original solution. Therefore, the burden of change lies with users, since the system does not know anything about the user's abilities. While we agree that this is the only option in some cases, particularly if we consider "non-intelligent" hardware, such as a cane, wheelchair, and hearing-aid, this definition becomes obsolete when considering computer-oriented technologies that are able to monitor users' performance and adapt (Gajos et al., 2007) or make suggesting for adaptations (Trewin, 2003)(Koester et al., 2007a).

Newer approaches to *Engineering Human Performance*, focus on building engineering models through quantification, measurement, and tracking of human performance to provide better system adaptations (Kondraske, 1995)(Persad et al., 2007). These processes typically require a large battery of tests with non-technology-related tasks, which assess users' low-level capabilities. For instance, Persad and colleagues (Persad et al., 2007) propose a Capability-Demand Theory to evaluate product accessibility, where sensory, motor, and cognitive dimensions of user capability are assessed to estimate the number of people potentially excluded from using a given product (demands). For instance, a remote controller with small labels (visual demand) may exclude a high percentage of the population from using it (e.g. older adults and visually impaired). By quantifying the visual abilities of the target audience, we can predict how many people will not be able to use the product. This highlights the variability of users as well as the need to build new solutions to accommodate a wide range of capabilities. Still, this approach is often static; that is, it assumes that the product is immutable. Thus, accommodations to technology require the development of completely new solutions or add-on's that lowers the device's demands.

In 1995, Newel proposed the concept of *Extra-Ordinary Human-Computer Interaction* (Newell, 1995), recognizing that all users have abilities, and some users have extra-ordinary abilities. For the first time, this drew the parallel between "ordinary" people operating in an "extraordinary" environment (e.g. adverse noise and lightning conditions, high work load) and "extra-ordinary" (disabled) users operating in an ordinary environment. The author acknowledges, but does not quantify, that context can temporarily reduce people's abilities in ways similar to health-related impairments. Similarly to Sears and colleagues (Sears et al., 2003) definition of context and situationally-induced impairments and disabilities, Extra-Ordinary Human-Computer Interaction was the first attempt to relate human abilities to context.

Around the same time, the concept of *User Interfaces for All* (UI4All) was proposed (Stephanidis, 1995) as a vehicle to efficiently and effectively address the numerous and diverse problems related to the accessibility of interactive applications in different contexts of use. Particularly, UI4All promotes the use of *unified user interfaces* (Stephanidis, 2001) to support user independent interface development. In a unified user interface, only the core functionality is developed, while abstract user interface representations map to one concrete interface template, either at configuration- or at run-time. Special purpose user interface software components automatically manage the specific issues and adapt the interface to a particular user or user group (Harper, 2007). Nonetheless, this approach lacks the means to measure and develop interfaces for these users, placing the burden on the developer.

Universal Design (Vanderheiden, 1998) and *Design for All* (Stephanidis and Salvendy, 1998) are very similar concepts that emerged in Europe and introduced the visionary goal of an Information Society for all. These approaches focus on applying a set of principles, methods, and tools to develop technological products and services that are accessible and usable by all citizens, therefore avoiding the need for adaptations or specialized design. These approaches emerged as a response to the limitations of add-on approaches like assistive technologies. Universal design advocates a “one size fits all” solutions, which in our view, may suit door handles or building entrances, but comes short on many technological domains.

After Universal Design and Design for All, emerged the concept of *Universal Usability*. Similarly, this approach provides guidelines for designing interfaces that are usable by the widest range of people that is possible (Harrin, 2008)(Shneiderman, 2000)(Vanderheiden, 2000). It attempts to answer questions, such as “what can everyone do?” . As universal design and design for all approaches, universal usability also focuses on a “one size fits all” ideal. However, it does not specifically aim at people with disabilities, and is more concerned with the gap in access due to other aspects, such as gender, literacy, economic status, culture, etc.

User-Sensitive Inclusive Design came primarily from the UK and was proposed by Newell and Gregor in 2000 (Newell and Gregor, 2000). It acknowledges that “design for all” is a very difficult, if not even impossible task. The authors state that:

> *“Providing access to people with certain types of disability can make the product significantly more difficult to use by people without disabilities, and often impossible to use by people with a different type of disability”.*

Inclusive design emerges from the premise that disabled users have a much greater variety of characteristics than able-bodied people and it is usually difficult to find “representative” user groups. The use of the term “inclusive” rather than “universal” reflects the view that “inclusivity” is a more achievable, and in many situations, appropriate goal than “universal design” or “design for all”. Also, the authors suggest that some significant differences must be introduced into the User Centered Design Paradigm. Therefore, “sensitive” replaces “centered” to underline the extra levels of difficulty involved when the range of functionality and characteristics of the user groups can be so great that it is impossible to produce a representative sample of the user group.

More recently, Wobbrock et al. (Wobbrock et al., 2011) proposed a new approach called *Ability-Based Design*, which consists of focusing on ability throughout the design process in an effort to create systems that leverage the full range of human potential. As user-centered design shifted the focus of interactive system design from systems to users, ability-based design attempts to shift the focus of accessible design from disability to ability. This concept provides a unified view of able-bodied and disabled users, as well as health-related impairments and situationally-induced impairments. The authors focus on how systems can be made to fit the abilities of whoever uses them, which is closely related with the notion of adaptation.

2.2.1. Challenges on Accessible Computing

Multiple approaches, same goal, different avenues. There are numerous approaches to accessible computing and they all share the common goal of providing access to technological solutions, thus improving quality of life and guaranteeing equality of rights in an increasingly digital world. How to achieve this goal, however, may vary between approaches. For instance, universal-related (Vanderheiden, 1998)(Stephanidis and Salvendy, 1998)(Harrin, 2008)(Shneiderman, 2000) approaches focus on lowering solutions demands so they can be used by everyone (Guerreiro et al., 2009), while others center their attention on individual needs (Newell and Gregor, 2000). Similarly, while some focus on measuring and modeling low-level human abilities (Persad et al., 2007) recurring to clinical assessment, others use human performance in computer-oriented tasks to accommodate users' needs (Wobbrock et al., 2011). The real challenge consists in knowing which approach is more appropriate to solve the problem at hand.

Universal interfaces, a utopian goal? Building universal interfaces that can be used by everyone is the "saint graal" of accessible computing. A world where computerized devices can be equally used by all people, independently of their motor, sensory and cognitive abilities is certainly a utopian world. Although "universal" approaches seem to believe in this world, much due to the inherited design principles that grew out of architecture (Mace et al., 1990)(Steinfeld, 1994)(Story, 1998), this is nearly impossible to achieve when considering computer-oriented interfaces. The founders of universal design were mainly concerned with physical spaces and while this may be a fruitful approach for architecture solutions, such as door handles and building ramps, designing "one size fits all" computer interfaces is unfeasible. The number of tasks and uses of current mobile devices, as well as the spectrum of users' profiles is too wide to be accommodated in a single interface. Even if we think of brain-computer interfaces, and we could explicitly assess users' intentions, cognitive abilities would still play a crucial role. Nonetheless, we believe that current interfaces can be greatly improved to cover wider audiences, and new solutions can be designed for those who are currently excluded from access to information.

A unified view of users' abilities. In order to provide interfaces that can be used by a wide range of users, designers and researchers should focus on users' abilities, rather than their dis-abilities (Wobbrock et al., 2011). Accessible computer interfaces need to be designed for "what a person can do", instead of "what disabilities does a person have?". Abilities vary across a limited range and are influenced by the person's characteristics as well as the surrounding context (Newell, 1995). In our research work, we adopted this unified view of users' abilities, dealing with the variability of health-related conditions (Chapter 5) and contextual factors (Chapter 4). Moreover, this approach enabled us to place both user groups on the same "ability scale" (Chapter 6) and design common solutions (Chapter 7).

A common need: assessing user's abilities. A common need of any accessible computing approach is the assessment of users' abilities. There are several ways of doing this: assessing demographic characteristics, such as age or height; performing clinical functional assessment (Cook and Hussey, 2001)(Oliveira et al., 2011a); or characterizing disabilities through subjective scales (WHO, 2009). While these approaches may be useful to have an overall view of a person's profile, it is usually very difficult to relate and generalize them to computer-skills. An alternative, which was followed in the remainder of this dissertation, was to detect user's abilities as they interacted with technology (Gajos and Weld, 2004) and, therefore, inform user interface at design-time. Another topic of interest is in when to perform this assessment. A "battery of tests" can be administered once, assuming that users' abilities will not change over time (or very little), or periodically for users whose abilities are constantly changing. A more difficult solution would be to test it "in the wild", without the need of assistance (Hurst et al., 2008) or previous knowledge about the task (Trewin et al., 1997). In this case a new set of challenges emerge, since we need to infer about users' intentions in order to measure their performance.

Improve solutions by understanding the context. In an increasingly mobile world, abilities are influenced by the context in which they are exercised (Gregor et al., 2002). Sensing context and leveraging collected data to improve users' performance with current technologies is still in its early stages. First, we need to understand how factors, such as light (Barnard et al., 2007), noise, or mobility (Schildbach and Rukzio, 2010)(Nicolau and Jorge, 2012c) influence users' abilities in order to remove the experienced difficulties. Some works in activity recognition hold great promise to achieve this goal (Choudhury et al., 2008)(Hinckley et al., 2000). Recent works have attempted to use contextual data to provide better suited interfaces (Kane et al., 2008b)(Goel et al., 2012). In our work, we went a little further and used context information (motion data) as a unifying measure between users with HIID and SIID (Chapter 7).

Modeling, modeling, modeling. After performance has been accurately measured and/or context has been sensed, there still remains the challenge of describing (model) users' abilities. This step is particularly important if adaptation is to take place. Describing abilities in terms of users' performance is an open research challenge, especially considering the variability of users with impairments, even with the same health condition. When considering both HIID and SIID, we expect this variability to increase. In Chapter 7, we propose several features and models that attempt to explain users' performance.

Apply all the knowledge. Once users' abilities have been measured and modeled through performance and/or context, we need to incorporate all that knowledge into the user interface. One option is to give full control to end-users enabling them to customize the interface to their needs and preferences (Koester et al., 2007a). However, users need to be aware of this feature. Still, there is no guarantee that they will, or more importantly, that they can do that, ending up with an interface that does not accommodate their abilities. An alternative is to automatically adapt the user interface, requiring little or no effort on the part of the user. Nevertheless, this approach also has a set of challenges of its own (Höök, 2000): lack of control, predictability, transparency, obtrusiveness, and privacy. Further research, should explore new ways of dealing with these issues, and if possible, combining the advantages of adaptive and adaptable interfaces.

2.3. Maximizing Accessible Computing

Most interfaces are designed for the "average user", typically a younger user with static abilities over time (Gregor et al., 2002). This approach completely disregards the fact that abilities vary widely across users, especially when considering disabled people, and that individual users are not static entities. The consequences are poorly designed interfaces that fail to deal with individual needs and preferences.

In this section, we provide a review of research that attempts to cope with users' wide range of abilities, instead of *all* abilities. Further detail on relevant literature, such as the effect of health- and situational-related impairments and disabilities, as well as proposed solutions, appears in the individual chapters to follow (Chapter 4 and 5). In this section, we describe relevant solutions that focus on adapting to each user's needs by placing the burden of change on the system, not the users.

Physical Keyboards

Works presented in this subsection can be divided in four types of solutions: 1) filters (Trewin, 2002)(Trewin et al., 2006), which are software solutions intended to reduce or eliminate specific kinds of typing errors (Novak et al., 1991); 2) configuration and suggestion systems responsible for recommending alternatives of new values for some parameters (Trewin et al., 1997)(Koester et al., 2007a); 3) orthographic correctors and word predictors (Kane et al., 2008a), which are language-dependent solutions that aim to reduce overall typing accuracy and improve speed; and 4) alternative input methods, which use new keyboard layouts (MacKenzie and Soukoreff, 2002b), interaction techniques (Wobbrock et al., 2003)(Kristensson and Zhai, 2005), or no keyboards at all (Guerreiro et al., 2008).

The Dynamic Keyboard, proposed by Trewin and Pain (Trewin et al., 1997), was able to dynamically model physical skills by monitoring users' typing performance and suggest keyboard parameters for a specific user. Typing data was captured "in-the-wild" and suggestions to several accessibility features such as key repeat delay, stickykeys[2], overlapping keys, and bouncing delay, were given. The user model contained measures like average key press length and variation, number of double presses, and number of accidental insertions. Language-dependent features (bigrams) were also included in order to estimate the

[2]http://windows.microsoft.com/is-IS/windows-xp/help/using-stickykeys

likelihood of errors. Note that no changes were made to the actual keyboard configuration in use, the model uniquely made recommendations to users. The authors applied their solution to motor impaired people and showed improvements in typing performance.

Later, the same authors observed motor-impaired users' typing behaviors and found that unintentional key presses were the most common cause of errors (Trewin and Pain, 1999). Thus, by leveraging the dynamic keyboard solution, Trewin proposed a filtering solution, the *Invisible Keyguard*, that modeled keystroke timing characteristics and accurately filtered 80% of errors (Trewin, 2002).

Koester et al. (Koester et al., 2007a) also presented a system, called *Input Device Agent* (IDA), whose goal was to optimally configure three keyboard features (repeat rate, repeat delay, and use of stickykeys) for people with physical impairments. Moreover, the model adapted itself to users' performance. Similarly to the dynamic keyboard, IDA only provided suggestions for changing keyboard features. Results showed that this solution was able to improve typing accuracy and efficiency.

More recently, Yesilada and colleagues (Yesilada et al., 2010a) adopted the dynamic keyboard and applied it to walking conditions. From their observations, mobile users experience similar types of errors of desktop motor-impaired people (Chen, 2008)(Yesilada et al., 2010b). Therefore, they applied a technology transfer solution, reusing a model built for motor-impaired users. The authors report that four types of errors were significantly reduced: long key presses, bounces, additional characters, and key ambiguity errors.

TrueKeys (Kane et al., 2008a) opted for a correcting approach by combining models of word frequency, keyboard layout, and typing error patterns to automatically correct typing mistakes. The system was envisioned to deal with motor-impaired users' abilities; however, able-bodied users also showed to perform significantly more accurately. After typing each word and hit the SPACEBAR key, the system automatically corrected the word or showed a list of the most probable words that users intended to type (Figure 2.3). Due to this correction stage, typing speed was reduced. Nevertheless, performance gain was similar to both motor-impaired and able-bodied users (0.8% - 1.35%).

Wobbrock and Myers proposed an alternative input method (Wobbrock et al., 2007) for people with spinal cord injuries, enabling them to write using gestures on a trackball device. The *EdgeWrite* technique and alphabet (Wobbrock et al., 2005) (see Figure 2.4) were originally designed for mobile devices and users with severe tremor (Wobbrock et al., 2003). This technique employs adaptive timeouts, slip detection, word prediction, and word completion times to improve users' typing performance. Moreover, the authors suggest, but do not demonstrate, that this solution can also be applied to users with SIID.

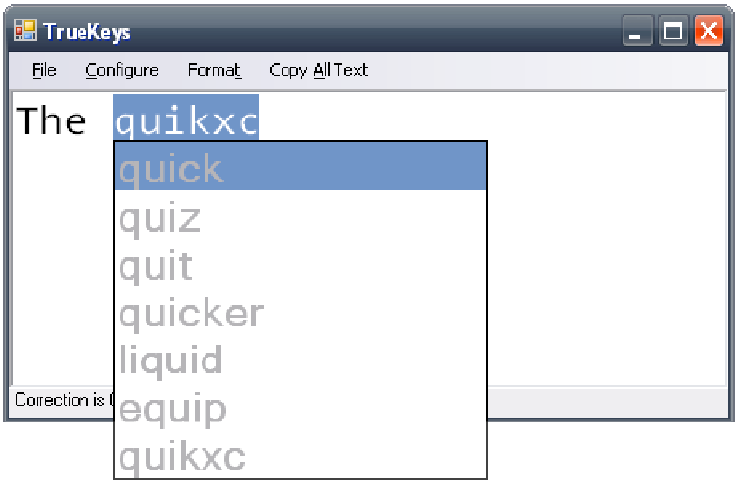


Figure 2.3: The TrueKeys (Kane et al., 2008a) user interface performing a correction from "quicxc" to quick.

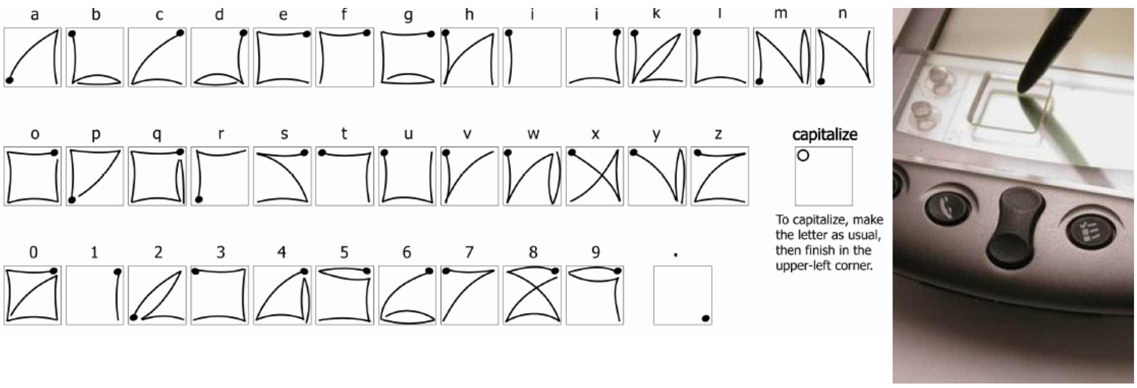


Figure 2.4: EdgeWrite's (Wobbrock et al., 2003) character forms. EdgeWrite template over the text input area (right).

Mouse Pointing

Slipping while clicking and accidental clicks are two of the most common types of errors for mouse users with motor impairments (Keates et al., 2005). The *Steady Clicks* (Trewin et al., 2006) suppresses these errors by freezing the cursor during mouse clicks. Once again, using simple and filtering approaches, Trewin allowed users with a wide range of abilities to improve their selection accuracy. In fact, this technique has also been used for pen-based devices showing positive results for older adults as well (Moffatt and McGrenere, 2010).

Koester et al. (Koester et al., 2005) also explored the effects of recommend mouse gain (a measure of how far the cursor moves relative to the amount of physical movement sensed by the input device) to individual pointing performance in target acquisition tasks. The system's recommendations showed no significant differences as compared to the operating system's default setting or users' preferred settings. This effect may be explained due

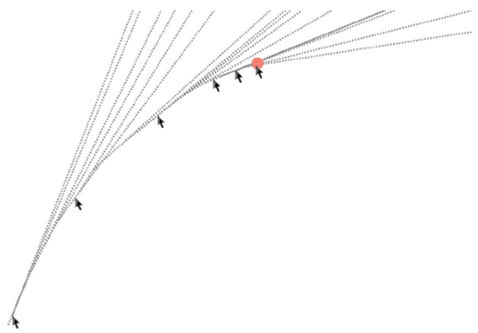


Figure 2.5: Mouse path showing 16 angle samples and their spread during movement. AngleMouse (Wobbrock et al., 2009) dynamically adjust gain in response to this spread.

to the experimental conditions (e.g. target sizes and distances) or unfamiliarity with the new gain value. However, the authors did find a significant effect of gain across all tested conditions, showing that by manipulating this feature we can improve or decrease pointing performance. Later, the authors extended their work and proposed an evaluation suite (i.e. test battery) to assess an individual's computer input skills based on pointing, text-entry, and switch tasks (Koester et al., 2006)(Koester et al., 2007b).

Hurst et al. built an *Automatic Mouse Pointing Assessment* (Hurst et al., 2007) tool to identify pointing problems during every day ("in-the-wild") computer use. They have studied real world pointing use for older adults and individuals with motor impairments. The authors were able to discern novice from skilled computer users with 91% accuracy in menu selection tasks. Moreover, this tool correctly identified 90% of people needing adaptation when tested in young and older adults, and users with Parkinson's disease (Hurst et al., 2008).

Wobbrock and colleagues (Wobbrock et al., 2009) proposed a method of dynamic gain adaptation that requires no knowledge of target locations or dimensions (see Figure 2.5). *AngleMouse* adjusts the mouse gain solely based on the movement performance, particularly angle deviations. This solution was designed based on observations that motor-

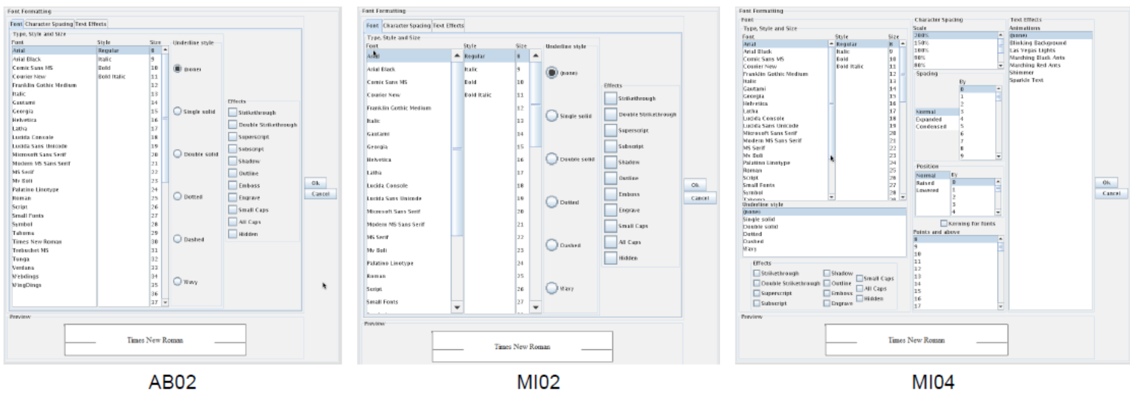


Figure 2.6: User interfaces automatically generated by SUPPLE++ (Gajos et al., 2007) for the font formatting dialog based on three users' individual motor abilities.

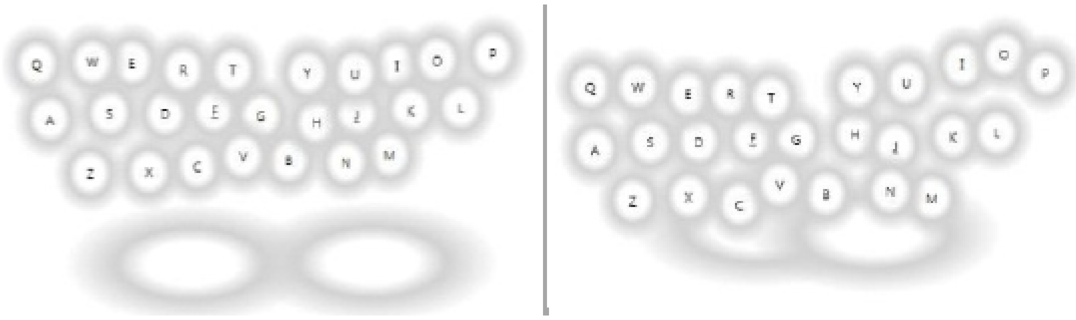


Figure 2.7: Personalization with a visually adaptive keyboard, showing the adapted layout for two participants (Findlater and Wobbrock, 2012).

impaired users are able to perform reasonably well in the initial phase of pointing, but fail when fine adjustments are needed (Hwang et al., 2004)(Wobbrock and Gajos, 2007). Results show that AngleMouse improved pointing performance for user with motor impairments, while remaining unobtrusive for able-bodied participants.

SUPPLE++ (Gajos et al., 2007) is probably one of the most known solutions in the adaptive interfaces research area. The system automatically generates interfaces which are tailored to an individual's motor and vision abilities (Figure 2.6). SUPPLE++ models users' abilities based on a "test battery" of pointing, clicking and dragging tasks (Gajos and Weld, 2004), generating a personalized interface (Figure 2.6). Unlike other approaches that adapt input device parameters, SUPPLE++ generates more accessible interfaces. Results show that the system is able to produce interfaces more adequate to its users, allowing them to complete tasks that otherwise would not be possible, and improving efficiency by 20% (on average).

Touch Input

Oliveira et al. (Oliveira et al., 2011a) showed that blind users' (low level) individual attributes have a significant impact on text-entry performance and that this effect is related with different methods' demands. Moreover, the authors suggest that small adaptations to user interfaces are often not effective, due to the high variance in abilities, and different input techniques may be needed depending on individual differences.

More recently, Findlater and Wobbrock (Findlater and Wobbrock, 2012) proposed a solution called *Personalized Input*, which leverages the potential of touch devices and dynamically adapts the QWERTY keyboard layout (Figure 2.7) and classification models to suit each user's typing pattern. The authors also explored the visual component of these keyboards by comparing visible and invisible adaptations. Results show that keyboards tailored to individual abilities improved typing speed, however visually adaptive keyboards can hinder performance.

Montague et al. (Montague et al., 2012) proposed the use of *Shared User Models* as a new

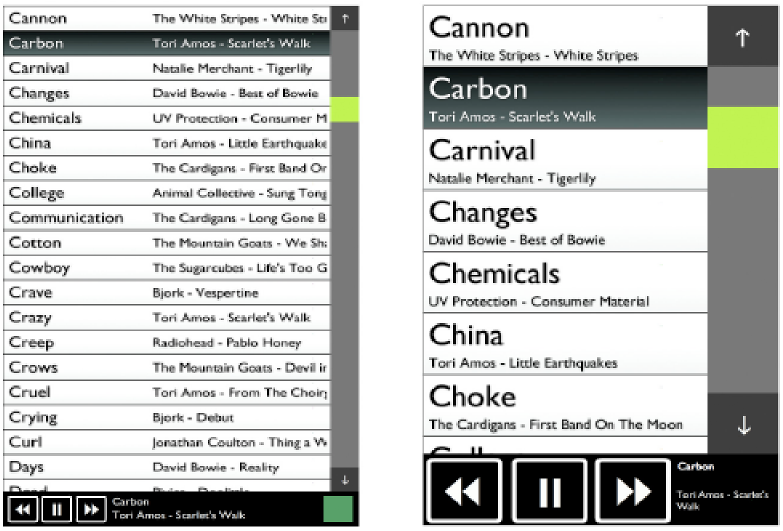


Figure 2.8: Music player user interface (Kane et al., 2008b) in two sizes: the player while standing (left); the player while walking (right).

method for providing cross-device and cross-application adaptations; that is, independent applications can learn from each other and share knowledge. This approach allows users' abilities to be collected during application use and be automatically applied to all interfaces in order to accommodate individual abilities/preferences. Preliminary results, gathered in both laboratory and "in-the-wild" were very promising, as adaptive approaches are able to compensate some errors experienced with static interfaces.

In 2008, Kane et al. (Kane et al., 2008b) coined the term *Walking User Interfaces* and explored an interface that adapted their layout accordingly to users' movements. The interface dynamically changed button sizes, based on accelerometer readings (Figure 2.8), allowing users to accurately select an item from a music list. Still, this solution showed to perform at the level of its static counterpart, mostly because larger buttons forced users to perform scrolling actions, thus counteracting the benefits of the adaptive interface.

Similar results were reported when trying to compensate vibration during reading tasks by constantly moving the screen content in the opposite direction (Rahmati et al., 2009). Yamabe and Takahashi (Yamabe and Takahashi, 2007) have also used an inertial sensor to automatically adapt font size and images in different mobility conditions, however with limited gain. On the other hand, *WalkType* (Goel et al., 2012) was able to accommodate situational impairments, using accelerometer data, by reducing typing errors; the authors propose a model that is able to compensate motion and cope with different walking speeds and typing behaviors.

2.3.1. Discussion on Maximizing Access

In this section, we review several projects that attempted to maximize access to computers by designing interfaces that cope with a wide range of abilities and situations. In this subsection, we compare the surveyed solutions following five criteria: *adaptability*, *adaptation source*, *context*, *effect*, and *ability range*. Table 2.1 presents the overall comparison. Each one of the criteria is explained and discussed below.

Adaptability

Adaptability refers to the approach used to accommodate users' abilities: self-*adaptive* or user-*adaptable*. Each of these approaches has its own strengths and weaknesses (Höök, 2000), however most of presented projects take an adaptive approach (see Table 2.1). Indeed, intelligent user interfaces seem to be the future of human-computer interaction. More than ever, most of us spend our day near computerized devices, both desktop and mobile, which now have the capabilities to sense and understand usage patterns. Adapting interfaces to each user is the next step. For what matters for this dissertation, adaptive interfaces have the potential to measure, model and automatically adjust themselves to users' abilities. SUPPLE (Gajos and Weld, 2004) is an example of a system with high adaptability, since it can transform the entire user interface, while others, such as trackball EdgeWrite (Wobbrock et al., 2009), only deal with small local adjustments.

On the other hand, some projects are user-adaptable; that is, they provide useful recommendations to activate or tune certain interaction parameters (Trewin et al., 1997) that will improve users' performance. End-users have complete control of how, when, and if they want to perform these changes. Privacy issues are also minimized when using this approach as users may choose what information that want to provide to the system (Trewin, 2000) .

Finally, while we acknowledge that interfaces can greatly benefit from adaptation, either self-adaptive or user-adaptable, this is not a crucial requirement in order to deal with a wide range of abilities. For example, SteadyClicks (Trewin et al., 2006) allow both motor-impaired and older adults to improve their clicking performance. Even though, both these user populations vary greatly in their range of abilities. Similarly, work by Hurst et al. (Hurst et al., 2008) showed that it is possible to automatically identify pointing problems from a wide range of user populations (young, older adults, motor-impaired, and users with Parkinson's disease) using a common solution.

Adaptation Source

All adaptive or adaptable works have to be able to measure, model and apply changes to the interfaces. Adaptation source represents the data that is used to perform those actions. In fact, most successful examples have something in common: they take advantage of users' performance with technology in order to improve it (see Table 2.1).

Obviously, the source of adaptation is tightly related with the task at hand. Moreover, data collection can be performed with a *controlled* set of tasks or *"in the wild"*. Reliably measuring users' abilities outside a controlled test battery is a challenging task. While in a controlled test designers know how to identify success and errors, accurately measuring tasks in the wild introduces uncertainty and requires inferring the users' intentions. For instance, knowing what users are looking at, trying to accomplish, trying to select, attempting to write, and so forth.

The automatic mouse pointing assessment utility (Hurst et al., 2008), needs to know where targets in the interface are located, segment mouse movements into discrete aimed pointing movements, and infer what mouse behaviors constitute an error. In the same way, understanding text-entry errors without knowing à priori what users intended to write; that is, without presenting to participants a well-known set of phrases, can be a daunting task. Some efforts have been made in this direction, using heuristics and language-based approaches (Wobbrock and Myers, 2006)(Baldwin and Chai, 2012).

Context

The context criterion refers to the ability to leverage the device's sensors to collect data and improve users' performance. Overall, this research area is still in its infancy, since inferring context through sensors such as accelerometer, GPS, or gyroscope can be a challenging task. Still, this is a very promising research and application field for the near future (Hinckley et al., 2000), mostly due to the ubiquity of mobile devices.

Sensing the context in which users are in can also be of great value to infer their abilities, particularly if they are the target of situational impairments. Furthermore, using that knowledge to change the interface and cope with their needs has great appeal. Current sensors can be the vehicle to deal with the dynamic and heterogeneous nature of SIID. Some projects have started to emerge in this area (Kane et al., 2008b)(Rahmati et al., 2009)(Goel et al., 2012); however, results are mixed: some are able to compensate for the problems imposed by context, while others showed to be ineffective.

In our research work, in addition to sensing context to compensate input errors, we also demonstrate that it can be an effective approach to unify the domains of SIID and HIID (see Chapter 7). Future work should investigate the effect of different contextual factors, while "on the go", such as environmental factors (light, temperature, noise, rain, wind, etc.), dual-tasking (meetings, working, walking in a busy street), or even social contexts (parties, concert, movie theater).

Effect

Accessible computing solutions can have a *visible* or *invisible* effect. If we consider self-adaptive approaches, especially continuous and dynamic adaptations, visual feedback can be unsettling and have a negative effect that counteract the effects of adaptations (Findlater and Wobbrock, 2012). Indeed, when visual adaptations are employed one should consider the effect of those adaptations on users' performance, which means that their abilities should be again measured and modeled (accounting for the adaptation effect). This creates an adaptation cycle that should be treated carefully, by ensuring that user performance converges rapidly.

In the case of invisible adaptations, prior work have reported benefits (Paymans et al., 2004)(Findlater and Wobbrock, 2012); still, transparency and predictability are important factors to consider. Users may not understand how the interface is adapting and may have trouble in predicting what the system's response will be to their actions. Nonetheless, it is not clear how much of the underlying adaptive process should be exposed to the user. Proving visual feedback can lead to increase trust towards the system, but not all users gain a greater understanding or even want that information (Bunt et al., 2007).

From all reviewed works, only Findlater and Wobbrock (Findlater and Wobbrock, 2012) compared visual versus non-visual adaptations. The remaining surveyed authors opted for one condition. For instance, SUPPLE++ used a visible approach, as it performed long-term changes to the interface. Kane et al. (Kane et al., 2008b) also opted for a visual change by dynamically increasing target sizes. On the other hand, the invisible keyguard (Trewin, 2002), or anglemouse (Wobbrock et al., 2009) did not make their adaptations visible.

Ability Range

As stated in the beginning of this section, our goal was to present some of the most relevant projects that aimed to address a wide range of abilities. Indeed, several interesting works were reviewed, some using adaptive or adaptable functionality, while others used static interfaces. Yet, many focused on the variability of abilities within a user group (Table 2.1), either motor-, visual- or situationally-disabled. Some explored more than one type of impairment, but still focused on health-induced conditions (e.g TrueKeys (Kane et al., 2008a), SteadyClicks (Trewin et al., 2006)).

Only Yesilada et al. (Yesilada et al., 2010a) investigated the use solutions for users with HIID and SIID, using a technology transfer approach. This dissertation extends this work focusing on touch-based devices and taking advantage of their full potential. Moreover, we use the device's sensors to access users' abilities, either affected by SIID or HIID, and improve their performance. From our knowledge, no one has ever explored this approach. We analyze each user group individually (Chapters 4, 5) and then assess their differences and similarities (Chapter 6) in order to accommodate tremor-induced impairments in a unifying manner (Chapter 7).

<table>
<tr><th></th><th>Author(s)</th><th>Adaptability</th><th>Source</th><th>Context</th><th>Effect</th><th>Ability range</th></tr>
<tr><td rowspan="5">Physical Keyboards</td><td>Dynamic Keyboard
(Trewin, 1997)</td><td>Adaptable</td><td>In-the-wild typing data</td><td>No</td><td>Visible</td><td>MI</td></tr>
<tr><td>Invisible Keyguard
(Trewin, 2002)</td><td>Adaptive</td><td>In-the-wild typing data</td><td>No</td><td>Invisible</td><td>MI</td></tr>
<tr><td>IDA
(Koester, 2007)</td><td>Adaptable</td><td>Controlled typing data</td><td>No</td><td>Visible</td><td>MI</td></tr>
<tr><td>TrueKeys
(Kane, 2008)</td><td>Adaptive</td><td>Controlled typing data</td><td>No</td><td>Visible</td><td>MI; AB</td></tr>
<tr><td>EdgeWrite
(Wobbrock, 2007)</td><td>Adaptive</td><td>Controlled gestural data</td><td>No</td><td>Invisible</td><td>MI</td></tr>
<tr><td rowspan="5">Mouse Pointing</td><td>Steady Clicks
(Trewin, 2006)</td><td>none</td><td>NA</td><td>No</td><td>Visible</td><td>MI; Older</td></tr>
<tr><td>Mouse Gain
(Koester, 2005)</td><td>Adaptable</td><td>Controlled pointing data</td><td>No</td><td>Visible</td><td>MI</td></tr>
<tr><td>Automatic Mouse Pointing Assessment
(Hurst, 2008)</td><td>none</td><td>In-the-wild pointing data</td><td>No</td><td>NA</td><td>MI; Older ; Parkinson's disease; AB</td></tr>
<tr><td>Angle Mouse
(Wobbrock, 2009)</td><td>Adaptive</td><td>Controlled pointing data</td><td>No</td><td>Invisible</td><td>MI; AB</td></tr>
<tr><td>SUPPLE
(Gajos, 2007)</td><td>Adaptive</td><td>Controlled pointing, clicking, dragging data</td><td>No</td><td>Visible</td><td>MI; VI</td></tr>
<tr><td rowspan="6">Touch Input</td><td>Blind Touch Input
(Oliveira, 2012)</td><td>Adaptable</td><td>Clinical assessment</td><td>No</td><td>Visible</td><td>VI</td></tr>
<tr><td>Personalized Input
(Findlater, 2012)</td><td>Adaptive</td><td>Controlled typing data</td><td>No</td><td>Invisible</td><td>AB</td></tr>
<tr><td>Technology transfer
(Yesilada,2012)</td><td>Adaptive</td><td>In-the-wild typing data</td><td>No</td><td>Visible</td><td>MI; SI</td></tr>
<tr><td>Walking UI
(Kane, 2008)</td><td>Adaptive</td><td>NA</td><td>Yes</td><td>Visible</td><td>SI</td></tr>
<tr><td>Shared-User Models
(Montague, 2012)</td><td>Adaptive</td><td>In-the-wild gestural, tapping, and reading data</td><td>No</td><td>Ability-dependent</td><td>MI; VI</td></tr>
<tr><td>Walktype
(Goel, 2012)</td><td>Adaptive</td><td>Controlled typing data</td><td>Yes</td><td>Invisible</td><td>SI</td></tr>
</table>

Table 2.1: Overall comparison of reviewed projects. In the ability range column: MI (Motor-Impaired), VI (Visually-Impaired), AB (Able-Bodied), SI (Situationally-Impaired).

2.4. Conclusion

Our survey of related work highlights the complexity of providing effective interfaces that deal with the wide range of human abilities. Indeed, abilities vary from person to person, especially when dealing with impaired people. Even people with similar health conditions can vary immensely in their abilities. Likewise, abilities are not static; they are influenced by the context in which they are exercised (Newell, 1995). These issues are relevant for Chapters 4 and 5, where we thoroughly analyze tremor impairments due to situational factors and health conditions, respectively. Moreover, prior work that focused on bridging the gap between these two user groups is scarce. A comparative analysis between HIID and SIID will be performed in Chapter 6. Finally, in Chapter 7, we propose and evaluate a solution that leverages accelerometer data to improve users' typing performance independently of their cause of impairment.

3 Preliminary Research

In this dissertation, we investigate how mobile technologies can be used by people with a wide range of abilities. Particularly, we focus on bridging the gap between health- and situationally-induced impairments and disabilities. In the previous chapter, we have reviewed approaches that aimed to provide access to technology to broader user groups (e.g. Universal Usability (Shneiderman, 2000), Inclusive Design (Newell and Gregor, 2000), Ability-based Design (Wobbrock et al., 2011)). Although these works give us a theoretical framework to work on, we still need to understand how users behave when interacting with applications' interfaces.

This chapter describes an exploratory stage of our research, where we took a holistic approach to users' abilities, investigating both motor and visual attributes. We present two preliminary studies that attempted to understand how mobile touch interfaces could be adapted to fit the needs and abilities of different user groups. First, we assess the overlap of interaction problems between motor-impaired and able-bodied users within a set of touch techniques and parameterizations. In the second experiment, we intentionally focused on visual demands and mobility settings.

Findings from both user studies allowed us to adjust our expectations and outline a research approach that focused on dealing with tremor-induced impairments, either due to health- or situationally-related factors.

3.1. Investigating Motor Demands

The spectrum of motor abilities is wide and diverse. In the last decades, efforts have been made to compensate this diversity and provide an inclusive access to technology, above all, to desktop computers. The same does not apply to mobile device accessibility, which is still in its infancy. Small devices and keys, less processing capabilities, along with the context it is designed to be used in may have been the main reasons for this lack of accessibility and overall limited understanding.

Meanwhile, mobile phone touchscreens are increasingly replacing their traditional keypad counterparts. As seen before, these interfaces present challenges for mobile accessibility: they lack both the tactile feedback and physical stability guaranteed by keypads, making it harder for people to accurately select targets. This becomes more relevant for people who experience lack of precision, such as tetraplegic users. However, touchscreens offer several advantages: they can easily display different interfaces in the same surface; the ability to directly manipulate interface elements provide a natural and engaging experience; and, it has been shown that the use of PDAs is a viable alternative to traditional input devices (i.e. mouse and keyboard), allowing the same interface to be used in different places and contexts (Myers et al., 2002). Furthermore, the high customization degree of touch screens makes them amenable to custom-tailored or adaptive solutions that better fit each user's needs. This may as well be a determinant factor for inclusive design as devices used by motor impaired people can be the same as the ones used by the able-bodied population, with slender interface tuning (Gajos et al., 2007)(Trewin, 2003)(Wobbrock et al., 2011).

However, there is no comprehensible knowledge of the values and flaws of each touch interaction technique in what concerns users' motor ability. To be able to provide flexible and customizable touch user interfaces, we first need to understand how users with dissimilar motor aptitudes cope with the different demands imposed by interaction techniques and interface parameterizations.

In this section, we present an exploratory evaluation with 15 tetraplegic and 18 able-bodied people aimed at understanding the differences and similarities between populations. We studied a set of interaction techniques (Tapping, Crossing, and Directional Gesturing - Figure 3.1) and parameterizations (Size and Position). The results show that despite the expected error rate disparity, there are clear resemblances, thus giving space for inclusive adaptive/adaptable user interfaces.

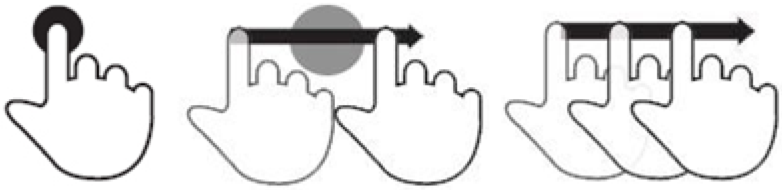


Figure 3.1: Interaction Techniques (from left to right): Tapping, Crossing, Directional Gesturing.

3.1.1. Evaluation of Touch Techniques

Touchscreen devices pose both challenges and opportunities for researchers. Recently, significant efforts have been applied to make these interfaces accessible to motor-impaired people. Wobbrock et al. (Wobbrock et al., 2003) proposed a stylus-based approach that uses edges and corners of a reduced touchscreen to enable text-entry tasks on a PDA. Similarly, Barrier Pointing (Froehlich et al., 2007) uses screen edges or 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.

Although these works insightfully explore the physical properties of the device to aid impaired people interacting with touchscreens, there is still little empirical knowledge about their performance with other interaction techniques. On the other hand, a great deal of research has been carried out to understand and maximize performance of able-bodied people using these devices (Lee and Zhai, 2009)(Park et al., 2008).

Our primary goal with this research was to evaluate different motor ability-wise participants with different interaction techniques, towards an adaptive/customizable inclusive touch design space. By understanding the limitations and needs of each population, along with the advantages and flaws of each technique and parameterization, we will be able to understand how to design interfaces that maximize each user's performance. Further, we will be able to build more inclusive interfaces.

Interaction Techniques and Variations

We considered two basic interaction methods: tapping the screen or performing a gesture. When performing a gesture, users could cross a target or just use directional gestures (Figure 3.1).

Tapping consisted in selecting a target by touching it. In this technique, targets were presented in 3 different sizes (7, 12, and 17 mm), derived from previous studies for able-bodied users (Lee and Zhai, 2009)(Park et al., 2008), and in all screen positions: edges, corners or middle, thus covering the entire surface.

Crossing, unlike Tapping, did not involve positioning one's finger inside an area. Instead, a target was selected by crossing it. Previous work, on desktop interaction, has shown that this technique offers better performance for motor-impaired users than traditional pointing methods (Wobbrock and Gajos, 2008). In our experiment, targets were shown at middle screen positions (see Figure 3.2 - left) in 3 different sizes.

Directional Gesturing was the only technique that did not require a target selection. Users could perform directional gestures anywhere on the device's surface. This technique was chosen due to its unconstrained nature and, as well as Tapping, because it is a common interaction technique in novel touch-based devices.

Table 3.1 summarizes all interaction techniques and their variations.

Technique	**Sizes**	**Positions**
Tapping	7, 12, 17 mm	Middle, Edges
Crossing	7, 12, 17 mm	Middle
Directional Gesturing	N/A	Middle, Edges

Table 3.1: Interaction techniques and variations.

Participants

Fifteen tetraplegic people were recruited from a physical rehabilitation center. The target group was composed by 13 male and 2 female participants with ages between 28 and 64 years with cervical lesions between C4 and C6. All had residual arm movement but no hand function. Regarding technological experience, all participants had a mobile phone and used it on a daily basis. However, none of them had a touchscreen mobile phone. Eighteen able-bodied participants (5 females) with ages between 20 and 45 years old were recruited using word-to-mouth in the local university. All of them had previous contact with mobile touch phones.

Apparatus

We used a QTEK 9000 PDA (Figure 3.2) running Windows Mobile 5.0. The mobile device screen had a resolution of 640x480 pixels for a size of 73x55 mm, with noticeable physical edges. The evaluation software was developed in C# using .NET Compact Framework 3.5 and Windows Mobile 5.0 SDK. The evaluation was video recorded and all interactions with the device were logged for later analysis.

Procedure

Participants were told that the overall purpose of the study was to investigate and compare different touch interaction techniques. We then conducted a questionnaire (see Appendix

Figure 3.2: QTEK 9000. Screen positions (left): white - middle; gray and black: edges. Vertical distances (right).

A2) and informed them about the experiment. All interaction techniques were explained and demonstrated.

To attenuate learning effects, participants were given warm-up trials before the evaluation of each technique. During these trials they were able to move the mobile device to a comfortable position. All sessions were performed in a quiet environment (the university, their homes or rehabilitation center facilities). Motor impaired participants carried out the trials sitting on their wheelchairs with a table or armrest in front of them. Able-bodied participants completed the trials sitting in a chair in front of a table and were free to choose how to hold the device. The interactions with the touchscreen were stylus-free; however participants were free to issue selections with any part of their hands/fingers.

Each participant was asked to perform target selections with each technique (Tapping and Crossing). For the Directional Gesturing condition, there were no targets and participants only had to perform a gesture in a particular direction (e.g. north). There were sixteen possible directions, including diagonals and repeated directions with edge support (e.g. north using the right edge as a guideline). For the Tapping condition participants were asked to select targets in all screen positions, as shown in Figure 3.2 - left. For the Crossing condition we only used the middle area (9 positions).

Participants were not informed as to whether the selection was successful or not. However, they received feedback that an action was performed. The next target appeared after each action with a two second delay. We selected tests in a random order to avoid bias associated with experience. In each method-size experience, target positions were also prompted randomly to counteract order effects.

The experiment varied interaction technique, target size and screen position. We used a within-subjects design, where each participant tested all conditions. For the position analysis, we created one extra factor, Vertical Distance (Figure 3.2 - left), which reflects the target position in relation to the users' support (level 1 refers to the closest position to participants' arm support while level 5 refers to the most distant ones).

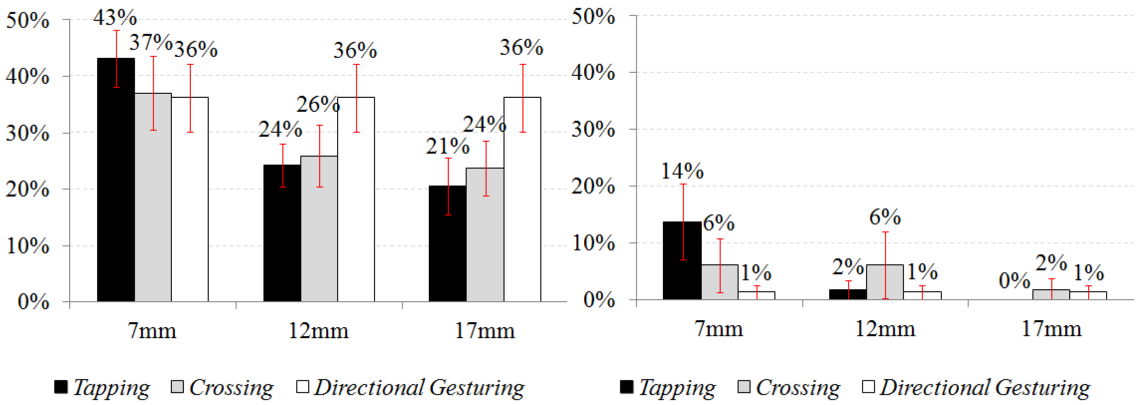


Figure 3.3: Error Rate for each Technique and Target Size (left: motor impaired, right: able-bodied). Error bars denote 95% confidence intervals.

Shapiro-Wilkinson tests of the observed values for Task Errors did not fit a normal distribution for able-bodied participants and all interaction techniques. Therefore, a Friedman test was used in further analysis. Post-hoc tests were performed using Wilcoxon signed rank pair-wise comparisons with a Bonferroni correction. On the other hand, observed values for Task Errors for motor impaired participants showed to fit a normal distribution. Thus, a repeated-measures ANOVA was used in the analysis.

3.1.2. Results

Our goal was to understand and relate the capabilities of both user populations (i.e. motor-impaired and able-bodied) when using different touch techniques. We present the results highlighting their main similarities and differences considering each technique, target size and interaction area. This knowledge will enable designers to predict how both motor-impaired and able-bodied users will perform using their touch interfaces and employed techniques.

Target Size

Motor Impaired. Considering Tapping, there was a significant effect of *Target Size* on *Task Errors* ($F_{1,42}$=25.10, p<.001). A multiple comparisons post-hoc test found significant differences between small and medium sizes, as well as between small and large sizes (Figure 3.3 - left). These results suggest 12 mm as an approximate suitable value for targets to be tapped by motor-impaired users. Regarding Crossing, a significant effect was also found ($F_{1,42}$=6.56, p<.01), however between the smallest and largest sizes.

A comparison between techniques revealed a significant effect in the medium ($F_{1,56}$=8.04, p<.001) and largest ($F_{1,56}$=3.83, p<.05) sizes, in which Directional Gesturing perform worst than Tapping and Crossing. This suggests that Directional Gestures are only worth considering when target size is small.

Able-bodied. There was a statistically significant difference in *Task Errors* depending on *Target Size* for Tapping ($\chi^2_{(2)}$=26.261, p<.001). Post-hoc analysis revealed significant differences between the smallest and both medium and largest sizes (Figure 3.3 - right). As for motor impaired participants, results suggest that Error Rate starts to converge at 12 mm. Regarding Crossing, no significant differences were found between target sizes.

Additionally, we found a significant effect of *Interaction Technique* in the smallest size ($\chi^2_{(2)}$=13.765, p<.001). Further analysis revealed that Directional Gestures is significantly more accurate than Tapping (Z=-3.237, p<.001), which suggests that when users are faced with small targets, gesture approaches are more adequate.

Differences and Similarities. Regarding each interaction technique, Tapping seems to be the most similar between user populations. Particularly, both perform worse with small target sizes (7 mm), and Error Rate begins to converge at 12mm. Nevertheless, we suspect that able-bodied users can achieve similar accuracy results with smaller targets (Lee and Zhai, 2009)(Park et al., 2008).

The main difference between these two types of users lies in the magnitude of errors, particularly in the Directional Gesturing technique. Motor-impaired users have great difficulty performing gestures in a specific directions, especially diagonals, while able-bodied users have no difficulty using this technique (see Figure 3.4). Indeed, results suggest that gesture approaches, either Directional Gesturing or Crossing, can be used as suitable alternatives to Tapping when the interface only has small targets.

Screen Area

In this section we will analyze participants' performance according to different interaction areas: edges, middle, and vertical distance.

Motor Impaired. Considering Tapping and interaction on edges, there were no significant differences on Task Errors, regardless of target size. Similar results were obtained for Directional Gesturing, as no significant differences were found on Task Errors for gestures supported by edges or anywhere else onscreen.

Regarding *Vertical Distance*, we found a significant effect for *Tapping* on medium ($F_{1,42}$=3.59, p<.05) and largest ($F_{1,42}$=5.19, **p**<.05) sizes. Post-hoc tests showed that targets closer to users' arm are easier to tap (Figure 3.5 - left). For Crossing and Gestures, no significant effect was found between vertical areas.

When comparing all interaction techniques in common ground, i.e. on the middle of the screen, there was no significant effect on Task Errors, suggesting that users have similar accuracy while interacting in the middle of the screen with Tapping, Crossing and

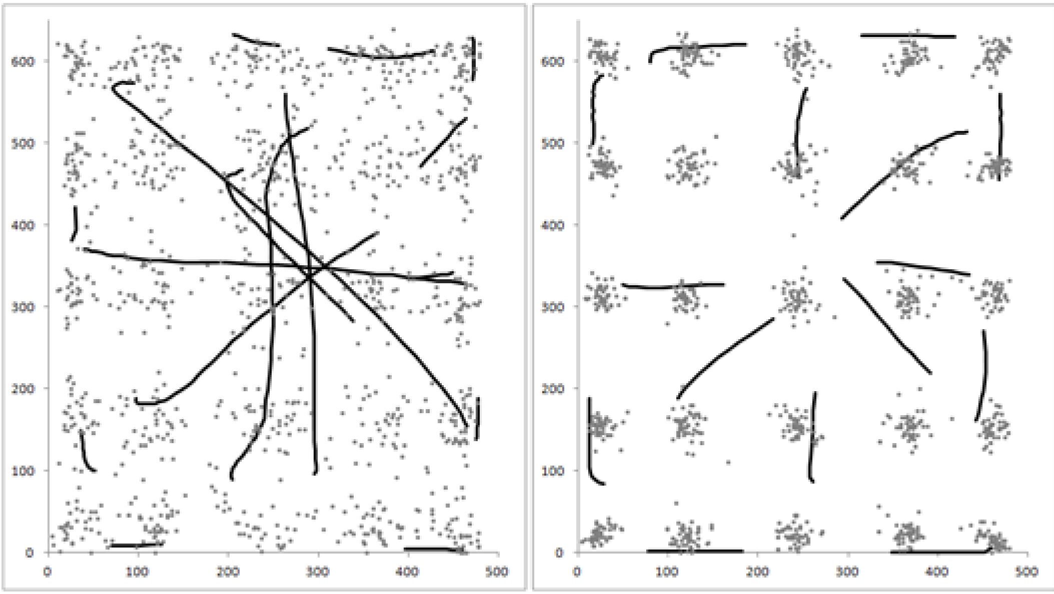


Figure 3.4: Overall Taps and Directional Gestures - left: motor impaired participant, right: able-bodied participant. Tapping dispersion is much higher for the disabled participants and Gestures are longer and more erroneous.

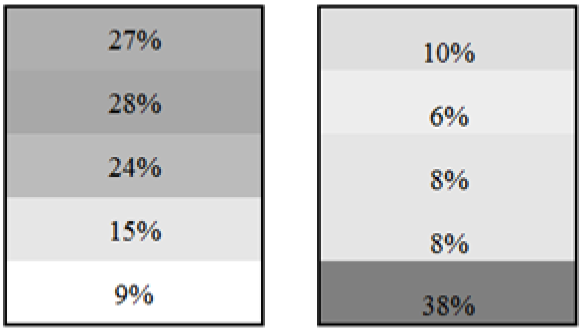


Figure 3.5: Error Rate by Vertical Distance: left - motor impaired (large size); right - able-bodied (small size)

Directional Gesturing. In this analysis we have discarded diagonal gestures as they were seen as drastically decreasing the success of Directional Gesturing approach for motor impaired users.

Able-bodied. Considering edges, we found a significant difference on *Task Errors* for *Tapping* (Z=-2.987, p<.05). Results showed that for small sizes, targets are easier to acquire on the middle of the screen. When considering *Directional Gesturing*, no significant differences where found between gestures on the edge or elsewhere on the screen. Moreover, this technique is shown to be more accurate than *Tapping* on screen edges (Z=-3.066, p<.05) for small target sizes.

Regarding *Vertical Distance*, there was a significant effect for *Tapping* in the smallest size ($\chi^2_{(4)}$=24.172, p<.001). A post-hoc analysis showed that targets near the bottom edge are significantly harder to acquire (Figure 3.5 - right). This result was not observed in

Crossing, since it had similar accuracies for all vertical areas.

Considering interaction in the middle of the screen, regardless of target size, we found no significant differences, which suggest that users have similar accuracy with Tapping, Crossing and Directional Gesturing.

Differences and Similarities. When considering interaction on the middle of the screen, both target populations perform equally with Tapping, Crossing, and Directional Gestures, suggesting that the main differences between interaction techniques are in the remaining of the screen (i.e. edges). Indeed, one could argue that Gestures performed with edge support would be more accurate to both able-bodied and disabled users. However, results have shown that both user populations have similar Error Rates performing a Gesture on the edge or anywhere else on the screen. Nonetheless, for able-bodied users, performing Directional Gestures on edges is significantly easier than Tapping small targets. In fact, when these were placed near edges, particularly the lower edge, Tapping accuracy was 3 times lower (18% Error Rate). On the other hand, motor impaired users had greater difficulties Tapping targets in the upper edge, due to restrictions of reach.

3.1.3. Towards Inclusive Design

After analyzing the results for both user populations regarding their performance with each interaction technique, we are now able to draw some conclusions about their main differences and similarities as it was proposed in the beginning of this study:

Use Tapping and Crossing as inclusive interaction techniques. Taking into account all interaction techniques, Tapping and Crossing have shown to be the ones with more resemblances between motor impaired and able-bodied users. These techniques presented low and very similar Error Rate within both target populations and, therefore can both be used in touch interfaces.

Error Rate starts to converge between 7mm and 12mm for Tapping. Tapping, the traditional selection method, was shown to be one of the most accurate Interaction Techniques. Moreover, 12 mm was revealed to be a good compromise for target size as Error Rate begins to converge for both user populations. Nevertheless, previous research suggests that able-bodied users can select smaller targets (between 7mm and 12mm) with similar accuracy (Park et al., 2008).

Keep in mind the magnitude of errors. Despite some similarities between motor impaired and able-bodied users' experiences with touch interfaces, one of the main differences resides in the magnitude of errors. As expected disabled users have a much lower accuracy rate. Overall, error rates are 5.6%, 6.1%, and 26.1% times higher for Tapping, Crossing and Directional Gesturing, respectively. Therefore, we believe that touch interfaces are still in need for new and more inclusive interaction techniques.

Avoid Directional Gesturing for motor impaired users. Directional Gestures have shown to be significantly more inaccurate than both Crossing and Tapping with 12mm and 17mm targets. Even when considering small targets, Gestures do not outperform any of the remaining techniques, thus showing no gain in its usage.

Middle of the screen consistency. Both user populations can use all interaction techniques on the middle of the screen with similar accuracy. Neither for able-bodied or motor impaired users was found a significant effect of Interaction Technique on Error Rate when using the center area of the display. This suggests that is the remaining of the screen (edges) that can favor or hinder interaction.

Take reach restrictions into account. One major difference between target populations is their ability to reach far-away targets. Motor impaired users have greater difficulties Tapping targets far from their arms' support, thus resulting in lower accuracy rate. Conversely, able-bodied users do not face this difficulty, however when targets are small they present some difficulties in Tapping targets near the bottom edge. This may be due to the restrictions imposed by the physical edges, preventing users from fully landing their fingers on targets.

Directional Gestures are a suitable alternative to small targets (only) for able-bodied users. Directional Gesturing proved to be an accurate interaction technique for able-bodied users. In fact, this technique has shown to be a suitable alternative to Tapping, particularly when small targets are placed near the edges. Unlike motor impaired people, who have many difficulties performing specific gestures, able-bodied people can easily take advantage of this technique.

3.1.4. Summary

Touchscreen mobile devices are able to exhibit different interfaces in the same display, allowing designers to create interfaces more suitable to their users' needs. These devices carry with them the promise of a new kind of user interfaces; one that is accessible to a broader user population. To fulfill this vision, we undertook an extensive evaluation with 15 tetraplegic and 18 able-bodied users in order to provide empirical knowledge to be used in the design of future touch interfaces. Our goal was to identify the main resemblances and differences between these two populations, while comparing different interaction techniques, target sizes and positions.

Results showed that traditional techniques, such as Tapping, can be used by motor impaired users, however with higher Error Rates than those obtained by able-bodied users. On the other hand, Directional Gesturing while extremely easy to perform by those with no impairments, proved to be quit inappropriate to the remaining. Crossing targets has also shown to be a suitable alternative to motor impaired, particularly if we consider the difficulty in Tapping small targets. Indeed, future touch interfaces have to take into account their users' capabilities and provide the most adequate mechanisms to ensure an efficient and effective experience.

Overall, we found several similarities between able-bodied and motor-impaired people. Nevertheless, able-bodied users experienced significantly lower error rates than motor-impaired participants. Still, the conditions studied in this work were always performed with the same device, in a controlled and quiet environment, and featuring target selection tasks. This is not representative of how mobile users interact with their devices. Mobility has been shown to induce situational impairments and decrease keying accuracy (Lin et al., 2007). In fact, previous work shows that motor-impaired and situational-impaired users' error rates start to converge (Chen, 2008)(Yesilada et al., 2010b). These findings open exciting new opportunities in the discipline of universal/inclusive design.

3.2. Investigating Visual Demands

In this user study, we explored the effect of mobility conditions, focusing on their visual demands. Indeed, operating devices in mobile environments pose new challenges to users since apparatus and context often compete for the same human resources, inducing situational impairments and disabilities (SIID) (Sears et al., 2003). For instance, texting while walking on a busy street can be quite challenging and prove hazardous since the visual system is both engaged on monitoring the surrounding environment and on interacting with the device. Similarly, reading text messages or email in public spaces can be difficult, or even impossible, due to screen glare caused by sunlight. In such situations we argue that users may become *functionally blind*, as their visual resources are overloaded and visual feedback is inadequate.

These problems become especially relevant when performing visually demanding tasks, such as text-entry. Indeed, text input is one of the most demanding tasks in mobile devices and one of the most common between applications, such as managing contacts, SMSing, emailing, note-taking, gaming, chatting, tweeting, and so forth.

In this preliminary work, we assess the effect of mobility on typing performance and explore a technology transfer approach (Yesilada et al., 2010a), where solutions initially created for blind people can be used by sighted people in mobile contexts. While we agree that both users' capabilities are different, there seems to be an overlap of interaction challenges. Focusing on text-entry tasks, we present a user study where participants used two alternative methods on three mobility conditions. Our main goal was to assess the effect of walking and visual demand on participants' performance with methods designed for blind users. We analyzed the obtained results for each method individually, and compare them with each other, in order to draw conclusions and suggestions for future work.

3.2.1. Background

The study of visual abilities is of upmost importance, since it is our primary sense to perceive real-world information. Mustonen et al. (Mustonen et al., 2004) showed that users, whilst mobile, perceive information differently, particularly, reading speed slows with increasing walking pace. Regarding the effect on cognitive resources, Oulasvirta et al. (Oulasvirta et al., 2005) performed a semi-naturalistic field study showing that visual attention is highly fragmented when interacting with mobile devices.

Over the years several solutions have been proposed in order to ease the visual demand of mobile interfaces. Pascoe et al.(Pascoe et al., 2000) proposed minimal attention user interfaces (MAUI) in order to minimize the amount of visual attention, though not necessarily the number of interactions required to operate the device. Other authors have abandoned screens entirely, allowing users to control their devices through alternative modalities (Sawhney and Schmandt, 2000).

Indeed, visual attention is a crucial resource when using devices whilst walking. Nevertheless, while these systems provide alternative interfaces, our work takes a different approach in that we try to reuse knowledge already available from users who can't use visual feedback and apply it on mobile contexts.

In mobile settings there is a competition for the users' attention between the surrounding environment and the mobile device (Oulasvirta et al., 2005). Users are constantly managing their attentional resources, switching tasks and gaze as needed. This behavior is usually aggravated due to visually demanding interfaces. Consequently, in visually demanding environments, users become *functionally blind*, as they cannot maintain performance on a given task due to an overload of their visual resources.

Solutions designed for blind people may ease the visual burden of mobile users. These include traditional screen readers, which replace visual feedback by its auditory representation. For example, when using an iPhone, VoiceOver[1] makes use of a text-to-speech tool to read interface elements touched by the user. Therefore, when using a virtual keyboard, users can navigate through letters and enter text without looking at the screen. Indeed, touch interfaces for blind people have been recently attracting a great deal of work (Yfantidis and Evreinov, 2006)(Guerreiro et al., 2008)(Bonner et al., 2010)(Frey et al., 2011)(Oliveira et al., 2011b)(Azenkot S. et al., 2012). For instance, the NavTouch (Guerreiro et al., 2008) method enables blind users to input text by performing directional gestures to navigate a vowel indexed alphabet (Figure 3.6). Gestures to right and left allow users to navigate the alphabet horizontally, while up and down allow them to jump between vowels. This technique requires no memorization beyond knowing the sequence of letters in the alphabet. Moreover, vowels can be used as shortcuts to the intended letter. Constant audio feedback reads each character to users as they select it, whereas double

[1] http://www.apple.com/accessibility/iphone/vision.html (last visited on 06/11/2012)

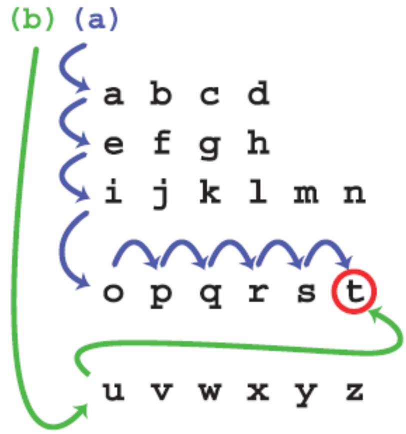


Figure 3.6: Vowel navigation (Guerreiro et al., 2008): two navigation alternatives to reach the letter 't'.

or split tap is used to accept the selection. Special actions (such as erase) are placed on screen corners.

While these methods can theoretically ease the visual demand of text-entry tasks, due to their auditory feedback, we still do not know how situationally impaired users behave on different mobility conditions. Our main goal is to provide this empirical knowledge to be used in the design of more usable mobile interfaces.

3.2.2. Evaluation of Knowledge Reuse

In this experiment, we try to reuse knowledge already available from users who cannot use visual feedback and apply it on mobile contexts. We hypothesize that mobile users become *functionally blind*, as they cannot sustain performance on a given task due to their visual system being overloaded. Therefore, in this experiment, we adopted solutions designed for those for whom graphical feedback is inappropriate (such as blind people), thus freeing some of the users' limited visual resources to their main task. According to the Multiple Resource Theory (MRT) (Wickens, 2005), this would make it easier for people to perform both tasks simultaneously with less interference and therefore with smaller performance penalty.

While we stress the similarities between blind and situationally-impaired users, we also acknowledge that the two groups' abilities are different in that SIID tend to be temporary and dynamic, as mobile users can always glance at their devices. Nevertheless, we believe that in visually demanding conditions, both populations experience the same problems, and could hence benefit from similar solutions. Therefore, perhaps a more appropriate question would be: when and how can mobile users benefit from assistive technologies? While previous research has focused on assistive technologies for motor impaired people (Yesilada et al., 2010a), visual demands are still unexplored. This user study sought to observe how users behave when using assistive technologies whilst walking.

Participants

Twenty three participants (15 male, 8 female) with ages between 18 and 37 years took part in the study. All participants owned a mobile phone, for at least five years. Only six of them did not use touchscreen technology. Regarding text-entry, two participants used it on a weekly basis, while the remaining did this task daily. As for preferred text entry methods, 15 participants used QWERTY layouts, 13 on virtual and 2 on physical keyboards, while 8 used MultitTap (2 virtual and 6 physical keypads).

Apparatus

This study used a Samsung Galaxy S device running Android 2.2 with a screen resolution of 480x800 (122.4x64.2 mm) pixels wide. We focused our research on QWERTY keyboards, since this is one of the most common mobile layouts, and picked one alternative input method. In summary, we used three text-entry methods: 1) a traditional QWERTY keyboard, used as a control condition; 2) a VoiceOver-like method (using QWERTY), since this is a common assistive technology for blind users; 3) NavTouch, because it uses a gesture approach. All text-entry methods were developed using Android SDK. In the QWERTY keyboards, letters were entered using a lift-off strategy, thus enabling participants to correct land-on errors. Speech feedback was given using SVOX Classic TTS. The evaluation was recorded on video and we logged all interactions with the device for later analysis.

Procedure

The study was conducted individually and started with a brief explanation about its overall purpose and procedure. After each participant completed a short questionnaire (see Appendix A3) to gather demographic data. All text-entry methods were explained, followed by a five minute practice trial for each method to counteract learning effects. Each participant was asked to perform two text-entry tasks using three different methods: QWERTY, VoiceOver alike (with QWERTY) and NavTouch (Guerreiro et al., 2008). Although two of the featured methods were designed for blind people, visual feedback was intentionally made available. Therefore, we could observe the participants' natural behavior when both visual and auditory modalities were present.

In order to realistically test these methods, we designed three mobility settings: 1) Control - participants were seated in a quiet and controlled environment; 2) Corridor - participants were asked to walk at their own pace in a straight path without obstacles; 3) Navigation

Figure 3.7: Participant typing whilst walking on navigation condition.

- participants had to orient themselves within an indoor track. The track featured poles with numbers and arrows indicating both the order and direction the participants had to walk along a prescribed route (similar to (Schildbach and Rukzio, 2010), see Figure 3.7). This setup was created to simulate the use of mobile devices in an urban environment. We picked mobility conditions in a random order to avoid bias associated with experience. Additionally, before testing the first mobility condition, we recorded each participant's preferred speed when walking in a straight line.

For each mobility condition, participants were asked to copy a set of sentences using all methods in a counter-balanced order. Each trial consisted of two sentences, each five words long with an average 4.48 characters per word. The sentences were extracted from a written language corpus, and each had a minimum 0.97 correlation with the language (see Apendix A1). We built the phrase set based on the procedure of MacKenzie and Soukoreff (MacKenzie and Soukoreff, 2003) applied to the Portuguese language. Each sentence was randomly selected and read aloud to participants. Also, the sentence was always visible on the device's screen in order to reduce misspelling errors.

Experimental Design and Analysis

The experiment varied both mobility condition and text-entry method. We used a within-subjects design, where each participant tested all conditions. We applied Shapiro-Wilkinson tests to observed values for words per minute, error (deleted characters) rate, minimum string distance (MSD) error rate (MacKenzie and Soukoreff, 2002b), and walking speed. However, the results did not show to fit a normal distribution. Therefore, we applied a non-parametric (Friedman) test to further analyses. For post-hoc tests, we used Wilcoxon signed rank pair-wise comparisons test.

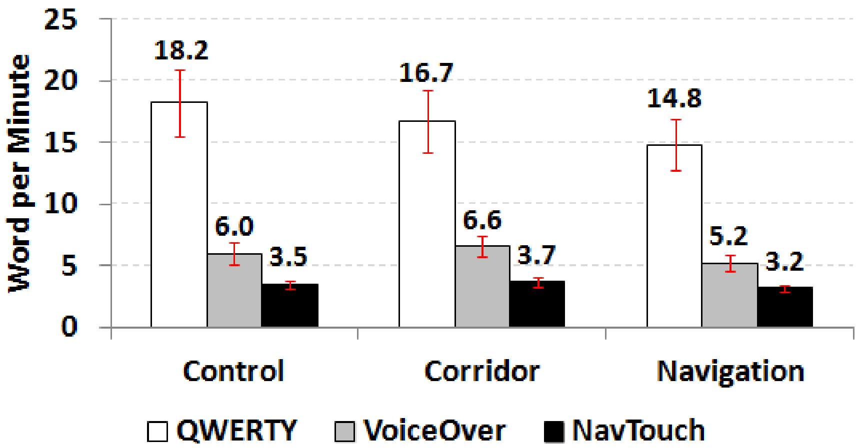


Figure 3.8: Words per minute for each condition. Error bars denote a 95% confidence interval.

3.2.3. Results

Our goal was to understand how users behave when using text-entry methods for the blind whilst on the move. In this section, we report the obtained results and analyze both *mobility* and *method* effects.

Text-Entry Speed

To analyze text-entry speed we measured words per minute (MacKenzie and Soukoreff, 2002b), calculated as:

$$(transcribed\ text\ length - 1) \times (60\ seconds \div time\ in\ seconds) \div 5\ characters\ per\ word$$

We measured the time to input each sentence from the moment the first character was entered to the last. Figure 3.8 illustrates the wpm for each condition.

Regarding the differences between text-entry methods, we found significant differences on *WPM* in the *seated* ($\chi^2_{(2)}$=96.93, p<.01), *corridor* ($\chi^2_{(2)}$=88.44, p<.01), and *navigation* ($\chi^2_{(2)}$=96.75, p<.01) conditions. A post-hoc test found significant differences between all methods. QWERTY keyboard was always faster, followed by VoiceOver and NavTouch. This result was probably due to two main reasons: QWERTY familiarity and the two-step selection process of assistive technologies. Both VoiceOver and NavTouch required navigation and confirmation actions for each letter, making these methods less efficient. As for *mobility*, we found significant differences for the *QWERTY* keyboard ($\chi^2_{(2)}$=9.92, p<.01), *VoiceOver* ($\chi^2_{(2)}$=7.06, p<.05) and *NavTouch* ($\chi^2_{(2)}$=4.7, p<.01). For the *QWERTY* keyboard, we found significant differences between the *control* (18.24 wpm) and

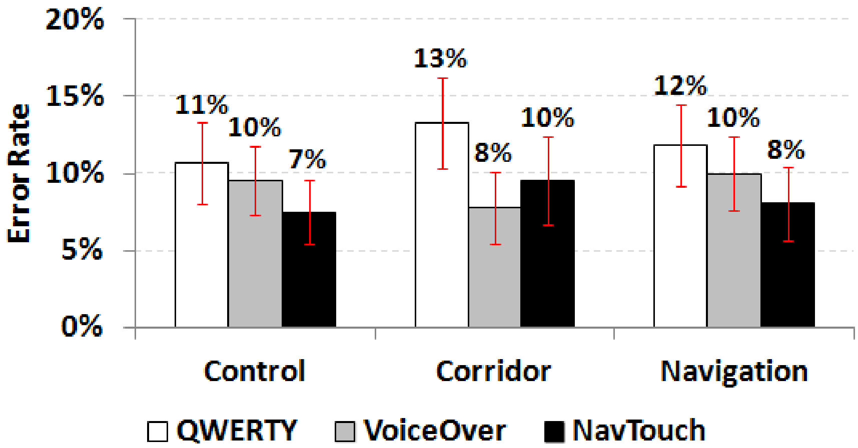


Figure 3.9: Error rate for each condition. Error bars denote a 95% confidence interval.

the *navigation* conditions (14.82 wpm); for the *VoiceOver* method we observed differences between the *corridor* (6.59 wpm) and *navigation* (5.23 wpm) conditions; as for *NavTouch* we saw differences between the *corridor* (3.68 wpm) and *navigation* (3.21 wpm) conditions.

These results suggest that all three methods were sensitive to visual demand conditions. However, assistive technologies were ineffective regarding input rate. On the other hand, the QWERTY keyboard performance varied the most with a loss of 3.42 wpm between the control and navigation conditions.

Error Rate

As a measure of effectiveness, we used error rate, calculated as:

$$(letters\ deleted \div letters\ inserted) \times 100$$

Comparing error rates between text-entry methods, we found differences in the *control* ($\chi^2_{(2)}$=4.54, p<.1), *corridor* ($\chi^2_{(2)}$=7.57, p<.05) and *navigation* ($\chi^2_{(2)}$=5.53, p<.1) conditions. After post-hoc analysis, we found that in the control and navigation situations the QWERTY keyboard had higher error rates (10.63% for the control and 11.85% for the navigation) than NavTouch (7.49% for the control and 8.07% for the navigation). In the corridor condition (see Figure 3.9) the QWERTY keyboard not only had a significantly higher error rate (13.27%) than NavTouch (9.55%), but was also higher than the VoiceOver method (7.79%). Regarding mobility, we did not find any significant effect.

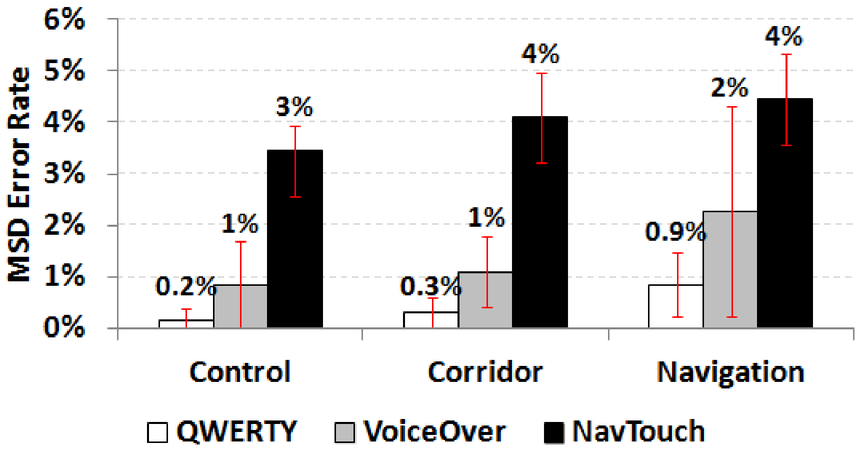


Figure 3.10: MSD error rate for each condition. Error bars denote a 95% confidence interval.

Quality of Transcribed Text

To measure the quality of transcribed text we used the Minimum String Distance Error Rate metric calculated as:

$$MSD(required\,text, transcribed\,text) \div Max(required\,text, transcribed\,text) \times 100$$

Concerning the effect of text-entry *method*, we obtained significant differences for the *control* ($\chi^2_{(2)}$=93.23, p<.01), *corridor* ($\chi^2_{(2)}$=73.51, p<.01) and *navigation* ($\chi^2_{(2)}$=64.77, p<.01) conditions. Overall, NavTouch produced the worst text quality in all mobility conditions (Figure 3.10). No significant differences were found between the VoiceOver and QWERTY keyboards. A detailed analysis on transcribed sentences revealed that most participants usually entered the letters correctly when using NavTouch; however, they forgot to double tap to insert white spaces between words, resulting in a MSD error rate around 4%. A possible explanation to this behavior may be the lack of practice.

Regarding the effect of *mobility*, we found a significant difference for the *QWERTY* method. After applying the post-hoc test we found significant differences between the navigation (0.85%) and control (0.16%) conditions, suggesting that the QWERTY keyboard is the most sensitive to visually demanding contexts.

Walking Speed

To measure walking speed we used the speed rate calculated as:

$$(Speed\,during\,trial \div Preferred\,walking\,speed) \times 100$$

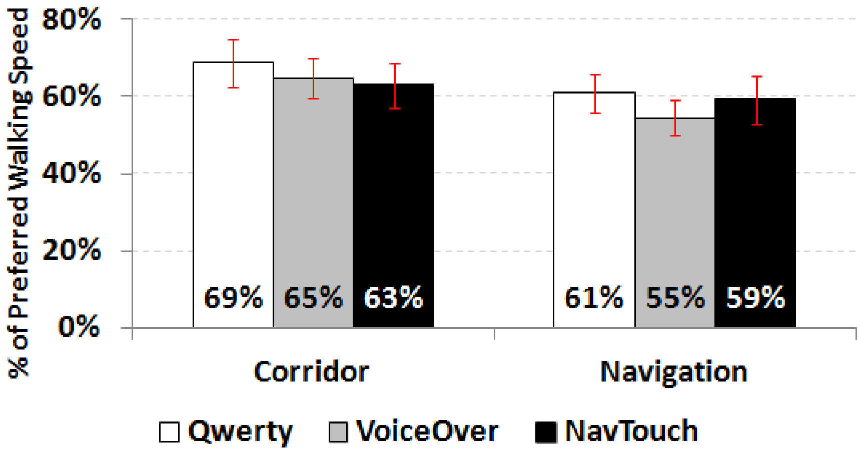


Figure 3.11: Percentage of preferred walking speed for each mobility condition. Error bars denote a 95% confidence interval.

Figure 3.11 shows the average decrease of preferred walking speed for each condition.

We found an effect of *method* on *walking speed* in the *corridor* ($\chi^2_{(2)}$=4.06, p<.1) and *navigation* ($\chi^2_{(2)}$=13.38, p<.01) conditions. In the corridor conditions differences were found between QWERTY (68.77%) and VoiceOver (64.54%), while in the navigation condition QWERTY was the method that allowed the fastest walking speed (60.86%). NavTouch came next (59.25%), followed by VoiceOver (54.53%).

As for *mobility*, users walked significantly faster in the corridor than in the navigation conditions for all text-entry methods: QWERTY decreased from 68.77% to 60.86%, VoiceOver decreased from 64.54% to 54.53%, and NavTouch decreased from 62.92% to 59.25%. These results suggest that the navigation course was more demanding, and therefore participants needed to decrease walking speed to compensate mobility challenges.

3.2.4. Lessons Learned

People reduce speed to compensate for task demands. Results show that users compensate the visual demand of contexts by naturally reducing walking speed. This was also observed in previous research (Mizobuchi et al., 2005)(Lin et al., 2007)(Bergstrom-Lehtovirta et al., 2011).

Users overlook audio feedback. Although results show the QWERTY keyboard as the most sensitive to mobility conditions, it still outperformed the remaining methods, both speed and text-quality wise. This suggests that audio-based methods are ineffective, at least, when visual feedback is available. Indeed, when debriefing participants they stated their preference to use the graphical interface and tended to dismiss audio feedback.

Assistive technologies are slow. Since participants continued to use their vision to interact with NavTouch and VoiceOver, the two-step, navigation and confirmation process needed for every character seemingly increased workload and consequently decreased performance. This suggests that modifications may be needed to effectively transfer solutions between health and situationally impaired user groups.

Mobility conditions were not demanding. Another reason for QWERTY outperform the remaining methods may lie in that our mobility conditions were not demanding enough to require users to stop looking at the mobile interface. Further research should focus on more demanding settings.

Experimental procedure. One of the main challenges when evaluating mobile users is guaranteeing consistency between participants. Although our conditions were controlled, we found large variations in both efficiency and effectiveness measurements between participants. While this may be due to individual differences, we believe that other factors may be involved. For instance, participants had different gaze behaviors, which can affect performance. Similarly, walking speed can also compensate for visual demands, thus introducing a lack of consistency between participants and text-entry conditions. Even though solutions should be evaluated in mobility settings in order to capture realistic data, performance should also be assessed in more controlled conditions (Lin et al., 2007).

3.2.5. Summary

In this study, we proposed the use of text-entry solutions for blind people on mobile contexts to reduce the visual demand of mobile interfaces and allowing situationally impaired users to maintain their performance on mobility tasks. Our experiment showed that users compensate for the challenges of mobility conditions by reducing walking speed. Moreover, the QWERTY keyboard outperformed the remaining methods, both speed and text-quality wise. This suggests that audio-based assistive technologies are ineffective, when visual feedback is available. Since users continued to use their vision to interact with NavTouch and VoiceOver, the two steps, selection and confirmation, needed for every key, seemed to increase workload and, consequently, decrease performance. Indeed, when debriefing participants they stated a clear preference for the graphical interface and tend to overlook audio feedback. Still, the QWERTY keyboard performance was the most affected by mobility conditions. We believe these errors being caused by general hand oscillations, which affected target selection accuracy.

3.3. Lessons Learned and Implications from Preliminary Research

Results from both user studies, described in this chapter, improved our understanding on three major topics: 1) differences and similarities between tetraplegic and able-bodied users; 2) effect of mobility and visual demand on overall typing accuracy; and 3) effect of applying solutions designed for blind people on sighted users in mobile contexts.

The first experiment called our attention to the similarities between tetraplegic and able-bodied users. Tapping, a traditional interaction technique, was among the most effective techniques for both target populations. This result suggests that the simple *touch-to-select metaphor is adequate for a wide range of abilities*. Nevertheless, we found some significant differences between user groups. For instance, motor impaired users showed many difficulties when selecting targets far from their arms' support. Additionally, performing gestures in some directions, namely diagonal gestures, was shown to be overwhelming. In general, although tetraplegic users experience some form of tremor, their *main difficulties were closely related with reach restrictions*. This was a valuable lesson for the remainder of our research. Finally, the difference in magnitude of errors between user groups was considerable. This revealed that if we wanted to design solutions that would have a significant effect on users' performance, we would need more *demanding conditions for able-bodied people* (i.e. SIID). Thus, we started to explore mobile settings.

In the second user study, in addition to exploring mobile environments, we also shifted our attention to visual demands in order to understand whether we could find similarities with blind users. Results were not satisfactory. We found that we were dealing with very different abilities; that is, blind people experience an impairment on a perceptual level, while mobile users, in navigation tasks, were attentional-impaired (i.e. cognitive ability).

Although attentional resources are important in many real-world situations, such as walking in a busy-street whilst texting, we found that users' (solely) walking motion was affecting typing performance. Participant struggled to maintain typing accuracy whilst walking in an obstacle-free course. This result suggests that even when users' attention is not stressed, *hand tremor has a significant effect on users' performance*, which was later reported by (Bergstrom-Lehtovirta et al., 2011).

3.4. Research Approach: A Unified View on Users' Abilities

In this dissertation, we hypothesize that hand movement decreases keying accuracy. Situational constraints, as well as health conditions, are two common settings where this effect can be observed.

Preliminary results, reported in Section 3.1, revealed some similarities between tetraplegic users and able-bodied within a set of touch techniques. Although these users experience some form of hand tremor, their dexterity and reach restrictions also play a significant role in target selection tasks. In order to assess the effect of hand motion on users' performance in an isolate manner, we investigated physiological tremor. This type of tremor occurs to everyone, but is especially visible in older adults, who experience increased physiological tremor (Van Den Eeden et al., 2003)(Strickland and Bertoni, 2004)(Benito-Leon et al., 2003). Thus, in the remainder of this dissertation, we investigate how mobile solutions can be designed for users with tremor-related impairments, due to either situational constraints or health conditions.

Bridge the Gap between Situationally- and Health-Induced Impairments

Our research approach views each user as a set of abilities, which can vary within a limited range. These abilities can be affected by either (or both) situational restrictions or health conditions. This framework of thinking allows us to abstract from the cause of impairment and focus on the effect on users' performance. More importantly, it enables us to bridge the gap between these two groups by designing mobile solutions that can fit users' abilities whether they are being affected by SIID or HIID. Bridging the gap between these user groups brings several advantages to the research community:

Avoid the duplication of work. Technology is constantly and rapidly evolving. New devices, products, and ideas are announced every day. Due to the dynamic nature of our research field, it is very likely that several authors do know about related or even overlapping works created by others. This leads to the well known phenomenon of "reinventing the wheel". Raising awareness and showing a relationship between SIID and HIID will mitigate the duplication of work.

Promote the reuse of knowledge. In addition to avoid doing work already created by others, this approach provides the means to a higher and more effective reuse of knowledge. Particularly, if common issues are to be found between users with situationally- and health-induced impairments, then similar solutions can be applied to both groups. This enables the unification of several domains, such as accessibility, mobile human-computer interaction, and general HCI, causing researchers to discuss common challenges as well as opportunities.

Leverage research. Building a relationship between HIID and SIID will create a "broader and better" research community. A larger number of researchers, with distinct backgrounds and skills, will enable the creation of interdisciplinary solutions. Research will be guided by common problems; however, seen through different perspectives, which will reduce the time taken to address these issues and leverage the quality of proposed solutions.

Reduce costs and increase availability. One of the main problems of solutions especially designed to users with health-impairments and disabilities is their target market. The cost that is inherent to these products is typically high and supported by end-users. By allowing solutions to be used by broader user populations, they will be available as "mainstream" products resulting in a decrease of prices.

Remove the negative connotation of the word accessibility. The field of accessibility or accessible computing is sometimes the target of negative connotations. It is usually seen as a very specific domain in the HCI field, which only addresses the problems of those with very specific motor or cognitive impairments. The introduction of situational impairments brings the opportunity to demystify this myth and build solutions to a wide range of abilities. Accessible computing is *NOT* a research field to the minorities, because at some point in our lives, we all experience some form of impairment or disability.

Concluding, although this approach can be applied to several human abilities, our research work focuses on tremor-related impairments. Particularly, we investigate the effect of walking and increased physiological tremor on typing performance, using hand motion as a unifying measure.

Assessment of Abilities

A common need of any successful accessible system is the assessment of users' abilities. Since we want to design common solutions to both users with SIID and HIID, we need to objectively assess users' abilities on common ground; that is, using the same "ability-scale". Clinical tests or non-computer oriented procedure are usually subjective/open to interpretation or difficult to generalize and relate with users' performance. Moreover, due to the dynamic and ephemeral nature of SIID, these approaches are inadequate.

Instead, we take an ability-based approach (Wobbrock et al., 2011) by characterizing users' abilities through their performance with technology. Particularly, in Chapters 4 and 5, we focus most of our analysis on accuracy results. We believe that a thorough understanding of issues and challenges encountered when using current devices is a valid proxy to characterize users' abilities. Three of the most commonly observed difficulties in typing are (MacKenzie and Soukoreff, 2002a):

Insertion errors. These occur when a key is unintentionally pressed resulting in additional characters. Users often press adjacent keys in addition to the intended key, especially if targets are small and no physical affordances are available. These errors can also occur when a key is unintentionally pressed more than once.

Substitutions. Users can miss the intended key and inadvertently press an adjacent one, resulting in an incorrect character. Similarly, the users' fingers can slip on the screen during a key press action and the wrong letter is inserted.

Omissions. These errors occur when users omit characters. A wrong mental model of words or forgetfulness can be the cause of this type of error. On the other hand, the device's inability of recognize a key press can also originate omission errors.

The proposed approach to measure users' abilities, as they interact with technology, has the advantage that it can be used in real-time and on-the-go. Additionally, results can inform interface design directly, since measurement takes place with the current interface. More importantly, it enables a true and fair comparison between SIID and HIID, as abilities are objectively and independently assessed of external factors (e.g. demographic profile, experience, and so on). Chapter 6 presents a comparative analysis between user groups, where their main differences and similarities are analyzed.

Context-Sensing

We believe that the main cause of disability of both health and situational impaired users is hand tremor. The overall movement of the device introduces new interaction challenges that decrease users' typing performance. Sensing capabilities of current mobile devices have the potential to measure this movement and counteract its effect.

Taking advantage of the tri-axis accelerometer to compensate for hand tremor is a very promising research opportunity and still in its early stages. This information can be used as a common measure of tremor for both user groups, enabling the development of broader and cause-agnostic solutions.

Moreover, leveraging the mobile device's accelerometer is also advantageous in order to deal with the dynamic nature and variability of SIID. Motion data can be used to predict the users' abilities in a multitude of situations. Together with touch information, context-sensing can provide an effective way to compensate entry errors and enhance typing accuracy.

Modeling Users' Abilities

After assessing users' performance and sensing the context, we need to model users' abilities. This is a particularly challenging task if we take into account the general variability of human behavior. This phenomenon is even worse when considering users with impairments. Two persons, even with the same health condition, can differ drastically on their abilities. In our case, we can only expect a higher variability since we are dealing with situationally- and health-induced impairments. We mitigate this effect by assessing users' abilities based on their performance with technology, thus no generalizations are needed to inform interface design/adaptation.

Regarding the modeling of users' abilities, in Chapter 7, we propose and evaluate different approaches. First, we describe their abilities considering the tremor-related errors they perform when typing: insertion and substitution errors. We propose the use of different typing features for each type of error in order to characterize their behavior. Second, the modeling itself is done by recurring to machine learning techniques and using real typing data. We opted to use machine learning algorithms, since they are typically more effective in unveiling hidden patterns in large amounts of data than simple decision procedures or heuristics.

Finally, users' abilities are modeled through the combination of different sub-models, each characterizing one type of error. Both touch and motion information is used to describe users' typing abilities.

Application of Knowledge

After assessing and modeling users' abilities, our goal is to apply that knowledge to improve typing accuracy of both user groups. All built and "learned" models are evaluated in Chapter 7 in order to maximize the gain of our final solution. We assess and compare the models' ability to describe users' typing skills by analyzing their accuracy results.

Finally, we perform a simulation using previously collected data in order to assess the effectiveness of our solution. We decided to apply our adaptations in a non-visual manner; that is, individual keys and keyboard adaptations are hidden from users. This approach has shown to prevent confusion and reduce cognitive load relatively to a visual-adaptive interface (Findlater and Wobbrock, 2012).

3.5. Conclusion

This chapter is an introduction to the remainder of the dissertation. It describes an exploratory stage of our work. We present two preliminary user studies that aimed to understand how touch interfaces could address the needs of users with different motor and visual abilities. From obtained results, we took several lessons and implications that had a direct influence on how we conducted the remainder of our work, allowing us to focus and adjust our goals.

We finish the chapter by detailing the research approach followed in this dissertation and how we addressed the challenges previously identified in the related work chapter (see Section 2.2.1).

4

Understanding Situationally-Induced Impairments and Disabilities

In this chapter, we present a thorough analysis of the effect of walking on typing performance. This corresponds to the first step of our research approach: characterizing mobile users' abilities. As seen in preliminary user studies (Chapter 3) and previous projects (Schildbach and Rukzio, 2010), walking conditions appear to induce motor-impairments, particularly hand oscillations (Bergstrom-Lehtovirta et al., 2011), which in turn decrease keying accuracy.

We first provide an overview of previous works that attempted to understand and deal with situationally-induced impairments and disabilities. Next, we describe our experimental protocol and report the obtained results, characterizing users' abilities through types of errors. Finally, we discuss the results and provide design implications for future work.

4.1. Background

In this section, we present and discuss previous research on mobile interaction. Particularly, we focus on understanding the challenges of mobile usage and proposed solutions to improve user performance in walking conditions.

4.1.1. Effect of Walking on Users' Performance

In a pioneer work, Kristoffersen and Ljungberg (Kristoffersen and Ljungberg, 1999) stated that mobile devices usually compete for the same human resources required for other mobility tasks. Since then, several empirical studies have tried to understand how users are affected by different mobility conditions. In particular, much work delved in walking scenarios, as this is a common activity.

Barnard et al. (Barnard et al., 2007) evaluated reading comprehension and word search tasks while walking under different lighting conditions. They found that contextual variations, such as light intensity and mobility lead to changes in user behavior and increased task times. Mustonen et al. (Mustonen et al., 2004) performed a similar user study, concluding that reading speed is significantly affected by mobility. Bergstrom-Lehtovirta et al. (Bergstrom-Lehtovirta et al., 2011) investigated how walking speed correlates to target acquisition performance, showing that to maintain selection accuracy, users need to reduce speed by 26%, as compared to their preferred walking pace. Similar results were found by (Barnard et al., 2007), reporting that users spontaneously reduce speed by 30-37% when performing reading tasks in a mobile device.

Lin and colleagues (Lin et al., 2007) carried out a Fitts' law experiment of stylus tapping whilst walking (Figure 4.1 - left) and found that the time to complete target selection tasks did not increase, however users compensated by reducing their walking speed and perceived an increased workload. Moreover, error rate increased. Schedlbauer and Heines (Schedlbauer and Heines, 2007) compared stylus-based target selection while seating, standing, and walking (at a normal pace) and found Fitts' Law to be valid under all conditions.

Brewster (Brewster, 2002) researched the use of soft keyboards under two different mobility conditions to assess the differences between realistic scenarios and laboratory settings: seated in a lab and walking outside. Results showed that walking significantly reduced usability (less data entered and increased perceived workload), indicating some difficulties associated with stylus-based tapping while walking. It is important to notice that the observed difficulties and decreased performance are likely to be even greater in a real mobile setting as the authors stated that the study scenario was still quite controlled ("reasonably quiet straight path").

Others (Chen, 2008)(Yesilada et al., 2010b) have shown that small devices, in particular

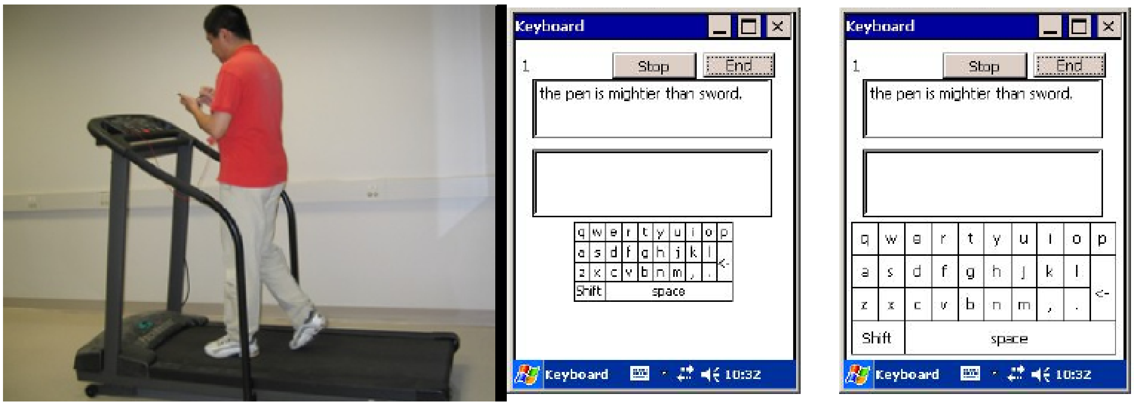


Figure 4.1: Left - Participant walking on the treadmill while doing target selection task (Lin et al., 2007); Right - Keyboard sizes as they appeared in (Mizobuchi et al., 2005).

QWERTY keypad-based devices, pose similar problems to non-impaired users as those felt by motor impaired users in a desktop setting, even in stationary seating settings.

Mizobuchi et al. (Mizobuchi et al., 2005) studied stylus text input (Figure 4.1 - right) and tried to reveal a relationship between walking speed and task difficulty. The authors found that text-input was slower whilst walking and observed that users generally decrease walking speed while typing. However, they found no relationship between these two variables. Additionally, error rates tended to increase and an effect of target size emerged: larger targets allow higher input rates and lower error rates. Although the authors also analyze the effect of walking and key size in text input, a key difference to our work lays in the fact that in their study a stylus was used and therefore, only two-handed interaction was analyzed. In our study, in addition to target size, we also considered both one-hand and two-hand interaction, in which participants use their thumbs to input text. In fact, this is currently one of the most commonly used techniques with touchscreen devices.

Interaction with thumbs has been studied in mobile target selection tasks (Schildbach and Rukzio, 2010), and static conditions (Parhi et al., 2006), yet there is a lack of knowledge pertaining text-entry tasks whilst mobile. The study presented in this chapter tries to bridge this gap by analyzing the effect of walking on text-entry performance, and secondly to understand the effects of two-hand interaction and target size. We decided to investigate both these effects to see if mobility challenges could be compensated by hand posture.

4.1.2. User Interfaces for Walking

Although we specifically focus on motor-impairments that originate from mobility demands, research on visual attention (Pascoe et al., 2000)(Oulasvirta et al., 2005)(Schildbach and Rukzio, 2010) and eyes-free techniques (Lumsden and Brewster, 2003)(Brewster et al., 2003)(Zhao et al., 2007)(Li et al., 2008)(Sawhney and Schmandt, 2000) are ac-

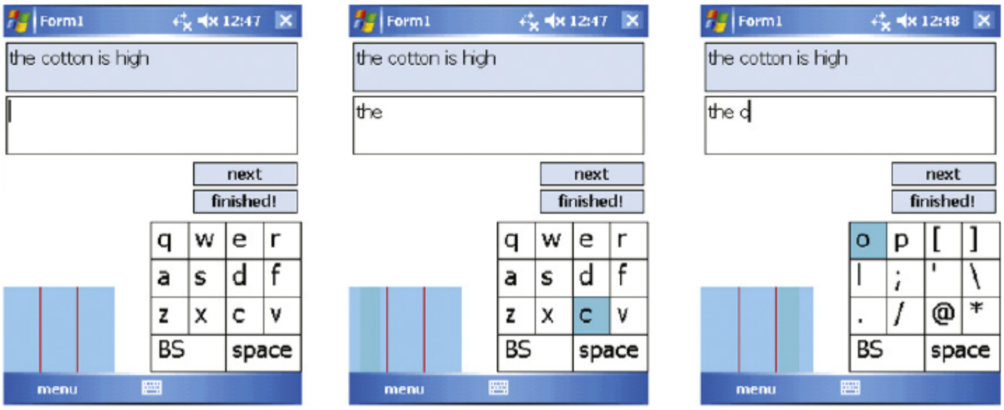


Figure 4.2: Examples of the two-handed chorded keyboard (Yatani and Truong, 2009) for right-handed users.

tive fields of work. Still, the projects reviewed in this sub-section focus on motor-related impairments.

Pirhonen and colleagues (Pirhonen et al., 2002) developed a music player controlled by directional gestures on the device screen. This solution was evaluated whilst walking, showing high performance improvements in comparison to a traditional GUI interface. More recently, Pielot et al. (Pielot et al., 2012) proposed a non-visual menu to be used in-pocket. By laying out menu items along the edges of screen users are guided by tactile features. Vibrotactile feedback and speech is used to give feedback. An evaluation "on the go" showed significant improvements over Apple's VoiceOver regarding efficiency, accuracy, and preference.

Yatani and Truong (Yatani and Truong, 2009) presented a new input technique taking advantage of users' thumb (Figure 4.2). A QWERTY chord-keyboard was created in order to increase target size and, consequently, decrease error rates. The authors evaluated this keyboard in an obstacle course and whilst walking up/down stairs, showing that the traditional keyboard has a clear advantage on typing speed, but the novel method allows lower error rates and greater consistency between static and mobile conditions.

Work by Kane et al. (Kane et al., 2008b), reviewed in Chapter 2, explored an adaptive interface that increased target size accordingly to users' motion. Although the authors did not report improvements in performance, this type of interfaces shows great potential to deal with the dynamic nature of mobile conditions. Interfaces can take advantage of context data and adapt in real-time to a wide range situations.

Similarly, others have used accelerometer data to compensate hand vibrations whilst walking (Rahmati et al., 2009)(Yamabe and Takahashi, 2007). One of the most successful approaches was presented by Goel et al. (Goel et al., 2012). The authors introduce WalkType, a text-entry model that uses the device's accelerometer to compensate typing errors.

While all these solutions attempt to deal with the effect of walking, our goal in this chapter is to understand users' performance in order to build new solutions that can bridge the gap between SIID and HIID, namely in tremor-induced impairments.

4.2. User Study

Touchscreen devices are increasingly replacing their button-based counterparts. The physical stability once provided by buttons is being lost, which makes it harder to accurately select targets. This is especially relevant in text-entry tasks whilst mobile due to both device and physical constraints, such as small targets and tremor, respectively. This user study aims at understanding the negative effect of mobility and hand posture on typing performance. The next sections describe our research questions and experimental protocol.

4.2.1. Research Questions

This user study aims to answer the following research questions:

1. *Does mobility affect text-entry performance? If so, how?*
2. *What are the most common types of errors whilst mobile?*
3. *Does hand posture (and target size) compensate for the negative effect of mobility?*
4. *Does a slow walking pace compensate for the effect of a normal walking pace?*
5. *What is the preferred and least preferred hand posture?*

4.2.2. Participants

Twenty two participants, 3 females and 19 males, took part in the user study. Their age ranged from 23 to 40 with a mean of 26.5 years old. They were recruited in the University campus. None of the participants had visual or motor impairments and all of the participants owned a mobile phone, however only 15 of them used touchscreen technology regularly. All participants were right-handed.

4.2.3. Procedure

At the beginning of the experiment participants were told that the overall purpose of the study was to investigate how text-entry performance was affected by walking conditions.

Figure 4.3: Participant following the pacesetter while entering text.

Next, participants filled in a pre-questionnaire about their demographic data and mobile phone usage (see Appendix A4 for details). Participants were then informed about the experiment and how to use our evaluation application.

We evaluated the participants' performance in three conditions: sitting, walking at average human pace (2 paces per second), and walking at 65% of average human pace (1.3 paces per second) (Barnard et al., 2005). The experiment was conducted in an indoor test track built-up at IST-Tagus Park campus (Figure 4.3). In the sitting condition, participants sat at a desk in a controlled and quiet environment. We instructed them to remain seated until they completed all text-entry tasks. In both walking conditions, we asked participants to follow a pacesetter while entering text. Although other designs could be chosen (Kane et al., 2008b), we opted to keep a fixed pace rather than measure it as a dependent variable in order to ensure a comparable level of walking demand across trials. The experimenter instructed participants to stay within 2 meters of the pacesetter as he walked. If the participant fell behind the pacesetter by more than 4 meters, the experimenter logged a walking deviation for that trial. The pacesetter carried a mobile phone, which gave him feedback through vibration, about the intended pace. Also, before each mobility condition participants had a 5 minute practice trial to get used to the pace and text-entry task.

For each mobility setting, they were asked to enter text with 3 hand postures (chosen randomly), always using their thumbs: portrait/one-handed, portrait/two-handed, and landscape/two-handed. For each condition participants copied seven different sentences (the first two sentences were practice trials), displayed one at a time, at the top of the screen (Figure 4.4 - right). 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.

We wanted to elicit natural typing behaviors and did not want participants to be concerned

Figure 4.4: Text-entry application and virtual keyboard on portrait model (left), and HTC Desire (right).

with the accuracy of their input. Thus, we followed a similar approach to Gunawardana et al. (Gunawardana et al., 2010) and Goel et al. (Goel et al., 2012), and created a keyboard in such a way that error correction was not available. On the other hand, if the prototype had a delete key it might introduce correcting strategies, which might vary across participants and upset the naturalness of the data. Participants were told that they could not correct errors and were instructed to continue typing if an error occurred. Once participants finished entering each sentence, they pressed the 'next' button to receive a new sentence. When the seven sentences were entered, we asked participants to perform the same tasks on a new mobility condition, until they performed all mobility settings.

Each participant entered a total of 63 different sentences. These sentences were extracted from a written language *corpus*, each with 5 words, an average size of 4.48 characters per word, and a minimum correlation with the language of 0.97 (see Apendix A1). Both sentences and mobility conditions were chosen randomly to avoid bias associated with experience.

4.2.4. Apparatus

A HTC Desire with a capacitive touchscreen (see Figure 4.4 - left) was used in the user study. A QWERTY virtual keyboard was used to simulate a traditional touch keyboard, where each key was 10x10mm on landscape mode, and 7x10mm on portrait mode. Neither word prediction nor correction was used. Acceleration data was captured through the device's accelerometer for later analysis.

Regarding the pacesetter, he also had a mobile device, which gave him feedback through vibration, so he could maintain a steady pace.

4.2.5. Dependent Measures

The performance during text-entry tasks was measured by several quantitative variables: *Words Per Minute* (*WPM*), *Minimum String Distance* (*MSD*) error rate, and character-level errors (*substitutions*, *insertions*, and *omissions*). Qualitative measures were also gathered in the end of the experiment by debriefing each participant. *Walking errors*, such as slowing down and stopping were recorded by the experiment supervisor.

We also gathered the motor demand (acceleration) of mobility conditions. This was achieved through the device's accelerometer sensor. Thus, with this measure we were able to objectively characterize the motor demand of all mobility conditions for further analysis.

4.2.6. Design and Analysis

We used a within-subjects design where each participant tested all mobile conditions. For each condition each participant entered 5 test sentences, resulting in a total of 45 sentences per participant. In summary the study design was: 22 participants x 5 sentences x 3 hand postures x 3 mobility settings (1 seated + 2 walking conditions) = 990 sentences overall.

Shapiro-Wilkinson tests of the observed values for *WPM*, *MSD* error rate, and types of errors showed to fit a normal distribution for all conditions. Therefore, a two-way repeated-measures ANOVA (*mobility* x *hand posture*) was used in further analysis. Greenhouse-Geisser's sphericity corrections were applied whenever Mauchly's test of sphericity showed a significant effect. Pairwise Bonferroni corrected t-tests were used for post-hoc tests.

4.3. Results

Our goal was to understand the effect of mobility on text-entry performance with touch-screen devices. In this study we focused on the use of thumbs with 3 hand postures (one-hand portrait, two-hand portrait, and two-hand landscape), enabling us to analyze the effect of two-hand interaction and key size. The results presented are threefold: text-entry performance, walking performance and users' preference.

4.3.1. Input Speed

To assess speed, we used the Words Per Minute(WPM) text input measure, calculated as:

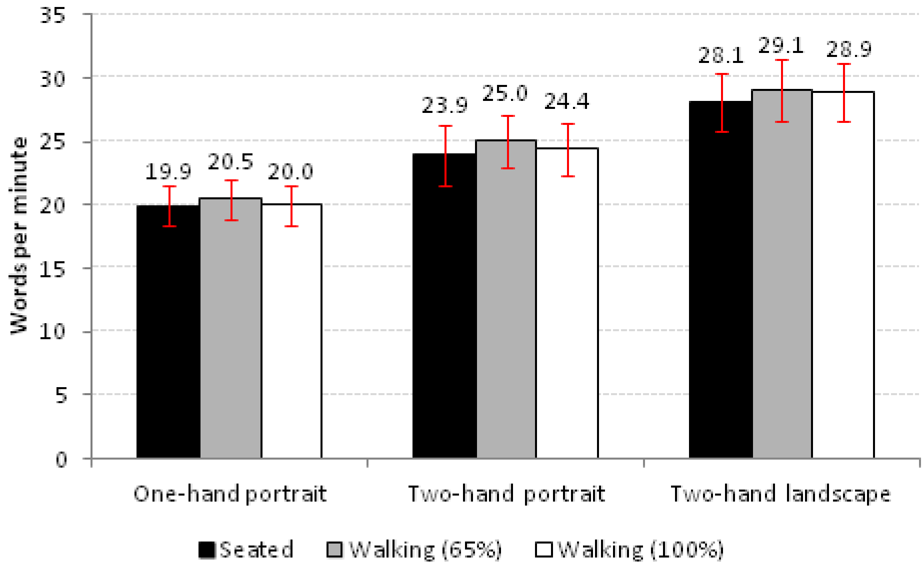


Figure 4.5: Words per minute across all mobility and hand conditions. Error bars denote 95% confidence intervals.

$$(transcribed\ text\ length - 1) \times (60\ seconds \div time\ in\ seconds) \div 5\ characters\ per\ word$$

Figure 4.5 shows the participants' average WPM for all mobility conditions and hand postures.

Effect of mobility on text-entry speed. Overall, input speed was very similar between mobility conditions. Participants obtained an average (confidence interval) of 24 wpm (1.4), 24.9 wpm (1.4), and 24.5 wpm (1.4) on seated, slow walking and regular walking conditions, respectively. Results show that there was no significant main effect of *mobility* on *WPM* ($F_{2,42}$=.97, p>.1). This effect was expected since participants could not perform corrections to transcribed sentences, thus they were continuously typing.

Effect of hand posture on text-entry speed. Participants wrote an average (confidence interval) of 20 wpm (0.9) with one-hand. Two-hand postures allowed participants to reach higher input speeds: 25 wpm (1.2) and 29 wpm (1.3) with portrait and landscape modes, respectively. A repeated measures ANOVA revealed a significant main effect of *hand posture* on *WPM* ($F_{2,42}$=84.878, p<.001). Bonferroni post-hoc tests showed significant differences between all hand postures. As expected, writing with a two-hand landscape posture is significantly faster, followed by two-hand portrait and one-hand modes.

Effect of mobility and hand posture on text-entry speed. Results show no interactions between *mobility* and *hand posture*, meaning that the main effects trends were

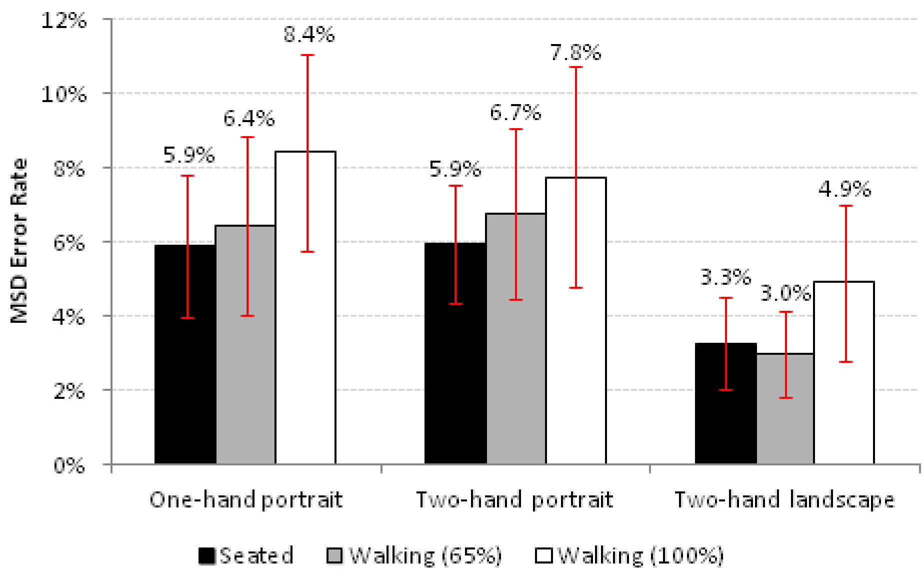


Figure 4.6: Minimum String Distance (MSD) across all mobility and hand conditions. Error bars denote 95% confidence intervals.

consistent between all factor levels ($F_{4,84}$=.373, p>.8).

4.3.2. Quality of Transcribed Sentences

The quality of the transcribed sentences was measured using the Minimum String Distance (MSD) error rate, calculated as:

$$MSD(required\,text, transcribed\,text) \div mean\,size\,of\,the\,alignments \times 100$$

Figure 4.6 presents the participants' MSD error rate for all mobility conditions and hand postures.

Effect of mobility on quality of transcribed sentences. In general, the *mobility* effect can be clearly seen in Figure 4.6, as MSD error rate increases between seated, slow walking, and regular walking. Accuracy decreases approximately 2.4% with participants achieving an average MSD error rate (confidence interval) of 5.1% (1.7%), 5.7% (2.3%), and 7.5% (3%) on seated, slow walking, and regular walking, respectively. Results show a significant main effect of *mobility* on *MSD* error rate ($F_{1.244,26.115}$=4.962, p<.05), with significant differences between seated and normal walking conditions.

Effect of hand posture on quality of transcribed sentences. Participants MSD error rate was very similar for portrait hand postures: 7.1% (2.5%) and 7.3% (2.7%) for one-handed and two-handed modes, respectively. This suggests that two-hand interaction

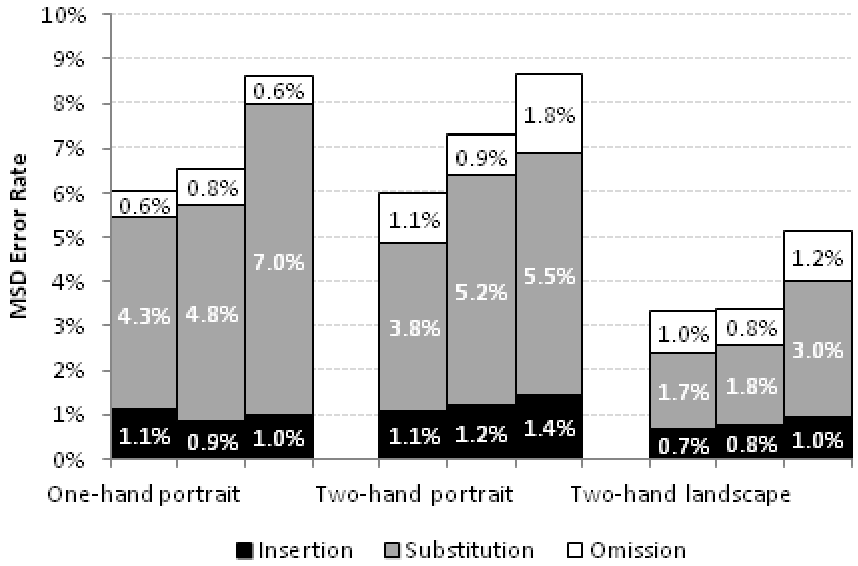


Figure 4.7: Types of errors (insertion, substitution, omission) across all mobility and hand conditions. For each hand posture: left bar is seated condition; middle bar is slow walking condition; and right bar is normal walking condition.

by itself does not provide additional accuracy. Nevertheless, in the landscape mode, accuracy increased nearly 3.4%, resulting on a significant main effect of *hand posture* on *MSD* error rate ($F_{2,42}$=16.546, p<.001). Post-hoc tests showed significant differences between two-hand landscape posture and both one-hand portrait and two-hand portrait postures. This suggests that key size has a greater influence than hand grip on text input accuracy and can compensate the negative effect of mobility.

Effect of mobility and hand posture on quality of transcribed sentences. No significant interactions between *mobility* and *hand posture* were found on *MSD* error rate.

4.3.3. Typing Errors

This section presents a fine grain analysis by categorizing the types of errors and characters that are problematic for each mobility condition and hand posture. Figure 4.7 shows the type of errors performed on all mobility conditions and with all hand postures.

Effect of mobility on typing errors. Regarding *insertions*, there was no significant effect of *mobility* ($F_{1.449,30.429}$=.674, p>.1). Participants committed an average of 1%, 0.9%, and 1.1% while seated, at slow pace, and at normal pace, respectively. On the other hand, average substitution errors were slightly higher: 3.3%, 4%, and 5.1% for seated, slow walking, and normal walking conditions, respectively. These results showed a significant effect of *mobility* on *substitution* errors ($F_{1.440,30.243}$=, p<.05). Significant differences were

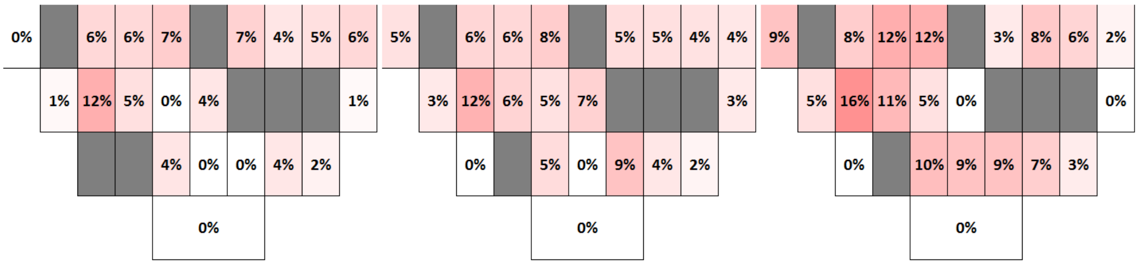


Figure 4.8: Substitution errors per key for one-hand portrait mode while seated (left), slow walking (middle), and normal walking (left). Key's "redness" illustrates higher error rates. Grey cells represent keys that were not used in the experiment.

found between seated and normal walking condition. Regarding *omission* errors, there was no significant main effect of *mobility* ($F_{1.453,30.511}$=1.167, $p>.1$). Participants committed an average of 0.9%, 0.9%, and 1.2% while seated, walking at slow pace, and walking at normal pace, respectively.

Effect of hand posture on typing errors. The average insertion error rate per participant was 0.9%, 1.2%, and 0.8% with one-hand, two-hand portrait and two-hand landscape modes, respectively. As with *mobility*, we found no significant effect of *hand posture* on *insertion* error rate ($F_{2,42}$=2.211, $p>.1$). On the other hand, we found a significant effect of *hand posture* on *substitution* errors ($F_{2,42}$=16.578, $p<.001$). Differences were found between two-hand landscape (2.1%) and both one-hand portrait (5.3%) and two-hand portrait (4.8%) postures. Moreover, results showed a significant effect of *hand posture* on *omission* Errors ($F_{2,42}$=3.393, $p<.05$) with significant differences between one-hand portrait (0.7%) and two-hand portrait (1.3%). This means that when using two hands users tend to commit more omission errors.

Effect of mobility and hand posture on typing errors. We found no significant interactions between *mobility* and *hand posture* for *insertion*, *substitution*, and *omission* errors.

4.3.4. Substitution Errors

In this section, we perform a more detailed analysis on substitution errors, since these were the most common type of errors. Based on our results, we created confusion matrices for each condition. We identify the most problematic characters and compare the results between all hand postures and mobility conditions. Also, we will analyze possible causes to substitution errors in order to draw some design recommendations and provide the knowledge to improve future designs of virtual keyboards.

One-hand portrait analysis

Substitution error rate per letter for one-hand portrait condition is shown in Figure 4.8. In the mobility baseline condition keys in the top row were the most difficult to hit accurately.

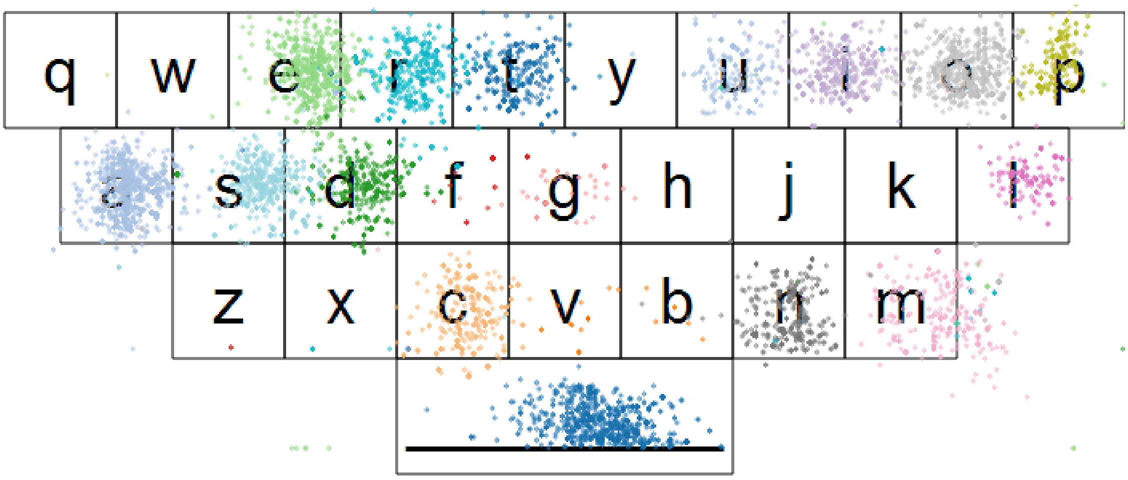


Figure 4.9: Touch (lift) points for all participants in walking / one-hand portrait condition.

The 'S' key also had a high error rate. Regarding the remaining mobility conditions, error rate seems to increase to all keys, proportionally to motor demand. This effect may be due to the lack of stability provided by one-hand interaction.

In the seated condition the most frequent substitution errors were: E $\rightarrow$ R (4.6%), R $\rightarrow$ T (6.5%), S $\rightarrow$ D (11.5%), T $\rightarrow$ Y (4%), U $\rightarrow$ Y (4.5%), U $\rightarrow$ I (3.4%). As we can see there is a clear predominance of same-row errors in the data, which suggests that participants found it easier to hit keys in the vertical direction than horizontally. This may be because key height is slightly higher than key width. Additionally, all errors are at a distance of one key and typically at the right (see Figure 4.9). These findings need further investigation, but may be related to hand dominance.

In the slow walking condition the most frequent substitution errors were: B $\rightarrow$ N (11.1%), B $\rightarrow$ SPACE (11.1%), S $\rightarrow$ D (10.6%), T $\rightarrow$ Y (6.2%). Unlike the baseline condition where most keys were on the top row, now they are spread across the keyboard. However, the same pattern of substitution (right key) is seen.

In the normal walking condition the most frequent substitution errors were: D $\rightarrow$ F (10.4%), R $\rightarrow$ T (9.3%), S $\rightarrow$ D (11.8%), T $\rightarrow$ Y (5.2%), T $\rightarrow$ R (4.6%). In this case, the keys with the highest error rates seem to be more distant from the users' dominant hand. Once again, the pattern of substitution is the same (right key). It is noteworthy that although mobility demand increased from seated to normal walking condition, substitutions remained at a distance of one key.

Two-hand portrait analysis

Figure 4.10 illustrates substitution error rates per letter in the two-hand portrait posture. In the seated condition error rates do not seem to follow a specific pattern, as they are scattered in several different keys. This result may be an effect of two-hand interaction. As mobility demand increases, higher error rates seem to cluster on the left side, suggesting

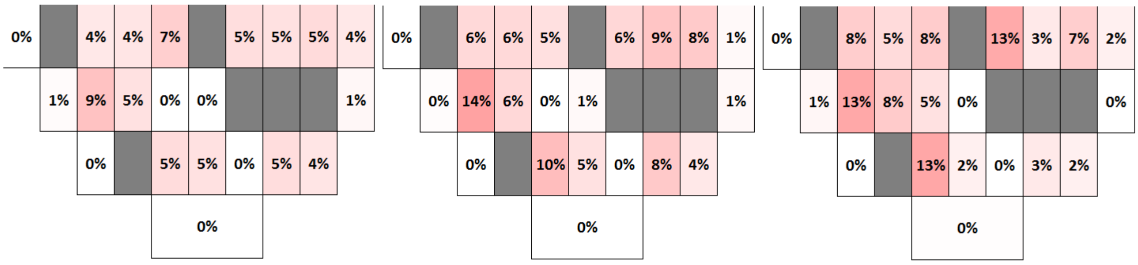


Figure 4.10: Substitution errors per key for two-hand portrait mode while seated (left), slow walking (middle), and normal walking (left). Key's "redness" illustrates higher error rates. Grey cells represent keys that were not used in the experiment.

that the non-dominant hand is less accurate.

In the seated condition the most frequent substitution errors were: C → X (3.1%), N → M (2.1%), N → B (2.1%), M → N (4.1%), S → A (6.8%), T → Y (4.4%), T → R (2.3%), U → Y (4.3%). As with one-hand posture, keys are usually substituted by others that are adjacent (one key distance) on the same row. However, unlike one-hand the substitution is not always on the right key. There seems to be an effect of the two-hand interaction.

In the slow walking condition the most frequent substitution errors were: C → V (4.2%), C → X (3%), I → U (5.2%), N → M (2%), O → I (5.8%), S → A (12.7%). The substitution pattern is the same as the seated condition.

In the normal walking condition the most frequent substitution errors were: C → X (4.4%), C → V (3.2%), S → A (12.6%), U → I (4.2%). Again, the substitution pattern is identical to seated condition.

Two-hand landscape analysis

Two-hand landscape hand posture conditions resulted in fewer errors, due to target size (Figure 4.11). Moreover, error rates were fairly distributed across keys. In the seated condition the most frequent substitution errors were: L → P (2.3%), V → G (5.9%). Unlike portrait postures, data from landscape mode suggests that substitution errors are in different rows, however in the same columns.

In the slow walking condition the most frequent substitution errors were: T → R (0.6%), V → C (7.1%). In this mobility conditions, substitution errors are, once again, in the

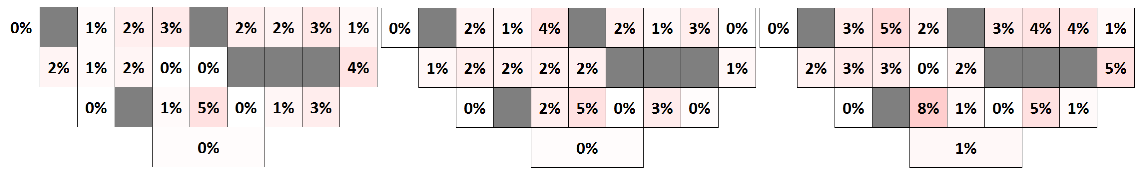


Figure 4.11: Substitution errors per key for two-hand landscape mode while seated (left), slow walking (middle), and normal walking (left). Key's "redness" illustrates higher error rates. Grey cells represent keys that were not used in the experiment.

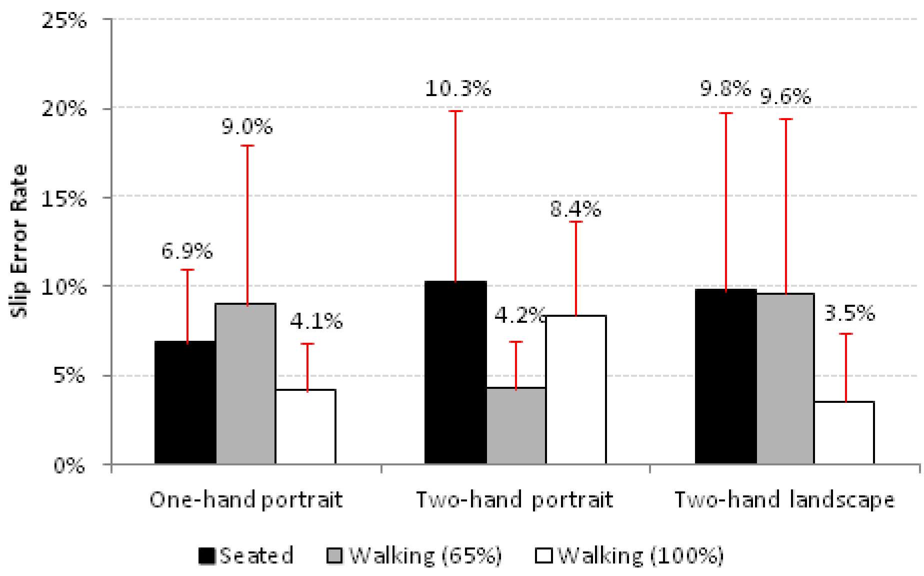


Figure 4.12: Slip error rate across all mobility and hand conditions. Error bars denote 95% confidence interval.

same row. This finding needs more investigation, however it may be related with hand tremor.

In the normal walking condition the most frequent substitution errors were: D $\rightarrow$ F (0.5%), R $\rightarrow$ T (2.2%), S $\rightarrow$ A (2.1%), T $\rightarrow$ Y (1.4%),T $\rightarrow$ R (1.2%). In this case, keys with higher substitution errors are placed on the left side of the keyboard, which suggest that the non-dominant hand performs worse.

Overall, mobility seems to only increase error rate magnitude; however, the substitution pattern is similar between conditions: same row errors and adjacent keys. Additionally, on one hand interaction, letters are usually substituted by their right-side keys, while while in two hand interaction this effect dissipates. Nevertheless, the non-dominant hand seems to have a lower accuracy, particularly when mobility conditions are more demanding.

Slip error rate

In addition to assessing substitution patterns, we were also interested in finding how these errors occurred: was it because users' fingers slipped during typing or they already landed on the wrong keys?

A slip error occurs when the users' finger slips to an adjacent key before lifting his finger, resulting in a substitution error. Since slips can be one of the main causes of substitution errors, in this section we analyze the amount of substitutions caused by slips as well as the effect of both mobility and hand posture (Figure 4.12).

Effect of mobility on slip error rate. The average slip error rate (confidence interval) was 9% (8%), 7.6% (7.7%), and 5.3% (6.5%) when seated, slow walking and normal walking, respectively. Results show no significant main effect of *mobility* on *slip* error rate ($F_{2,42}$=1.005, p>.1), mainly due to the large variance of results (i.e. high standard deviation).

Effect of hand posture on slip error rate. Regarding hand posture, the average slip error rate (confidence interval) was 6.7% (5.9%), 7.6% (6.5%), and 7.6% (8.3%). We found no significant main effect of *hand posture* on *slip* error rate ($F_{2,42}$=0.089, p>.1), which suggest that this type of error is not affected by key size or interaction mode (one-hand vs. two-hand).

Effect of mobility and hand posture on slip error rate. No interactions between *mobility* and *hand posture* were found on *slip* error rate ($F_{4,84}$=0.637, p>.1).

The results described in this section are very interesting. Although substitution errors are the most common type of error and highly sensitive to both mobility and hand posture, slips do not seem to be the main cause of this effect. One could assume that due to hand oscillation the users' fingers could slip to an adjacent key before entering the letter, resulting in a substitution error. However, on average, slips accounted for a minority of substitutions and were not affected by mobility. A thorough analysis revealed that, on average, 81-96% of substitution errors are due to poor aiming, i.e. incorrect land-on target. Although participants could compensate for land-on errors, most of them performed quick taps as if they were typing with a physical keyboard, resulting in wrong characters.

4.3.5. Insertion Errors

In this section we explore one possible cause of insertion errors - bounces. A bounce error occurs when a key is unintentionally pressed more than once, producing unwanted characters. Also, we analyze what percentage of insertion errors are caused by bounce behaviors. Moreover, we examine the effect of mobility as bounce errors may be affected by tremor (Figure 4.13).

Effect of mobility on Bounce error rate. On average bounce errors account for 10%, 8.8%, and 8.5% of insertions when seated, slow walking and normal walking, respectively. An analysis of variance showed no significant main effect of *mobility* on *bounce* error rate ($F_{2,42}$=0.183, p>.1). As insertion errors, bounce errors do not seem to be affected by mobility.

Effect of hand posture on Bounce error rate. Regarding hand posture, results show no significant effect of *hand posture* on *bounce* error rate ($F_{2,42}$=0.972, p>.1). This result was somewhat expected since insertion errors were not affected by hand posture.

Effect of mobility and hand posture on Bounce error rate. No interactions between *mobility* and *hand posture* were found on *bounce* error rate ($F_{4,84}$=0.516, p>.1).

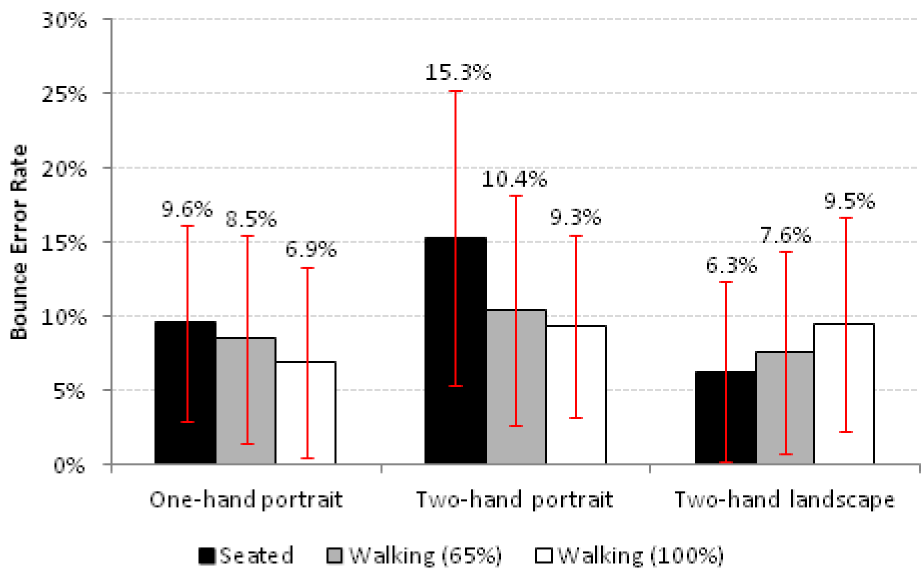


Figure 4.13: Bounce error rate across all mobility. Error bars denote 95% confidence interval.

4.3.6. Time-based Analysis

The main goal of this section is to analyze time-based events, such as key presses and key holding events, in order to draw some recommendations for future keyboard designs. As expected we examine the differences between mobility and hand conditions.

Key Press Time

Key press time consists in the time elapsed between the key down and key up events. This measure illustrates the average time needed for a key selection, allowing us to draw some insights regarding the time needed for selecting a key in different mobility and hand conditions. Figure 4.14 illustrates the results obtained from the user study.

Effect of mobility on key press time. The average key press time (standard deviation) was 143 ms (30.9 ms), 150 ms (40.7 ms), and 154 ms (38.1 ms) for seated, slow walking and normal walking conditions, respectively. Results show a significant effect of *mobility* on *key press time* ($F_{2,42}$=4.283, p<.05) between the seated and normal walking conditions. This means that the delay for key holding events should be slightly higher whilst mobile in order to prevent activation of an unintended interaction mode.

Effect of hand posture on key press time. Regarding the effect of *hand posture* on *key press time* there were no significant effects ($F_{1.444,30.334}$=0.811, p>.1). Average key press time for one-hand portrait, two-hand portrait, and two-hand landscape was 151 ms (43 ms), 150 ms (33.1), and 145 ms (33.8), respectively. Participants took, approximately, the same time to press a key independently of hand posture or target size.

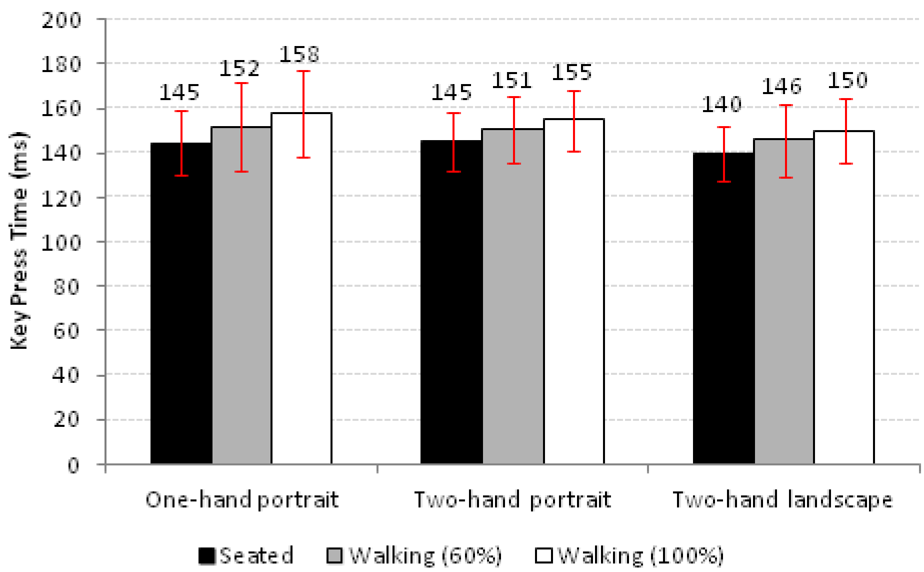


Figure 4.14: Key press time (ms) across all mobility and hand conditions. Error bars denote 95% confidence interval.

Effect of mobility and hand posture on key press time. We found no significant interaction between *mobility* and *hand posture* on *key press time.*

Long Key Press error rate

A long key press error consists of a key that was pressed for longer than the active key hold delay, producing unwanted characters. In this user study we did not use long key presses to enter alternative characters(such as punctuation, accentuation or numbers); nevertheless, we analyzed the participants' typing behaviors in order to draw some recommendations about the most appropriate holding delay, regarding hand posture and mobility condition. Thus, we selected 9 delay values: 100ms, 150ms, 200ms, 250ms, 300ms, 350ms, 450ms, 600ms, 750ms (used in HTC keyboard), and 900ms. Error rates that would occur are shown in Figure 4.15.

Effect of delay on long key press error rate. Results show a significant main effect of *delay* on *long key press* error rate ($F_{1.747,36.690}$=184.664, p<.001). Differences reveal that error rates begin to converge at 250ms with an average of 4% error rate.

Effect of mobility on long key press error rate. We found a significant main effect of *mobility* on *long key press* error rate ($F_{2,42}$=3.869, p<.05), with significant differences between seated and normal walking conditions. Results show that, in general, error rates are higher whilst mobile.

Effect of hand posture on long key press error rate. We found no significant main effect of *hand posture* on *long key press* error rate ($F_{1.424,29.905}$=1.504, p>.1). This means that users commit the same amount of long key press errors independently of how they hold the device (one-hand, two-hand portrait or two-hand landscape).

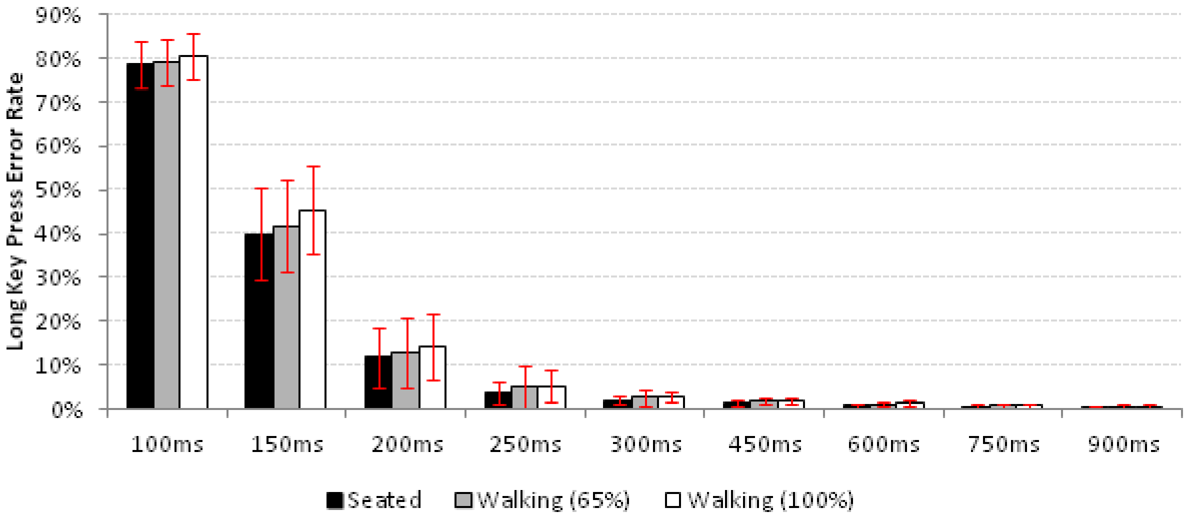


Figure 4.15: Average long key press error rates across mobility conditions for 100, 150, 200, 250, 300, 450, 600, 750, and 900 ms. Error bars denote 95% confidence interval.

Effect of delay, mobility, and hand posture on key press error rate. No interactions were found regarding *delay*, *mobility*, and *hand posture* on *key press* error rate.

Results show that hold delay could be smaller than traditional values, enabling a faster insertion of alternative characters. For example, adopting a delay of 300ms would result in less than 5% error rate.

4.3.7. Walking Performance

In this user study, participants had to follow a pacesetter on mobility conditions. Walking errors were counted when users stopped or lagged behind the pacesetter by more than 4 meters. When that happened, the pacesetter waited for the participant and then resumed the pace. Figure 4.16 illustrates the participants' walking errors during this user study.

Effect of mobility on walking errors. The average number of walking errors (confidence interval) per participant was 0.05 (0.05) and 1.1 (0.4) for slow walking and normal walking, respectively. Results show that participants committed significantly more errors whilst walking at a normal pace than whilst walking at slow pace ($F_{1,21}$=10.010, p<.05).

Effect of hand posture on walking errors. Regarding hand postures, the average number of walking errors (confidence interval) per participant was 0.5 (0.3), 0.4 (0.3), and 0.3 (0.2) for one-hand portrait, two-hand portrait, and two-hand landscape, respectively. Results show a significant effect of *hand posture* on *walking* errors ($F_{1.575,33.082}$=3.847, p<.05). Significant differences were found between landscape and both portrait postures.

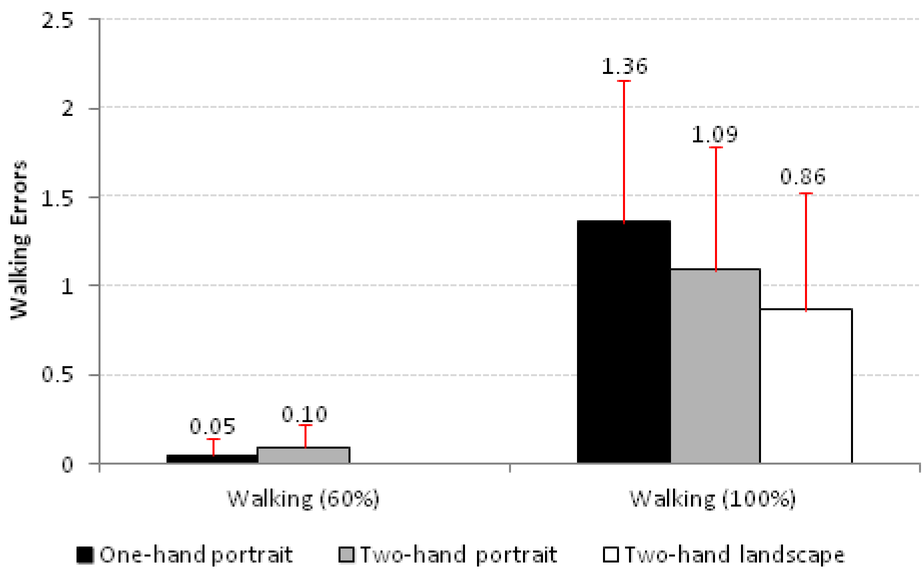


Figure 4.16: Number of walking error for all walking and hand condititons. Error bars denote 95% confidence intervals.

This result suggests that easier text-entry methods allow for better performance on walking tasks.

Effect of mobility and hand posture on walking errors. No interactions between *mobility* and *hand posture* were found ($F_{1.541,33.354}$=2.327, p>.1), which shows that all hand postures equally affected walking errors for all levels of mobility.

4.3.8. Hand Oscillation

In this user study, we measured the oscillation of the dominant hand that held the device, through its built-in accelerometer. Amplitudes of hand oscillation were calculated as standard deviations of acceleration readings (Bergstrom-Lehtovirta et al., 2011).

Hand oscillation varied from 0.1 m/s^2 in the seated condition to nearly 2 m/s^2 whilst walking at normal pace, showing that oscillation clearly increases with mobility. Figure 4.17 shows the relationship between MSD error rate and hand oscillation. As expected, there is a certain degree of dependency between these two measures, since error rate increases with oscillation. As reported in Section 4.3.3 there is a main effect of *mobility* on *MSD* error rate ($F_{1.244,26.115}$=4.962, p<.05). Moreover, each mobility condition is easily identifiable through oscillation intervals, which implies that future mobile interfaces can take advantage of mobile sensing capabilities to improve user performance.

From Figure 4.17 we can also see the relationship between hand oscillation and hand posture. Both portrait postures (one and two hand) have similar hand oscillations and, consequently, similar error rates. This result suggests that two-hand interaction by itself

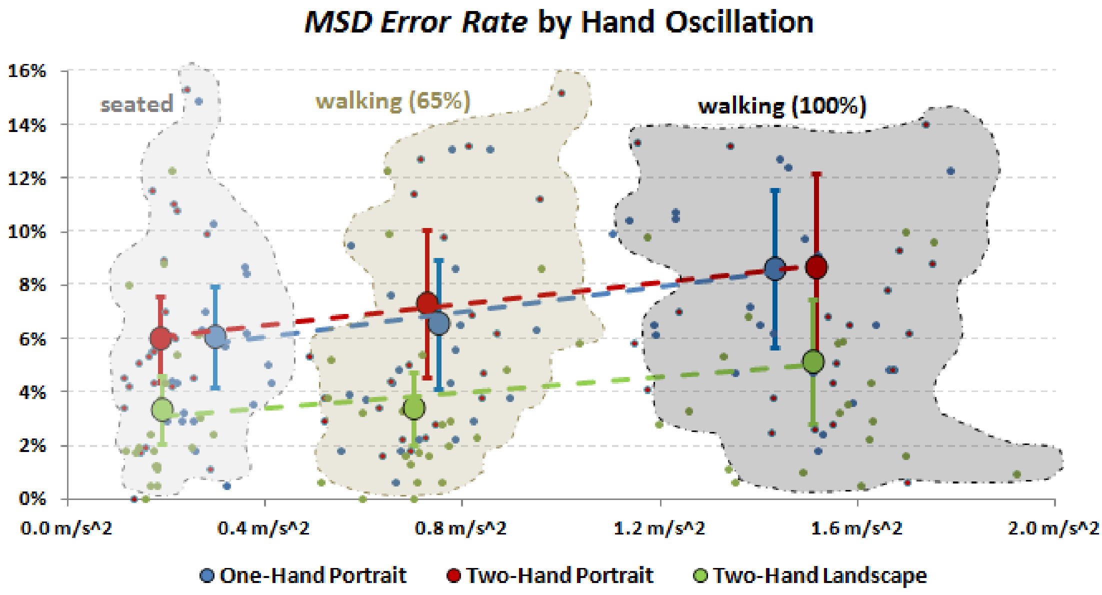


Figure 4.17: MSD error rate by hand oscillation amplitude. Vertical bars denote 95% confidence intervals. Lines illustrate linear regression.

does not provide additional physical stability. On the other hand, key size seems to positively influence the quality of transcribed text. These results are consistent with the ones presented in Section 4.3.3, where significant differences were found between two-hand landscape and both portrait postures.

4.3.9. Participants Preference and Comments

In the end of the user study participants were debriefed and asked about their preferred hand posture, as well as general comments about their performance.

Regarding preference, 81.8% of participants (18 out of 22) preferred the two-hand landscape mode (Figure 4.18), with a 95% adjusted-Wald binomial CI ranging from 60.9% to 93.%, a lower limit well above the three-choice chance expectation of 33.3%. None of the participants chose this interaction mode as the least preferred, showing that generally users prefer larger keys to input text.

> "In portrait mode the size of the keys is too small to hit with my thumbs."

> "I prefer the two-hand landscape posture since the keys are larger."

On the other hand, four participants preferred the two-hand portrait posture. Those participants stated that they liked this hand posture due to speed (near targets are faster to hit) or convenience, i.e. they usually do not spend time rotating their devices to write. Interestingly, this preference was not correlated with a better performance.

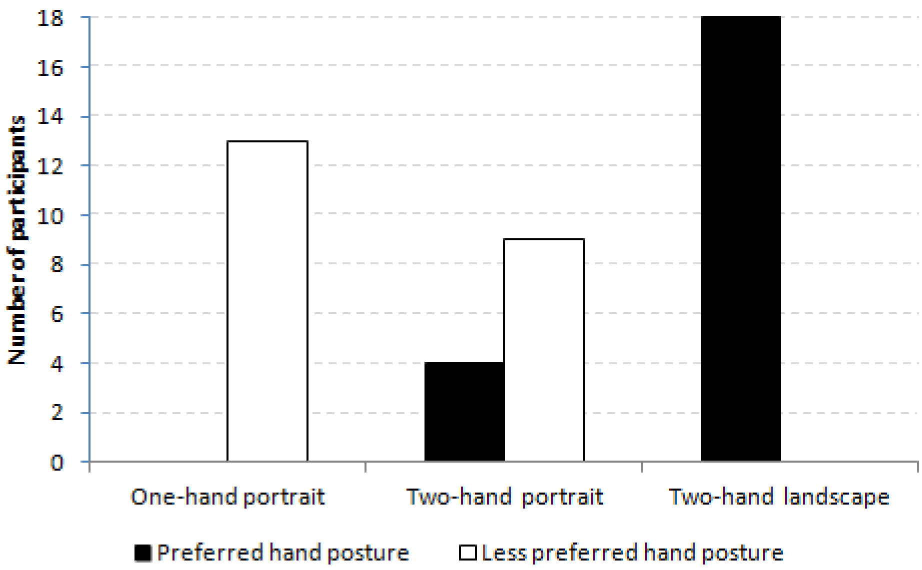


Figure 4.18: Participants preference regarding hand postures.

> "Using two hands is definitely faster and the compactness of the portrait layout makes it even faster to type."

However, 9 participants chose this posture as the least preferred, stating that interacting with two hands on such a small keyboard is very difficult. Overall, this hand posture is preferred by a minority, while generally disliked by most users.

> "In portrait mode I prefer to use only one hand."

> "Interacting with two hands on portrait position is as bad or event worst than with one-hand."

> "In portrait mode is easier to write with one hand."

> "In portrait the keys are too small."

> "In this position (portrait) the fingers occlude almost all the keys."

The one-hand portrait posture showed to be the least preferred for 13 of our participants. Moreover, none chose this interaction mode as his/her preferred.

> "Hitting keys near the right bottom size, such as 'm' and 'n' is very difficult."

> "Writing with only one hand is much more difficult."

> "With one hand is nearly impossible to hit the more distant keys."

When asked about their main difficulties during the user study, participants pointed out several issues related to: 1) mobility conditions (e.g. "The tremor, due to walking, makes text input harder."); 2) task (e.g. "Sometime I forgot to copy some of the words on the required sentence."); and 3) keyboard ("It is weird to use a QWERTY keyboard with our thumbs. I usually use other mobile keyboards.", "I frequently press adjacent keys, because they are too small and have no texture", "Keys are too close to each other.")

4.4. Discussion

In this section we discuss our results by: 1) answering the research questions proposed at the beginning of this study, 2) presenting design guidelines, and 3) describing the limitations of the experiment.

4.4.1. Answering Research Questions

After analyzing all data from different mobility conditions and hand postures, we are now able to answer the research questions proposed at the beginning of this user study.

1. *Does mobility affect text-entry performance? If so, how?*

   Regarding words per minute, we found no significant effect of *mobility*, meaning that participants maintained their input speed across all mobility conditions (around 24.5 wpm). This finding is consistent with other text-entry user studies. Nevertheless, this effect may be due to our experiment design as participants were not able to delete incorrect letters, thus they were always typing. On the other hand, typing quality significantly decreased with mobility. While insertion and omission errors were constant between all conditions, substitution errors were the most sensitive to mobility. Thus, we found significant differences between seated and normal walking conditions.

2. *What are the most common types of errors whilst mobile?*

   The most common types of errors are substitution errors. Also, these were the most sensitive to mobility conditions as significant differences were found between the baseline and normal walking conditions. Results show that these errors are mainly due to poor aiming. In fact, this finding has been consistently reported by the HCI community. Holz and Baudisch (Holz and Baudisch, 2011) explain this effect as a "paralax" artifact between user control and interface sensing; that is, users' (visual) control features are misaligned with sensing assumptions of current touch devices. Moreover, in a large-scale user study, Henze and colleagues (Henze et al., 2011) analyzed more than 1 million touch events, showing that finger-touch positions are systematically skewed and originate aiming errors. In touch typing tasks, Findlater et al. (Findlater et al., 2011) report similar results to ours: incorrect keys most often occur as adjacent keys in the same row. A key difference to our work lays in the fact that we analyzed the mobility effect and cause of these errors, showing that finger slips only account for a minority of error rates. Users' poor aiming was the most common issue whether seated or whilst walking.

3. *Does hand posture (and target size) compensate for the negative effect of mobility?*

   We found a significant main effect of *hand posture* on *words per minute*, showing that two-hand landscape posture is significantly faster, independent of mobility condition, followed by two-hand portrait, and one-hand portrait postures.

   Regarding *MSD* error rate, target size seems to have a greater influence, since significant differences were only found between two-hand landscape and the remaining postures. This means that holding the device with two hands does not necessarily imply more stability and fewer errors. However, the two-hand landscape mode effectively compensates the negative effect of mobility, since participants achieved a lower error rate than with the two-hand portrait posture. This means that by increasing target size, we are able to compensate for typing errors.

4. *Does a slow walking pace compensate for the effect of a normal walking pace?*

   We did not find any significant differences on *MSD* error rate between seated and slow walking conditions, which may suggest that users maintain performance and can compensate for the effect of mobility by reducing their pace. However, we neither found significant differences between slow and normal walking conditions. Therefore, further research is needed to accurately answer this question. So far our findings are inconclusive.

5. *What is the preferred and least preferred hand posture?*

   Overall, participants preferred the two-hand landscape posture, mainly because of the larger target sizes. On the other hand, the least preferred was the one-hand posture. Participants complained about the small target size and ergonomic factors, such as hitting near and distant targets. Also, some highlighted the fact that whilst mobile, one hand interaction is much less precise due to tremor.

4.4.2. Design Implications

Do not over rely on two-hand interaction for physical stability. Future mobile text-entry solutions should not rely on hand grip to improve typing effectiveness. Data shows that two handed interaction does not decrease hand oscillation nor does it improve input quality. As an alternative, increasing key sizes allow users to compensate the negative effect of mobility. However, this solution is not always possible due to limited screen size.

Adjacent substitutions. Results show that substitutions typically occur on the same row and adjacent keys. Therefore, predictive text-entry methods and orthographic correction algorithms should take into account this typing behavior, by assigning higher weights on same row adjacent keys. Moreover, for one-hand interaction, right-side keys (assuming the right hand as the dominant one) should have a lower weight than left-side, since these are the most probable of being accidentally selected. These solutions will mostly likely work for different mobility conditions, since the substitution pattern remains unchanged.

Design for poor aiming, especially whilst mobile. Alternative modalities have been used to improve touch typing and reduce slip-based errors (Hoggan et al., 2008). However, these solutions only address a minority of substitution errors. Future designs should focus in dealing with poor aiming, in order to significantly enhance text-entry performance.

Relative key sizes. Results suggest that users are less accurate hitting targets near the non-dominant hand, especially whilst mobile. This is true for both one-hand and two-hand interactions. Nevertheless, since space is a limited resource on mobile devices, further investigation is needed to see if decreasing key size of one side of the keyboard while increasing the other one is beneficial for users. Additionally, future research should assess whether this finding holds true for left-handed users.

Smaller key hold delays. Traditional virtual keyboards frequently use key hold based techniques to provide new entry modes (e.g. symbols and numbers). Some keyboards, such as HTC virtual keyboard, have a default key hold delay of 750 ms. Although this value guarantee error rates below 1%, our results show that these start converging at 250ms with 4% error rate. Indeed, this delay would allow for a more fluid interaction while maintaining similar accuracy. Moreover, results from the user study revealed that key hold delays should be slightly higher whilst mobile in order to prevent activating unintended entry modes. This effect is especially seen with short delays, i.e. below 250ms.

Sense hand oscillations. Future text-entry methods should take advantage of mobile sensing capabilities by measuring hand tremor and automatically compensating touch locations.

4.4.3. Limitations

In this user study participants were not able to correct errors, i.e. use the delete key. While this decision was necessary to achieve our goals and measure natural typing patterns, allowing users to correct sentences would certainly impact behaviors and results (on both WPM and MSD error rate). Future research should explore new experiment designs and present participants with more realistic sentences, featuring punctuation and symbols in order to gather new error patterns and interaction difficulties.

Although our participants included several levels of expertise, from those who had never used a touchscreen mobile device to a couple years of experience, most participants belong to a specific age group (20-30). Also, in this user study we did not have any left-handed participants, which made it impossible to draw strong conclusions about hand dominance effects.

4.5. Conclusion

We have investigated text-entry performance of 22 participants whilst mobile. Our user study featured 3 mobility conditions (seated, slow walking, and normal walking) and 3 hand postures (one-hand portrait, two-hand portrait, and two-hand landscape), which used only thumbs.

Our results demonstrate that two-hand text input is faster, especially when on landscape mode due to the larger targets. However, when analyzing text quality, two-hand interaction is not always better. In portrait mode, error rates are as bad as one-hand interaction. Conversely, in landscape mode, error rates are significantly lower, which suggest that target size has a greater effect on text quality than hand posture (i.e one-hand vs. two-hand). This is also true for more demanding mobility conditions. Although mobility affects all hand postures equally regarding text quality, the landscape mode can compensate the negative effect of mobility when compared with one-hand interaction.

The most common type of errors, accounting for more than half of total errors, were substitutions, which consists in replacing the intended key by an adjacent one. Moreover, this type of error was the most sensitive to mobility conditions. As mobility demand increased, error rates also increased. To better understand substitution errors and prevent them in future designs, we analyzed participants' typing behaviors thoroughly. Results show that substitutions usually occur in adjacent keys of the same row. However, slips are not the main cause of these errors, accounting on average for less than 10%, which shows that participants usually land on the wrong key.

Regarding the walking task, participants slowed down their pace in order to maintain text-entry performance. Again, key size seems to have a greater influence than hand posture, since participants committed fewer walking errors in the landscape mode.

The results presented here should encourage researchers to pursue more accessible one-handed interaction methods, since two-hand interaction and large targets are not always feasible when considering mobile contexts. Those results, in addition to the described design implications, can improve future mobile designs towards more effective text-entry solutions.

5

Understanding Health-Induced Impairments and Disabilities

After characterizing mobile users main difficulties in typing tasks (Chapter 4), we now present the second stage of our research, in which we characterize the abilities of users with health-induced hand tremor. We focus on older adults, who experience an increase of physiological tremor (Sturman et al., 2005). The lack of physical affordances on current mobile touch-based devices makes it harder for these users to accurately select targets; especially when considering the small target size and narrow spacing of virtual keyboards (Jin et al., 2007).

Our goal, in this chapter, is to provide the knowledge needed to design effective and efficient text-entry solutions for older people. Moreover, we wanted to characterize users' abilities when touch typing and assess their relationship with users' hand tremor profile. We conducted evaluations with 15 users and two touch-based devices (Figure 5.1), analyzing the effect of hand tremor on text-entry performance. Moreover, we thoroughly describe the users' typing behaviors and performance errors, as well as their comments.

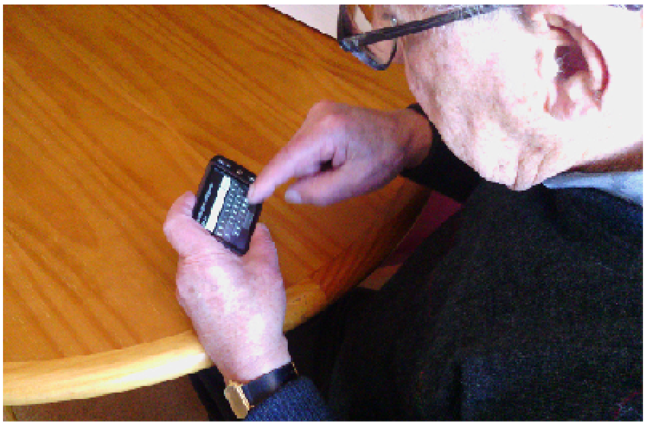


Figure 5.1: Participant typing on a touchscreen device.

5.1. Background

The related work presented in this section is two-fold: firstly, we describe previous works that try to better understand tremor and how it affects older adults; secondly, we discuss HCI research that focus on creating new touch-based solutions for this target population.

5.1.1. Older Adults and Tremor

In the most general sense, tremor is defined as any involuntary, approximately rhythmic, and roughly sinusoidal motion around a joint (Raethjen et al., 2000). Tremor is present in all individuals and is the most common form of movement disorder with an increased prevalence among older individuals (Benito-Leon et al., 2003)(Van Den Eeden et al., 2003)(Strickland and Bertoni, 2004).

There are two classification systems used in evaluating tremor: type of movement and cause (Bain, 1993). The first is based on whether tremor occurs at rest (resting tremor) or with action. Tremors with movement (action) are subdivided into 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 a highly skilled activity. A postural tremor is usually tested by having a patient hold the arms stretched out in front, while kinetic or intention tremor can be tested by 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 classically characterized as a low amplitude postural tremor with a modal frequency of 8-12 Hz (Elble and Koller, 1990) in the hands, but may be as slow as 6.5 Hz in other body parts (Hallett, 1991). Previous research has demonstrated a decrease in tremor frequency with age (Birmingham et al., 1985)(Kelly et al., 1994)(Marsden et al., 1969)(Wade et al., 1982), especially after the fifth decade, however more recent studies have not found any decline (Elble, 2003)(Raethjen et al., 2000). Pathological tremor is the most extended movement disorder and can be seen in several pathologies, such as: Essential Tremor (ET), Parkinson's Disease (PD), dystonic disorders, cerebellar disease, as a consequence of head trauma or as a symptom of alcohol withdraw. Most of these pathologies are more prevalent among older individuals (Elble, 1998)(Kelly et al., 1994)(Waite et al., 1996) with prevalence rates for ET reaching 6.3% in individuals between 60 and 65 years old (21.7% for people with 95+ years) (Louis and Ferreira, 2010), and the incidence of PD increasing from 0.50/100,000 in individuals aged 30-39 years old to 44/100,000 in individuals with more than 50 years old (Van Den Eeden et al., 2003).

Current gold 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, with higher scores indicating greater severity (Goetz et al., 2008). However, there are certain limitations in the utility of this rating instrument because scores are subjective and imprecise.

Objective motor assessment and continuous patient symptom monitoring is an open challenge for movement disorder specialists (Burkhard et al., 1999). There have been many attempts to objectively rate movement disorder severity. Previous methods include electromyography (EMG) (Bacher et al., 1989)(Carboncini et al., 2001)(Spieker et al., 1998) handwriting and drawing samples (Elble et al., 1990)(Riviere et al., 1997), accelerometer readings (Gerr et al., 2000)(Hoff et al., 2001)(Van Someren, 1997), gyroscope data (Burkhard et al., 1999)(Burkhard et al., 2002)(Salarian et al., 2007), electromagnetic tracking (O'Suilleabhain and Dewey Jr, 2001)(Rajaraman et al., 2000) and a laser system for transducing velocity (Norman et al., 1999). Some methods of tremor quantification, such as electromagnetic and laser systems, are not feasible for home use due to the large and expensive equipment. Also, they require highly trained professionals to interpret results, even if they correlate well with clinician scores (O'Suilleabhain and Dewey Jr, 2001)(Rajaraman et al., 2000)(Elble et al., 1996). Handwriting and drawing samples have long been used to quantify tremor during movement due to their simplicity (Elble et al., 1990)(Riviere et al., 1997); however, it is not applicable for measuring resting or postural tremor. Accelerometers and gyroscopes are currently the most commonly used instruments in tremor studies, since they are capable of providing reliable and objective indexes. Accelerometers measure linear acceleration and are influenced by gravity, whereas gyroscopes measure angular velocity independently of gravity.

Similar to the method employed in our work, many tremor quantification algorithms use power spectral analysis in the frequency domain (Lukhanina et al., 2000)(Riviere et al., 1997)(Gerr et al., 2000)(Hoff et al., 2001)(Salarian et al., 2007) and define tremor amplitude as the amplitude of a peak in the power spectrum in the 3 to 7 Hz range. Also, the logarithm of tremor amplitude has shown to correlate well with UPDRS scores when measured using handwriting samples (Elble et al., 1996) and gyroscopes (Salarian et al., 2007). Other analysis techniques have also been used and include cumulative distance traversed (Rajaraman et al., 2000) and adaptive modeling (Riviere et al., 1997).

Objectively measuring the severity of tremor disorders has been a topic of much research to clinicians in the last decade and it will be of significant relevance to the HCI community as well. With the global increase of the senior population (Watkins et al., 2005), understanding, modeling and dealing with tremor will be a significant concern in designing future assistive technologies.

5.1.2. Older Adults, 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 (Beaton and Weiman, 1984)(Wilson and Liu, 1995)(Martin, 1988)(Bender, 1999)(Colle and Hiszem, 2004)(Sun et al., 2007)(Beringer, 1990)(Scott and Conzola, 1997). Past research has investigated optimal target size, spacing and position. Regarding able-bodied users, several works suggest that targets smaller than 10 mm in width strongly reduced performance (Parhi et al., 2006)(Park et al., 2008)(Lee and Zhai, 2009). The spacing between keys is also an important feature on touch interfaces, yet controversial. Scott and Conzola (Scott and Conzola, 1997) support the use of 2 mm or less spacing in keypad designs. Martin (Martin, 1988) recommends 6 mm spacing for square keys, while Colle and Hiszem (Colle and Hiszem, 2004) suggest 1 mm spacing should be used if sufficient space is available and spacing as small as 0 mm might be acceptable if space is very limited. Regarding on-screen target positioning; users often prefer targets near the center of the screen, because it is easier and more comfortable to tap. However, the highest accuracy and precision rates occur for targets on the edge of the screen (Park et al., 2008)(Guerreiro et al., 2010).

Jin et al. (Jin et al., 2007) investigated optimal target size and spacing for older adults, measuring time, accuracy, and preference. Results show that 19.05 mm targets allow high accuracy rates and reaction times. Regarding preference, older adults prefer targets between 16.51 mm and 19.05 mm.

While the aforementioned studies derive some recommendations and general guidelines on target sizes, spacing and position, they do not consider the particular challenges of text-entry: large number of targets, small key size and narrow spacing. Researchers have explored linguistics models to deal with typing inaccuracy. Kristensson and Zhai (Kristensson and Zhai, 2005) proposed a method whereby the overall geometric shape formed by all hit points for a word is considered in linguistic matching. Gunawardana et al. (Gunawardana et al., 2010) proposed a method to expand or contract key areas for each press based on linguistic models, building on previous work by Goodman et al. (Goodman et al., 2002) and Al Faraj et al. (Al Faraj et al., 2009). In similar work, Himberg et al. proposed adaptation through the movement of individual keys (Himberg et al., 2003). In addition to language models, others have proposed the use of touch models to adapt to individual typing patterns (Goodman et al., 2002)(Rudchenko et al., 2011) and improve overall input accuracy. Alternatives to tapping a virtual QWERTY keyboard have also been proposed, including alternative key layouts (MacKenzie and Zhang, 1999)(Zhai et al., 2000)(Zhai et al., 2002)(Dunlop and Levine, 2012), gestures (Guerreiro et al., 2008)(Tinwala and MacKenzie, 2010), and methods that enable users to stroke between keys (Kristensson and Zhai, 2004)(Zhai and Kristensson, 2003).

Yet, few researchers have explored the specific needs of older adults in text-input tasks. Wobbrock et al. (Wobbrock et al., 2003) proposed a stylus-based approach that uses edges and corners of a reduced touch screen to enable text-entry tasks on a PDA. Results showed that EdgeWrite provides high accuracy and motion stability for users with motor impairments. Similarly, Barrier Pointing (Froehlich et al., 2007) uses screen edges or 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. Others (Mertens et al., 2010)(Wacharamanotham et al., 2011) took a similar approach by proposing a technique that uses swipe gestures towards the screen edges in order to select targets (swabbing).

Although these works insightfully explore the device physical properties to aid impaired people interacting with touchscreens, there is little empirical knowledge about older adults performing text-entry tasks with traditional virtual keyboards. Previous research does not take into consideration older users abilities (such as tremor) that might affect the use of virtual keyboards. The user study reported in this chapter bridges this gap by thoroughly analyzing their performance when typing with both mobile and tablet devices, enabling designers to take advantage of this knowledge when building future virtual keyboards.

5.2. User Study

Accurately selecting targets on current touchscreen devices can be a hard task to accomplish. This is especially relevant in text-entry tasks for older adults, due to both small target size and space between keys. In this user study, we evaluated two different types of touch devices - mobile phone and tablet - and thoroughly analyzed how older users inputted text. Moreover, we analyzed the relationship between hand tremor and typing performance.

5.2.1. Research Questions

We aim to answer the following research questions:

1. *How do older adults type on touchscreens regarding accuracy, speed and typing strategies?*
2. *What are the most common types of errors for older adults? What are their main causes?*
3. *Do tablet devices allow lower error rates?*
4. *Does tremor affect text-entry performance? If yes, how is users' performance correlated with hand tremor?*

5.2.2. Participants

Fifteen participants, 11 females and 4 males, took part in the 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 (Centro de Dia Algueirão Mem-Martins) 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; with corrected vision all participants were able to read the screen content. The only exception was participant #9. She had myopia (also known as nearsightedness) in the right eye and a cataract in the left eye. Although text font was large (Figure 5.2), this participant was unable to correctly perceive the mobile device's characters, and therefore did not perform the user study in this condition. We present her results for the tablet condition; however, she was removed from the within-subjects analysis.

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 never entered text. Regarding QWERTY familiarity, six participants had used this type of keyboard whether in typewriters (4 participants) or personal computers (2 participants). Table 5.1 summarizes all demographic data of participants.

5.2.3. Procedure

The 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 their demographic data and mobile phone usage (see Appendix A5 for more details). We then explained and showed 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 participants were not familiar with touch devices and QWERTY keyboards, we asked them to perform two

Participant	Age	Gender	Mobile experience	QWERTY experience	Touchscreen experience	Visual characteristics
#1	88	Male	Yes	No	No	Corrected vision
#2	78	Female	Yes	Yes	No	Corrected vision
#3	89	Female	No	Yes	No	No correction
#4	72	Female	Yes	No	No	Corrected vision
#5	86	Male	Yes	Yes	No	Corrected vision
#6	81	Male	Yes	Yes	No	Corrected vision
#7	77	Male	Yes	No	No	Corrected vision
#8	84	Female	Yes	No	No	Corrected vision
#9	86	Female	Yes	No	No	No correction: Myopia and cataract
#10	75	Female	No	No	No	Corrected vision
#11	69	Female	Yes	No	No	Corrected vision
#12	67	Female	Yes	Yes	No	No correction
#13	80	Female	Yes	Yes	No	Corrected Vision
#14	86	Female	No	No	No	No correction
#15	70	Female	Yes	No	Yes	Corrected Vision

Table 5.1: Participants profile. Participant #9 was not able to input text in the mobile condition.

familiarization tasks with each device. The first one consisted of entering single letters. Participants had to copy a letter, displayed at the top of the screen, to a text box. If the right key was pressed an earcon was played to convey that the users' action was correct. On the other hand, if the key was not correct, then a different earcon was played to convey that the users' action was incorrect. Participants performed this task for 10 minutes (to guarantee an equal amount of training across individuals). The second task consisted of copying single words and sentences. Participants started by copying one word and if it was correctly transcribed the next sentence had one more word. Error correction (delete) was not available. The sentences had a maximum of 5 words, similar to those presented on the evaluation phase. Participants performed this task for 20 minutes (10 minutes per device).

In the evaluation phase, which occurred in the next day to avoid fatigue, we started by assessing the users capabilities regarding tremor (rest, postural and action tremor) applying three different methods. We first performed a small set of the UPDRS in order to characterize tremor severity (Martinez-Martin et al., 1994); we then asked participants to draw a spiral with each hand without leaning hand or arm on table (Bain et al., 1993); and finally, we 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 (Selker et al., 2011). Participants were then informed about the experiment and how to use our evaluation application. We evaluated participants' performance with two devices on landscape mode: mobile phone and tablet.

Figure 5.2: Screen shot of the test application. Participants were not able to correct errors. The button 'Avançar' allowed them to continue to the next sentence.

Before each condition participants had a 5 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; for the tablet device condition, it was placed in the table in front of them. As a result, all participants consistently used their dominant index finger to select the intended keys.

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 5.2). 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.

We wanted to elicit natural typing behaviors and did not want participants to be concerned with the accuracy of their input. Thus, we followed a similar approach to Gunawardana et al. (Gunawardana et al., 2010) and Goel et al. (Goel et al., 2012), and created a keyboard in such a way that error correction was not available. On the other hand, if the prototype had a delete key it might introduce correcting strategies, which might vary across participants and upset the naturalness of the data. 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 participant, during the evaluation phase, entered a total of 10 different sentences. These sentences were extracted from a written language corpus, and each one had 5 words with an average size of 4.48 characters and a minimum correlation with the language of 0.97 (see Apendix A1). Sentences were chosen randomly and no two sentences were written by the same participant.

5.2.4. Apparatus

An HTC Desire and ASUS EEE Pad Transformer TF101 tablet with a capacitive touch screen were used during the user study. A QWERTY virtual keyboard, similar to Android's traditional keyboard, was used in both devices (Figure 5.2); for the HTC Desire each key was 10x10mm on landscape mode, while for the ASUS tablet each key was 20x10mm. A letter was entered when the user lifted his finger from the key. Neither word prediction nor correction was used. Acceleration data was captured through the mobile device's accelerometer for posterior analysis. The user study was filmed to observe the participants' input behaviors.

5.2.5. Dependent Measures

Performance during the text-entry task was measured by several quantitative variables: *Words Per Minute* (*WPM*), *Minimum String Distance* (*MSD*) error rate, and character-level errors (substitutions - incorrect characters -, insertions - added characters -, and omissions - omitted characters) (MacKenzie and Soukoreff, 2002a). Qualitative measures were also gathered in 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 objectively characterize their level of impairment: 1) *self-reported tremor*; 2) small set of the UPDRS (rest and action tremor); 3) Archimedes spiral test (action tremor); and 4) Postural acceleration sensed by the mobile device accelerometer (action-postural tremor).

5.2.6. 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 10 sentences per participant. In summary the study design was: 14 participants x 5 sentences x 2 devices.

We performed Shapiro-Wilkinson tests of the observed values for *WPM*, *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.

5.3. Results

Our goal was to understand how older adults inputted text with traditional touch-based 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 (mobile and tablet), focusing on type of errors and main causes.

5.3.1. Tremor Profile

For self-reported measures of tremor we used a 5-point likert scale (Goetz et al., 2008) (0 - Absent; 1 - Slight and infrequently present; 2 - Moderate, bothersome to user; 3 - Severe, interferes with many activities; 4 - Marked, interferes with most activities). Regarding self-reported results, 10 (66.7%) participants stated that they had no tremor, 4 (26.7%) had slight and infrequently present tremor, and 1 (6.7%) had severe tremor.

Regarding the UPDRS results, we tested for both resting and action tremor. Resting tremor occurs when the affected part is relaxed, still, and totally supported by gravity and the muscles are not voluntarily actuating (Anouti and Koller, 1995). This tremor usually disappears when a movement or action begins (Wyne, 2005). For that reason, in most cases in which the static tremor is the main symptom, there are no large problems for the user, except for embarrassment. In order to measure resting tremor, we asked participants to place their hands on the table. Only one (6.7%) participant had resting tremor (slight and infrequently present) in both hands, while the remaining did not show any signs of resting tremor (93.3%).

In addition to resting tremor, we also measured action tremor, which occurs during a voluntary muscular contraction. Particularly, we measured the kinetic tremor, which occurs only during the accomplishment of any action with the affected limb (Anouti and Koller, 1995). We asked participants to perform the finger-to-nose maneuver with both hands, i.e. participants had to touch their noses with the index finger. Results showed 7 participants (46.7%) with a slight and infrequently present kinect tremor for both hands. Exception has to be made for only one participant that showed a moderate kinect tremor severity for the non-dominant hand. The remaining exhibited no tremor.

Task-specific tremor, which is a type of action tremor, was also measured in both hands, using the Archimedes spiral test. The drawings were then classified by a clinician as Absent, Slight, Moderate, Severe or Marked. For the dominant-handed drawings, 7 participants (46.7%) showed no tremor, 4 (26.7%) showed slight tremor, 1 participant (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

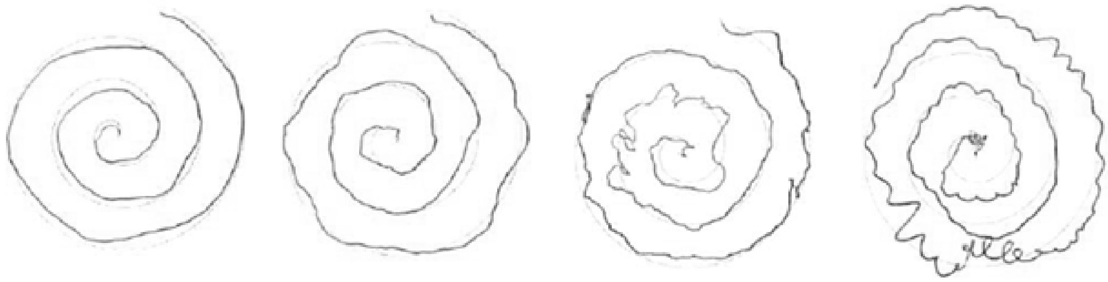


Figure 5.3: Examples of Archimedes spiral drawings.

(6.%) demonstrated moderate tremor, 1 (6.7%) showed severe tremor, and 2 participants (13.3%) demonstrated marked tremor.

Figure 5.3 illustrates some examples of the Archimedes spiral test. Table 5.2 summarizes the subjective tremor assessment.

Participant	Self-reported	Rest (both hands)	Kinect Action (right/left hand)	Task Specific Action (right/left hand)
#1	Absent	Absent	Slight	Absent/Absent
#2	Slight	Absent	Slight	Absent/ Absent
#3	Absent	Absent	Absent	Slight/Slight
#4	Slight	Absent	Absent	Slight/Moderate
#5	Absent	Absent	Slight	Moderate/Marked
#6	Absent	Absent	Absent	Absent/Slight
#7	Slight	Absent	Slight	Slight/Slight
#8	Slight	Absent	Slight	Moderate/Severe
#9	Severe	Slight	Slight	Severe/Marked
#10	Absent	Absent	Absent	Absent/ Absent
#11	Absent	Absent	Absent	Slight/Slight
#12	Absent	Absent	Slight/Moderate	Slight/Slight
#13	Absent	Absent	Absent	Absent/Slight
#14	Absent	Absent	Absent	Absent/ Absent
#15	Absent	Absent	Absent	Absent/ Absent

Table 5.2: Subjective tremor results for each participant.

In addition to subjective measures, we also measured tremor objectively, through the device's accelerometer. Particularly, we measured the postural - a type of action - tremor, by asking participants to hold the mobile device at the arm's length for 30 seconds with each hand and remain still (Selker et al., 2011). From the captured data we analyzed four main values: the acceleration standard deviations, which correspond to hand oscillations (Bergstrom-Lehtovirta et al., 2011); and the peak amplitude in the power spectrum in the 3 to 7 Hz, 7 to 12 Hz and 1 to 12 Hz range. We report the peak amplitude in different frequency ranges since physiological and pathologic tremors (such as Parkinson's disease) are usually distinguishable. Nevertheless, we also report the peak amplitude in all frequencies spectrum, which for our participants is always located between 1 and 12 Hz.

Participant	X	Y	Z	XYZ
#1	0.18	0.11	0.16	0.12
#2	0.22	0.1	0.19	0.08
#3	0.2	0.1	0.16	0.1
#4	0.12	0.14	0.24	0.15
#5	0.33	0.14	0.3	0.16
#6	0.32	0.15	0.23	0.14
#7	0.23	0.16	0.29	0.18
#8	0.23	0.2	0.44	0.22
#9	0.17	0.24	0.43	0.23
#10	0.12	0.09	0.17	0.1
#11	0.1	0.14	0.25	0.09
#12	0.11	0.12	0.44	0.11
#13	0.22	0.32	0.62	0.15
#14	0.13	0.12	0.21	0.12
#15	0.11	0.11	0.31	0.11

Table 5.3: Dominant hand oscillation (in m/s^2) for each axis and participant.

Results for hand oscillation showed a mean magnitude of 0.186 m/s^2 (sd=0.074), 0.15 m/s^2 (sd=0.63), 0.3 m/s^2 (sd=0.13), and 0.137 m/s^2 (sd=0.044) for X, Y, Z, and XYZ axis, respectively. Table 5.3 illustrates the oscillation for the dominant hand for each axis and participant.

Regarding the non-dominant hand, due to a logging bug we were only able to save 9 of the 15 participants' accelerometer data. Mean oscillation was 0.174 m/s^2 (sd=0.07), 0.115 m/s^2 (sd=0.024), 0.3 m/s^2 (sd=0.149), and 0.101 m/s^2 (sd=0.03), for X, Y, Z, XYZ axis, respectively. Table 5.4 illustrates the oscillation for the non-dominant hand for each axis and participant.

Participant	X	Y	Z	XYZ
#4	0.18	0.09	0.16	0.1
#5	0.33	0.14	0.3	0.16
#6	0.2	0.11	0.19	0.08
#8	0.17	0.14	0.47	0.14
#10	0.18	0.14	0.53	0.09
#11	0.19	0.11	0.22	0.07
#12	0.12	0.12	0.44	0.1
#14	0.09	0.09	0.11	0.08
#15	0.11	0.08	0.25	0.08

Table 5.4: Non-dominant hand oscillation (in m/s^2) for each axis and participant.

Regarding the frequency analysis, results showed a mean peak magnitude of 0.362 m/s^2 (sd=0.429), 0.17 m/s^2 (sd=0.17), 0.489 m/s^2 (sd=0.445) for the 3 to 7 Hz, 7 to 12 Hz and 1 to 12 Hz, respectively. It is worth noticing that the results for each of the frequency ranges show high standard deviations, suggesting that tremor severity varies widely among participants. Concerning the non-dominant hand, results showed a mean peak magnitude of 0.17 m/s^2 (sd=0.162), 0.105 m/s^2 (sd=0.175), 0.287 m/s^2 (sd=0.189) for the 3 to 7 Hz, 7 to 12 Hz and 1 to 12 Hz, respectively. Again, as in the dominant hand results, tremor magnitudes for each of the frequency ranges show high standard deviations, suggesting that tremor severity is highly user dependent. As we can see, from all tremor profile results, our participants formed a heterogeneous user group; some with no visible tremor while others experienced marked tremor.

Therefore, we investigated the relationship between age and each of the tremor measures. Preliminary analysis was performed to assess normality, linearity and homoscedasticity. Pearson product-moment correlation was performed on variables that did not violate any of its assumptions previous described (Oscillation X and Oscillation XYZ), while Spearman correlation was used for the remaining (self-report, UPDRS Rest, UPDRS Kinect, Task-specific, Oscillation Y, Oscillation, Z, Magnitude 3-7 Hz, Magnitude 7-12 Hz, and Magnitude 1-12 Hz). Table 5.5 illustrates the correlation coefficients and significance levels. Although we present this analysis, our goal in this user study was not to see in detail how these measures were correlated, nor could we do it due to the small number of participants. Instead, we are just reporting theses values so the reader can gain some understanding of our data.

Results show that all subjective measures present a small correlation with age. On the other hand, objective measures show a higher correlation level. Particularly, there was a strong, positive correlation between *age* and *hand oscillation on the X axis* [Spearman rho=0.561, n=15, $p<0.05$] (with 31.47% of shared variance), with high oscillation levels on older people. Also, there was a medium correlation between *age* and *hand oscillation on the Z and XYZ axis*, and *peak magnitude on the 3 to 7 Hz* range.

	Effect of Age	Age (r/rho, sig)	
Subjective measures	Self-report	0.013	0.963
	UPDRS Rest	0.248	0.372
	UPDRS Kinect	0.186	0.506
	Task Specific	0.096	0.735
Objective measures	Oscillation X	0.561	0.029
	Oscillation Y	0.036	0.899
	Oscillation Z	-0.358	0.19
	Oscillation XYZ	0.366	0.179
	Magnitude 3-7 Hz	0.325	0.238
	Magnitude 7-12 Hz	0.236	0.397
	Magnitude 1-12 Hz	-0.091	0.746

Table 5.5: Correlation coefficients and significance levels between age and tremor measures.

Moreover, we investigated the relationship between perceived tremor (both from participants and experimenter) and objectively measured tremor. Therefore, a correlation analysis was performed between subjective and objective measures, using Spearman correction coefficient. Table 5.5 shows all coefficients and significance levels between the variables.

Self-reported measures showed to be strongly and positively correlated with the *peak acceleration magnitude in the 1 to 12 Hz* range [Spearman rho=0.558, n=15, $p<0.05$] (31.14% of shared variance). On the other hand, there were no strong correlation between *Rest Tremor* and any of the objectively measured variables. Regarding the *Kinect tremor* there was a strong, positive correlation with the *peak acceleration magnitude in the 7-12 Hz* range [Spearman rho=0.59, n=15, $p<0.05$], which is often associated with physiological tremor (Elble and Koller, 1990). Finally, *task-specific tremor* measures were strongly and positively correlated with several objective measures: *oscillation XYZ* [Spearman rho=0.527, n=15, p=0.043] (27.77% of shared variance), *peak magnitude in the 3 to 7 Hz* range [Spearman rho=0.584, n=15, p=0.022] (34.12% of shared variance), *peak magnitude in the 7 to 12 Hz* range [Spearman rho=0.618, n=15, p=0.014] (38.19% of shared variance), and *peak magnitude in the 1 to 12 Hz* range [Spearman rho=0.706, n=15, p=0.003] (49.84% of shared variance).

The strong relationship between some subjective and objective measures, suggest that results for one given test can be predicted by other. For instance, task-specific tremor could be predicted by observing the peak acceleration magnitude between 1 and 12 Hz.

(r/rho, sig)	Self-report		UPDRS Rest		UPDRS Kinect		Task Specific	
Oscillation X	0.23	0.409	-0.062	0.827	0.402	0.137	0.154	0.583
Oscillation Y	0.426	0.113	0.371	0.173	0.217	0.438	0.466	0.08
Oscillation Z	0.194	0.489	0.247	0.374	0.217	0.438	0.411	0.128
Oscillation XYZ	0.467	0.079	0.433	0.107	0.34	0.215	0.527	0.043
Magnitude 3-7 Hz	0.498	0.059	0.433	0.107	0.449	0.093	0.584	0.022
Magnitude 7-12 Hz	0.459	0.086	0.373	0.171	0.59	0.021	0.618	0.014
Magnitude 1-12 Hz	0.558	0.031	0.433	0.107	0.325	0.237	0.706	0.003

Table 5.6: Correlation coefficients and significance levels between subjective and objective tremor measures.

5.3.2. Input Speed

In this section we thoroughly analyze input performance regarding speed for both device conditions (mobile and tablet). To assess speed, we used the *Words Per Minute* (*WPM*) (MacKenzie and Soukoreff, 2002b) text input measure calculated as:

$$(transcribed\ text\ length - 1) \times (60\ seconds \div time\ in\ seconds) \div 5\ characters\ per\ word$$

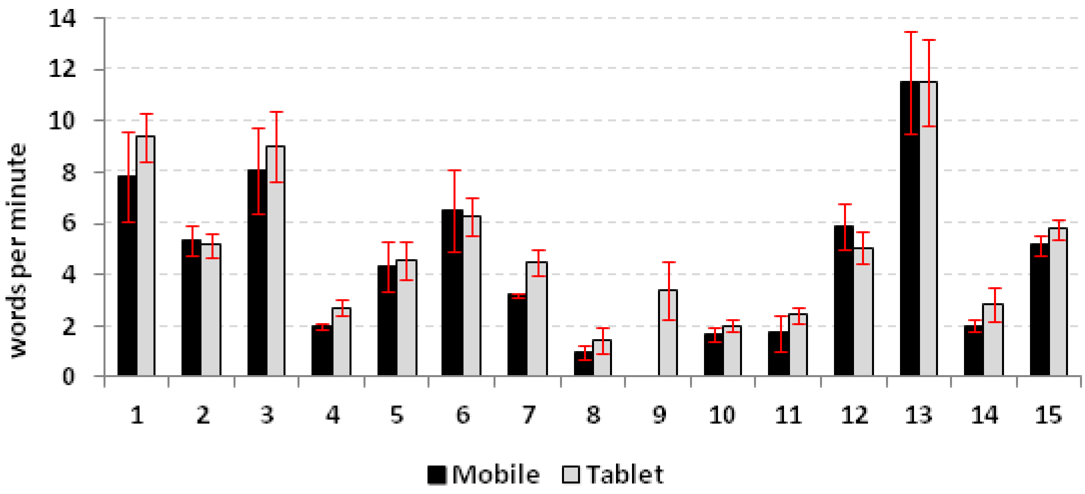


Figure 5.4: Words per minute for each participant and device condition. Error bars denote 95% confidence intervals.

Figure 5.4 shows the participants' average WPM for mobile and tablet conditions.

Mobile device. Participants obtained a mean 4.73 (sd=3.06) wpm using the mobile virtual keyboard. Notice that participants #4, #7, #8, #10, #11, #14 stayed well below the average (Figure 5.4). All these participants never used a QWERTY keyboard before (except for the training session), which can explain the low input rate. Indeed, there was a strong positive relationship [Spearman rho=.648, n=14, p<.05] between *QWERTY experience* and *input rate*; that is, participants that used a QWERTY keyboard before, inputted text faster (46.24% of shared variance). Regarding tremor (dominant hand), the peak acceleration magnitude between 1 and 12 Hz also had a relevant relationship with input speed. Spearman's coefficient revealed a medium negative correlation between the two variables [Spearman rho=-.418, n=14, p=.137], with low peak magnitudes (lower tremor) associated with higher input rates. Other measures of tremor also showed a medium negative correlation with *WPM*: *self-reported tremor* [Spearman rho=-.392, n=14, p=.165], *task-specific tremor* [Spearman rho=-.397, n=14, p=.159]. Since the mobile device was held during the text-entry task, we also analyzed the relationship between the participants' performance and non-dominant hand tremor. Generally, all measures had low correlations. The only exception was the peak acceleration magnitude between 1 and 12 Hz [Spearman rho=-.393, n=9, p=.235] with a medium correlation, however with low significance.

Tablet device. Regarding performance using the tablet device, participants achieved a mean 5.07 (sd=2.93) wpm. The overall pattern is similar to the mobile device conditions; that is, participants with low input rate (#4, #8, #10, #11, and #14) never used a QWERTY keyboard previously. In fact, we found a strong positive correlation between *QWERTY experience* and *input speed* [Spearman rho=.534, n=15, p<.05]. There were also several tremor measures (for the dominant hand) with a medium inverse correlation: *self-*

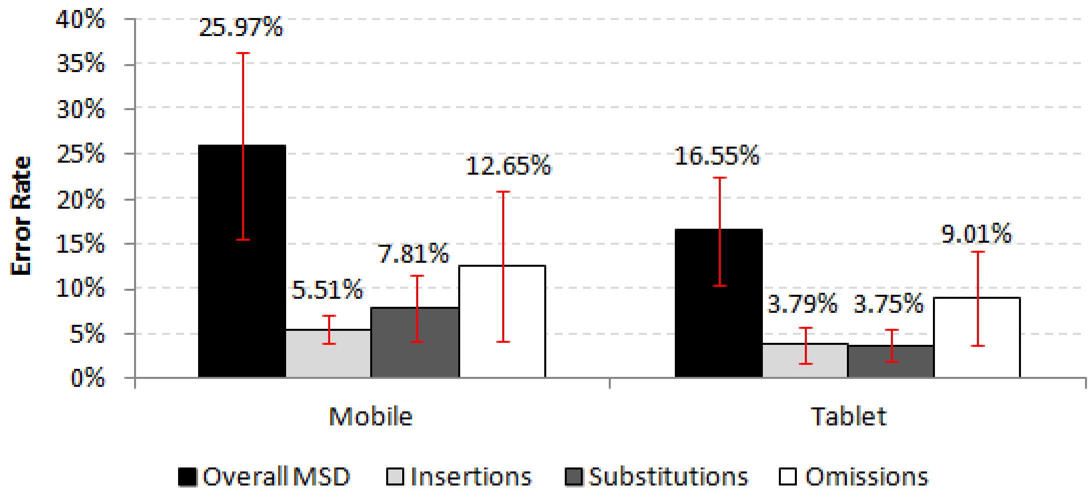


Figure 5.5: Overall MSD, insertion, substitution, and omission error rate for each device condition. Error bars denote 95% confidence intervals.

report tremor [Spearman rho=-.383, n=15, p=.158], *UPDRS Kinect tremor* [Spearman rho=-.31, n=15, p=.913], *task-specific tremor* [Spearman rho=-.459, n=15, p=.117] and *peak acceleration magnitude between 1 and 12 Hz* [Spearman rho=-.379, n=15, p=.164]. These results suggest that tremor is indeed related with mobile input rate; particularly, higher tremor is associated with lower input rates.

Summary. Overall, input rate was strongly correlated with previous experience with QWERTY keyboard, which explains 46% and 29% of shared variance for mobile and tablet conditions, respectively. A paired-samples t-test was conducted to evaluate the effect of device on text-entry speed. We found a statistically significant increase in wpm [t(13) = -2.752, p<.05] with a large effect size (eta square statistic =.4), suggesting that participants can achieve higher input rates with tablet devices.

5.3.3. Quality of Transcribed Sentences

The quality of the transcribed sentences was measured using the *Minimum String Distance* (*MSD*) error rate, calculated as:

$$MSD(required\, text, transcribed\, text) \div mean\; size\; of\; the\; alignments \times 100$$

Figure 5.5 illustrates participants' MSD error rate for both mobile and tablet conditions along with error types (insertions, substitutions and omissions). In this section we will only analyze the MSD error rate for both device conditions.

Mobile device. Participants achieved an average MSD error rate of 25.97% (sd=19.72%)

in the mobile device condition. Most participants (#1, #2, #6, #7, #12, #14, #15) had an average MSD ER between 0 and 20%; five (#3, #4, #5, #11, #13) were between 20% and 40%; one (#10) between 40% and 60%; and one (#8) between 60% and 80%. As opposed to the results obtained on input speed, there was a weak correlation between quality of transcribed sentences and *QWERTY experience* [Pearson r=.145, n=14, p=.621]. Instead, there was a higher correlation with tremor measures. Particularly, task-specific tremor for both dominant [Spearman rho=.505, n=14, p=.066] and non-dominant [Spearman rho=.506, n=14, p=.065] hands was strongly correlated with *MSD* error rate and accounts for approximately 26% of shared variance. On the other hand, objective measures of tremor had an even higher correlation coefficient. *Hand oscillation* of the non-dominant hand in the Y [Pearson r=.751, n=9, p<.05] and Z [Pearson r=.613, n=9, p=.079] axes 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 37.6% - 56.4% of shared variance, which is higher than the results obtained for the dominant hand.

Tablet device. In this condition, participants had an average MSD error rate of 16.55% (sd=11.9%). Eleven participants (#1, #2, #4, #5, #6, #7, #9, #10, #12, #14, #15) achieved a mean MSD ER between 0 and 20%, 3 participants (#3, #11, #13) obtained results between 20% and 40%, while one (#8) was between 40% and 50%. As in the mobile device condition, there was only a small correlation between *MSD* error rate and *QWERTY experience* [Pearson r=.155, n=15, p=.58]. Contrary to input speed, the quality of transcribed text cannot be explained by previous experience with keyboards. Moreover, both subjective and objective measures of tremor did not show any strong correlation with *MSD* error rate in this device condition (tablet).

Summary. Overall, *MSD* error rate was not correlated with any specific demographic measure, particularly *QWERTY experience*, which suggests that previous experience with QWERTY keyboards is not enough to compensate for typing errors. However, tremor measures had a higher relevance, particularly in the mobile device condition, explaining nearly 26% of performance variance. Comparing error rates between conditions, there was a statistically significant decrease of 9.42% from mobile to tablet device [Z=-2.417, p<.05] with a medium effect (eta squared statistic = .3). This suggests that older users indeed benefit from tablet devices, either due to the larger key sizes or its static position (on the table).

5.3.4. Typing Errors

This section presents a fine grain analysis by categorizing the types of input errors: insertions - added characters; substitutions - incorrect characters; and omissions - omitted characters (MacKenzie and Soukoreff, 2002a). Figure 5.6 and Figure 5.7 shows the type of errors performed on mobile and tablet devices, respectively, for each participant.

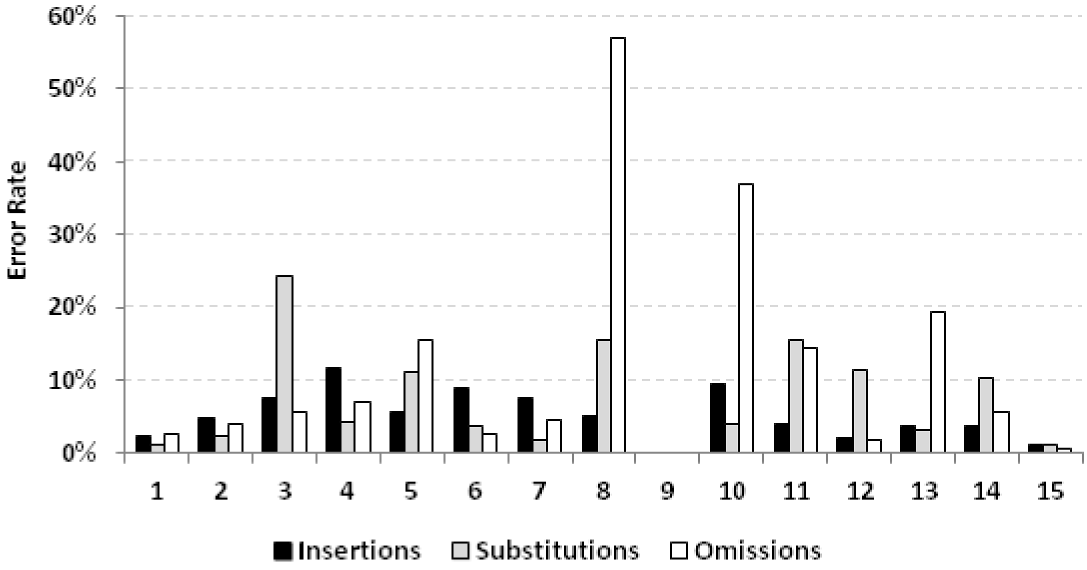


Figure 5.6: Types of error for each participant in mobile condition.

Mobile device. Results show that on average 6% (sd=3%) of input errors are insertions. Overall, this is the least common error type (although no significant differences were found to other types of errors). Moreover, generally speaking, correlations between *insertion* errors and tremor measures (both objective and subjective) were weak (r<0.3). However, there were some medium correlations: *self-report tremor* [Spearman rho=.392, n=15, p=.165], right-hand *task specific tremor* [Spearman rho=.312, n=15, p=.278], left-hand *task specific tremor* [Spearman rho=.401, n=9, p=.156], left-hand *oscillation X* [Pearson r=.356, n=9, p=.348], and left-hand *oscillation XYZ* [Spearman rho=.333, n=9, p=.381] axes.

Regarding substitution errors, participants obtained a mean 7.8% (sd=7%) error rate. This was the second most common error type and consisted in pressing incorrect keys. Results show a high standard deviation, which reflects difficulties of some participants (see Figure 5.5). While eight participants had a substitution error rate between 0 and 5% (#1, #2, #4, #6, #7, #10, #13, #15), six achieved an error rates higher than 10% (#3, #5, #8, #11, #12, #14). Moreover, we found a large positive correlation between *substitution* error rate and both right [Spearman rho=.624, n=15, p<.05] and left [Spearman rho=.541, n=9, p<.05] hand *task-specific tremor*, which accounts for 39% and 29% of shared variance, respectively. Indeed, visual inspection of participants' profile revealed that only one of participants with average substitution error rate above 10% was absent of tremor; all remaining had slight or moderate task-specific tremor. Moreover, only one of the participants (#4) with (moderate) tremor did not obtain a mean substitution error rate above 10%; instead their main cause of errors were insertions (Figure 5.6).

Concerning omission error rates, results show an average of 12.65% (sd=16%). This error type is often described as a cognitive error, since it does not depend on motor abilities

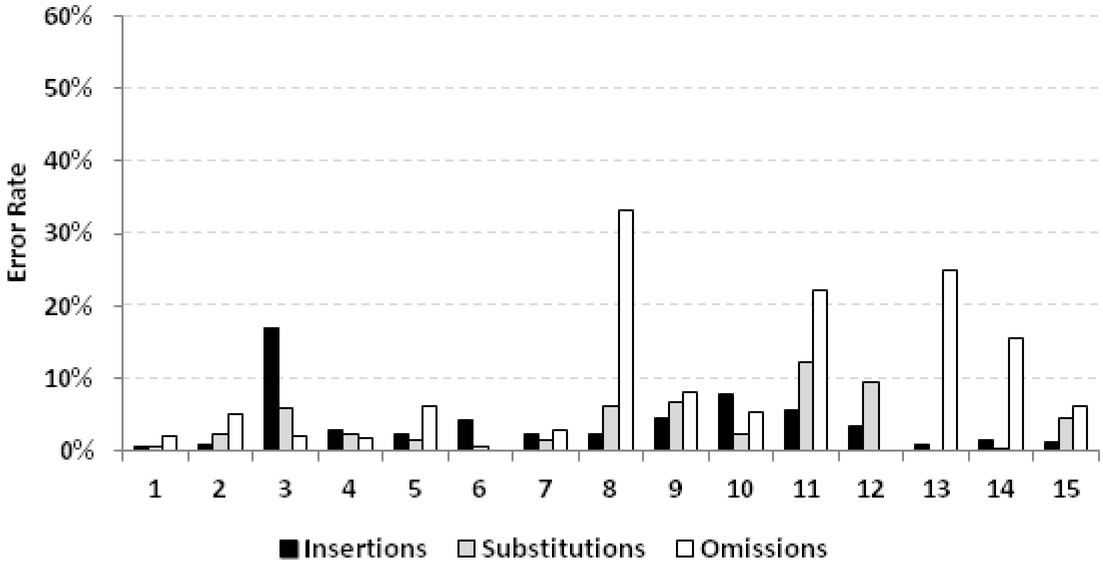


Figure 5.7: Types of error for each participant in tablet condition.

(Kristensson, 2009). Instead, users forget to type the intended characters or misunderstand the required sentence. Although we did not account for cognitive differences and therefore cannot confirm this hypothesis, we did find a large positive correlation between *omission* error rate and left-hand *oscillation Y* [Spearman rho=.733, n=9, p=.025], which suggest that this error type may (also) be related to tremor. Also, we should notice that the high variance (16%) is suggestive of a user-dependent measure. Indeed, most participants (nine) obtained an error rate below 10%, while participant #8 and #10 obtained a 57% and 37% error rate, respectively (Figure 5.6). These results will be further investigated in Section 5.3.7 (Omission Errors in Detail).

Tablet device. Participants obtained an average of 3.8% insertion rate (sd=4.1%) in the tablet condition. As in other text-entry measurements, insertion error rate shows a high variance. Particularly, three participants achieved an error rate above 5%: participants #3 (16.89%), #10 (7.75%), and #11 (5.48%). A medium size positive correlation was found between *insertion* error rate and *task-specific tremor* [Spearman rho=.398, n=15, p=.142]. All other tremor measures were weakly correlated with user performance, suggesting a weak relationship between insertion rates and tremor when using tablet devices.

Regarding substitution error rate, participants achieved a mean of 3.75%, however with high variance (sd=3.61%). While nine participants are below 3% error rate, six (#3, #8, #9, #11, #12, #15) are between 3% and 13% (see Figure 5.7), resulting in a user-dependent measure. Indeed, we found a large positive correlation with *task-specific tremor* [Spearman rho=.539, n=15, p=.038]; that is, participants with higher tremor had higher substitution error rates.

As for omission error rate, this is also the most common error type in the tablet condition, accounting for an average of 9% (sd=10%). Again, results show high variance with nine participants below 10% error rate and four participants (#8, #11, #13, and #14) above 15%. No large correlations were found between *omission* error rate and both subjective and objective tremor measures in the tablet condition.

Summary. Overall, results for each type of error showed a high variance, which suggests that older adults text-entry performance is highly user-dependent. As for insertion errors, we found small relationships with all tremor measurements, thus its variance could not be explained by participants' motor abilities. Nevertheless, we found significant differences between *device* conditions [Z=-2.103, $p<0.05$, eta squared statistic=.25], in which tablet devices show a decrease of 1.72% of insertion errors. Regarding *substitution* error rate there was a large correlation with task-specific tremor for both device conditions, explaining 29-39% of users' performance. Also, we found a statistically significant decrease of 4% from *mobile* to *tablet* device [Z=-2.731, $p<.01$] with a medium effect (eta squared statistic = .36). Concerning omission error rate results are inconclusive. This error type is often associated with cognitive errors, thus it would be expected that performance across device conditions would be similar. Indeed, no statistical significant differences were found between *mobile* (m=12.65%, sd=16%) and *tablet* (m=9%, sd=10%) conditions [Z=-.722, $p>.4$]. On the other hand, we found a large correlation between *omission* error rate and *hand oscillation* on mobile condition, which suggests that this measure may be related to motor abilities. We will further explore these results on Section 5.3.7(Omission Errors in Detail), analyzing each participant individually.

5.3.5. Substitution Errors

Mobile device. In general, participants had similar difficulties across all keys. All characters are below 15% error rate and none highlights as the most problematic. Moreover, no row, column or side patterns emerge from Figure 5.8.

In order to analyze the most common substitutions, we created confusion matrices (MacKenzie and Soukoreff, 2002a). Some of the most frequent errors were: C → SPACE (6.83%), C → V (3.17%), O → P (4%), P → Q (3.83%), R → T (3.83%), S → Z (4.34%), S → D (2.9%), T → Y (3.96%). As we can see there is a clear predominance of right or bottom key substitutions in the data, which suggests that participants found it easier to hit keys in the right-bottom (southeast) direction. Additionally, errors are at a distance of one key. Indeed, this pattern can be seen in Figure 5.9, which illustrates all lift points of mobile condition. These findings may be related to hand dominance, but further investigation is needed to confirm this hypothesis.

A common error that cannot be explained by this pattern is the substitution of P → Q. We believe that this error is not related with motor abilities. Since characters are very similar, indeed they are symmetrical (p → q), this error can be either a visual or cognitive

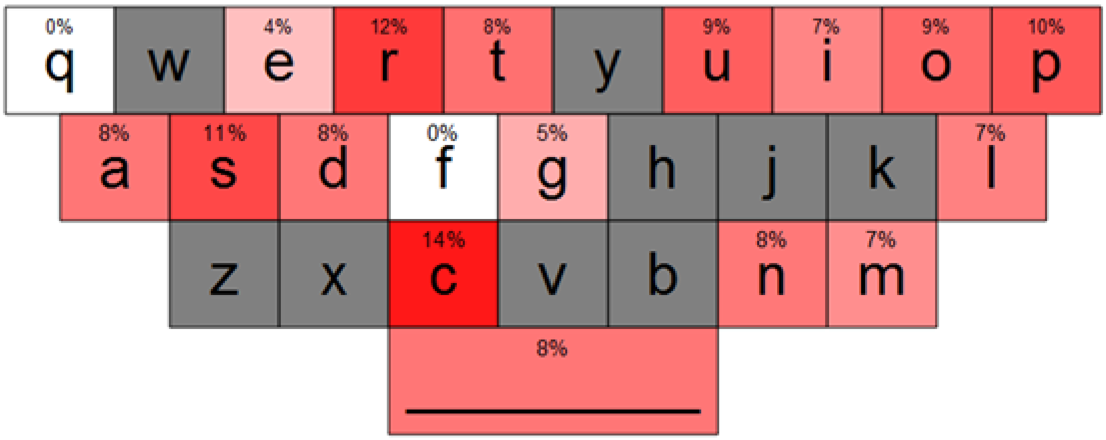


Figure 5.8: Substitution error rate per key with mobile device. Key's "redness" illustrates higher error rates. Grey cells represent keys that were not used in the experiment.

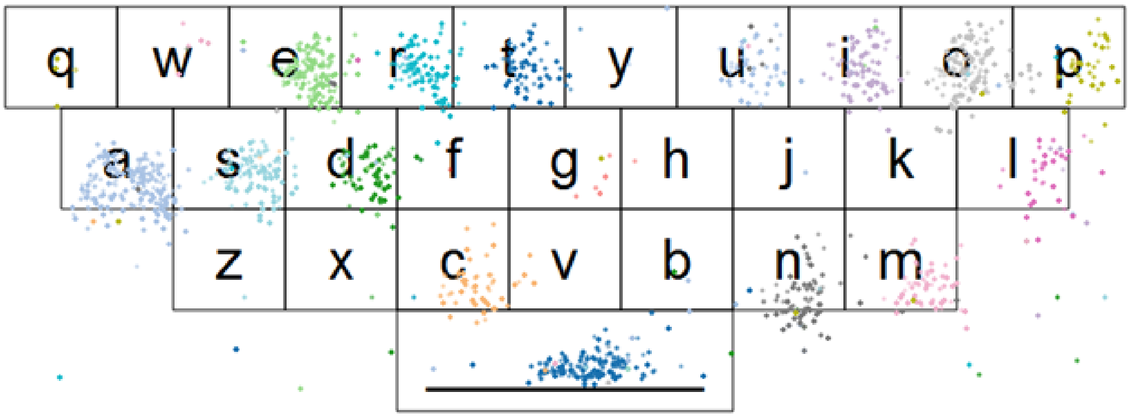


Figure 5.9: Touch (lift) points for all participants with mobile device.

error; that is, confusion may be due to visual perception - as older participants tried to copy exactly what they perceived from the required sentence - or due to an improper mental model of the letter.

Similar problems occurred with the letter 'i', which was frequently replaced by the letter 'l' (3.1%). Participant #11 consistently replaced symmetrical letters: m → w (66.7%), n → u (52.85%). These results suggest that some substitutions are not due to motor errors alone; visual perception and cognitive errors may play an important role in text input for older adults and should be taken into account.

In this user study, we were also interested in finding why substitution errors 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. Figure 5.10 shows the proportion of correct and incorrect land-on targets for substitution errors. As shown, most errors were due to incorrect land-on (i.e. poor aiming), with an average of 6.71% (sd=5.65%). On the other hand, *slips* accounted for an average of 1.1% (sd=2.1%),

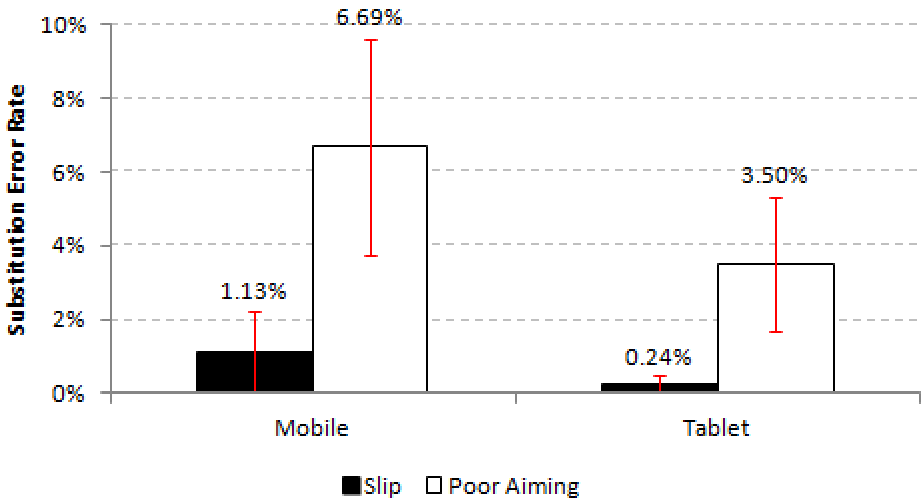


Figure 5.10: Land-on error rate across conditions. Error bars denote 95% confidence intervals.

which was significantly lower than *poor aiming* errors [Z=-3.107, p<0.01, eta squared statistic=.43]. In fact, only six participants performed slip errors; yet, we found large positive correlations between *slip* error rate and: right-hand *task-specific tremor* [Spearman rho=.609, n=14, p=.021], left-hand *oscillation on XYZ axis* [Spearman rho=.714, n=9, p=.031], left-hand *peak acceleration magnitude between 1-12 Hz* [Spearman rho=.752, n=14, p=.02]. Tremor measures accounted for 37-56.6% of slip error rate variance. Regarding *poor aiming* error rate we found a large correlation with right-hand *task-specific tremor* [Spearman rho=.541, n=14, p=.046], although lower than slip error rate.

Moreover, we were interested in the overall virtual keyboard layout that would emerge from older adults' touch points. For this analysis, we calculated key centroids for each key across all participants, which are shown in Figure 5.11. Clear cases of visual/cognitive errors (substitution of 'q' for 'p') or transposition errors (typing 'e' then 'm' instead of the opposite) were identified. To account for these types of errors, we removed outlying points for each key that were more than one key distance away from the center of the required key in either x or y direction (31.44%). Contour ellipses illustrate standard deviations in x and y axis. As a result, the overall keyboard shape is shifted to the bottom right in comparison to the traditional QWERTY keyboard (light-grey lines), which was expected from previous analysis.

From Figure 5.11, we can see that some keys have larger deviations, thus a more detailed analysis of the spread of hit keys would allow us to identify individual keys that may benefit from an increase in size. We calculated the standard deviation of finger-lift points for each key per participant 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.

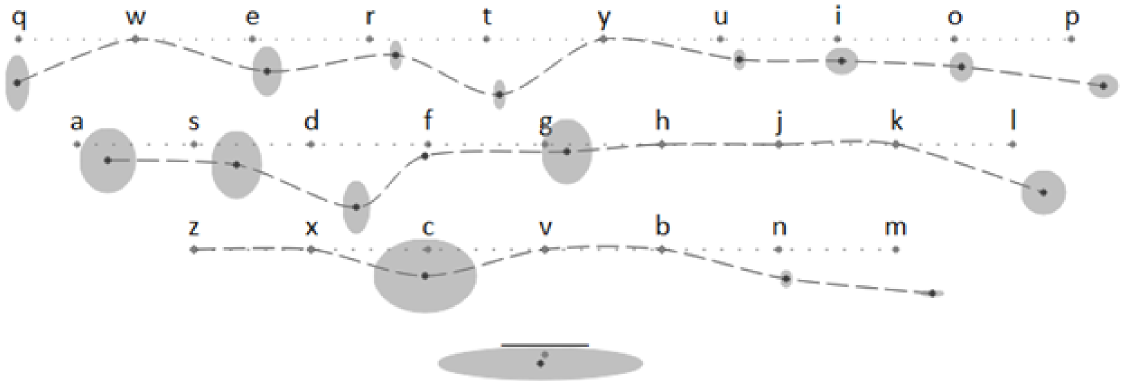


Figure 5.11: Centroids of mobile key hits from all participants with contour ellipses of standard deviations in x and y directions.

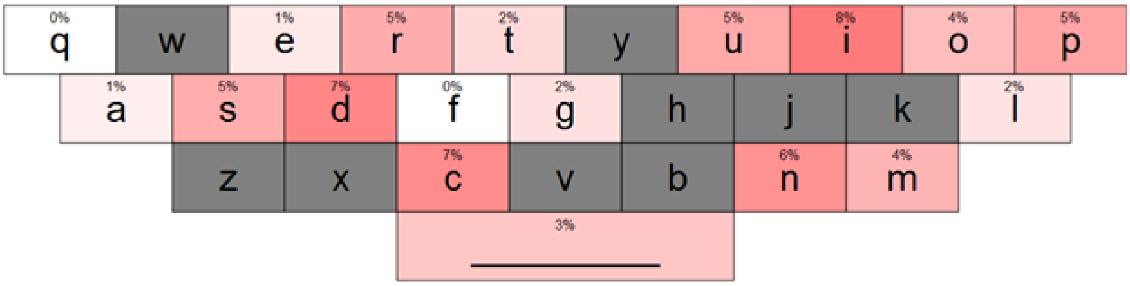


Figure 5.12: Substitution error rate per key with tablet device. Key's "redness" illustrates higher error rates. Grey cells represent keys that were not used in the experiment.

There was no significant effect of row on deviations for both x and y directions. However, we found a significant decrease of deviations from left to right side of keyboard for the *x-direction* [t(13)=-3.043, $p<.01$, eta square statistics=.42]. This finding suggests that keys on the left side of the keyboard should be slightly wider (see Figure 5.11 - keyboard layout). Again, this may be related with hand dominance.

Additionally to this analysis, we calculated the distance to the center of intended key for each centroid per participant in x and y directions. However, we did not find any significant effects on distance to center of key for row or keyboard side.

Tablet device. Overall, participants obtained a substitution error rate below 10% for all keys. Moreover, analysis of keyboard layout did not reveal any clear patterns of substitution errors (Figure 5.12). This result suggests that participants had similar difficulties across all keys. Nevertheless, some of the most common substitution errors were: C $\rightarrow$ SPACE (6.8%), P $\rightarrow$ Q (3.8%), R $\rightarrow$ T (4.2%), S $\rightarrow$ Z (4.3%), S $\rightarrow$ D (3%), U $\rightarrow$ J (1.9%). As in the mobile conditions, there is a predominance of bottom or right key substitutions. Also, most substitutions are at a distance of one key. Non motor-related errors occurred again as letters were replaced with symmetrical ones (e.g. p $\rightarrow$ q). As expected these errors seem to be device independent. Figure 5.14 illustrates touch points for all text-entry trials in the tablet condition.

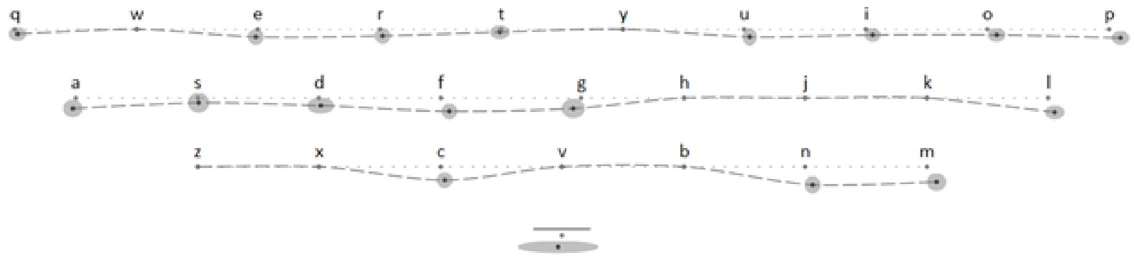


Figure 5.13: Centroids of tablet key hits from all participants with contour ellipses of standard deviations in x and y directions.

Most substitution errors were due to poor aiming rather than finger slips as shown in Figure 5.10. On average, 3.5% (sd=3.5) of substitutions occurred when participants already landed on incorrect keys, while only 0.24% (sd=0.4%) occurred when participants landed on correct keys. This difference has shown to be statistically significant [Z=-2.944, $p<.01$, eta squared statistics=.4]. This means that finger slips account for a minority of substitution errors; that is, it is easier to land on an incorrect key than performing a finger slip. Moreover, *poor aiming* errors are strongly correlated with *task-specific tremor* [Spearman rho=.563, n=14, $p=.029$], accounting for 31.7% of shared variance.

Regarding keyboard layout, we performed a similar analysis to the one described in the mobile condition subsection; that is, key centroids were calculated and outlier points were removed (7.95%). Figure 5.13 illustrates the new keyboard layout. As in the mobile condition, the overall keyboard shape is slightly shifted to the bottom right relative to the standard QWERTY keyboard.

Although we did not find any effect of row or keyboard side, we found a statistically significant increase of x-axis dispersion relative to *y-direction* [t(14)=4.039, $p<.001$, eta square statistics=.56]. This finding suggests that users are more susceptible to x-direction deviations from their touch centroids. We also analyzed distance to center of key and found a significant effect of row on *y-distance* [$F_1.404, 19.653=27.350$, $p<.001$]. Post-hoc analysis revealed significant differences between all rows: mean distance of 8.5 pixels (sd=4.4), 11.5 pixels (sd=6.2), and 16.9 pixels (sd=7.7) for first, second and third row, respectively. Top row centroids presented smaller distances to the center of keys. Additionally, we also found an effect of *keyboard side* on both *x-* [Z=-2.953, $p<.01$, eta squared statistics=.4] and *y-* [t(14)=2.468, $p<.05$, eta squared statistics=.32] *distance directions* in which right-side centroids are closer to the key center. This suggests that participants are more precise when selecting top-right keys. Hand dominance may be related to these findings, however further research is needed to validate this hypothesis.

Summary. In both device conditions, we found a consistent substitution pattern. Letters were mainly substituted by their bottom or right side keys. We hypothesized that it may be related to hand dominance, however our data is insufficient to confirm this. Additionally, we found that participants replace symmetrical or very identical letters, such as p $\rightarrow$ q

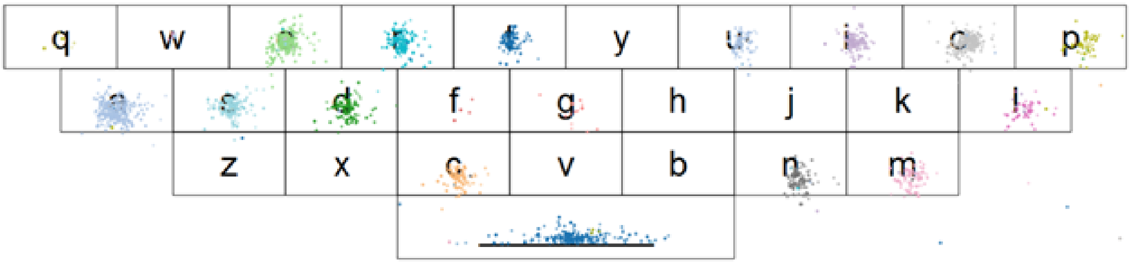


Figure 5.14: Touch (lift) points for all participants with tablet device.

and i → l. These are called non-motor errors and seem to be device independent; that is, users either perceive the characters incorrectly or may have an improper mental model of the required letter. Since we did not account for either of these hypothesis, we are not able to confirm them.

Other similarity between device conditions was the cause of substitutions. In both cases, most substitutions are due to poor aiming, i.e. participants already landed on the wrong keys. Similar results were seen in situationally impaired users (see Chapter 4). The main differences between device conditions are the magnitude of errors and relationship between poor aiming and slip errors. While in the mobile condition there are 5.9 times more poor aiming errors, in the tablet conditions this ratio increases for 14 times (Figure 5.10). Indeed there was a significant decrease of *slip* [Z=-1.826, p<.1, eta squared statistics=.2] and *poor aiming* errors [Z=-2.497, p<.05, eta squared statistics=.32] in the *tablet* condition. This result means that both error types are less probable to occur whilst typing on a tablet device, especially slips.

Regarding obtained keyboard layout, based on calculated key centroids, we found a downward and slightly to the right deviation for both device conditions. Nevertheless, dispersion, particularly in the *y-direction* [t(13)=5.835, p<.001, eta squared statistics=.7] and *distance to the center of intended key* [Z=-2.726, p<.01, eta squared statistics=.36] are significantly lower on *tablet* devices (see Figure 5.11 and Figure 5.13). This may be due to the larger targets and/or device positioning; i.e. in the tablet condition, the device was placed in front of participants in a static position, while the mobile device was held during text-entry tasks, thus more susceptible to hand tremor.

5.3.6. Insertion Errors

Participants obtained an average insertion error rate of 6% (sd=3%) and 3.8% (sd=4.1%) for the mobile and tablet device conditions, respectively. Moreover, we found that this type of error is device-dependent since significant differences were found between conditions [Z=-2.103, p<0.05, eta squared statistic=.25].

Insertion errors have two main causes: 1) accidental touches, for instance when users

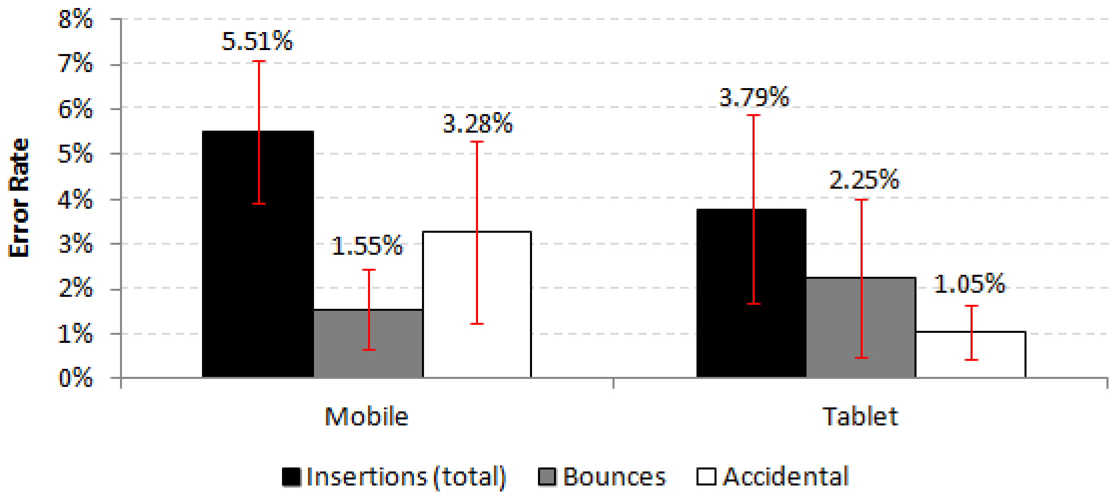


Figure 5.15: Overall Insertion, bounces, and accidental touches error rate with both device conditions.

are scanning the keyboard for the intended key and accidental touch another key; and 2) bounce errors, which occur when a key is unintentionally pressed more than once, producing unwanted characters. In this section we analyze in detail these two types of errors for each device condition. Knowing how to identify these errors whilst users type can be a valuable feature in order to prevent incorrect characters from being entered (Trewin, 2003)(Clawson et al., 2008). Our goal is to provide interface designers new empirical knowledge that allows them to build more effective virtual keyboards. 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.

Mobile device. Concerning bounce error rate (Figure 5.15), participants obtained a mean of 1.55% and high variance (sd=1.7%). We found large positive correlations with tremor measures: right hand *oscillation on the X axis* [Spearman rho=.596, n=14, p=.025] and left hand *peak magnitude acceleration between 7 and 12 Hz* range[Spearman rho=.532, n=9, p=.14]. Moreover, we found a negative large correlation between *bounce* error rate and *QWERTY experience* [Spearman rho=-.623, n=14, p=.017].These results suggest that bounce errors are strongly related to tremor, yet expert QWERTY users may be able to compensate for their difficulties. Regarding accidental touches, participants achieve a slightly higher error rate with an average of 3.28% (sd=3.9%). Conversely to bouncing errors, accidental touches did not correlate with QWERTY experience. However, we found again a large positive correlation with right hand *oscillation on the X axis* [Spearman rho=-.559, n=14, p=.038]. Moreover, left hand oscillation also showed a correlation with several tremor measures: *oscillation Y axis* [Spearman rho=.762, n=9, p=.017], *oscillation Z axis* [Spearman rho=.536, n=9, p=.162], *peak magnitude acceleration between 3 and 7 Hz* [Spearman rho=.508, n=9, p=.162]. Results show that accidental touch variance is

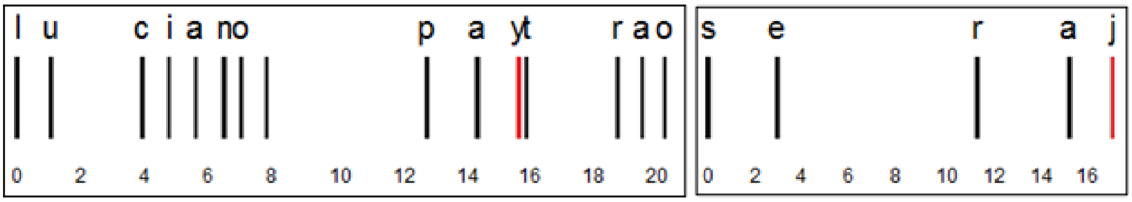


Figure 5.16: Individual typing pattern to illustrate accidental touches (in red). Touch events are plotted on a timeline, 0 s marking the start of a touch down event. Black bars denote touch duration on screen, white represents inter-key interval. Top-bar characters are the inserted letters. Time in seconds is represented in the x axis. Left - accidental insertion was near intended key, Right - accidental insertion far from intended key, suggesting that is was due to "finger scanning" behavior.

largely explained by tremor measures. A detailed inspection of participants' performance showed that these errors usually occurred near the intended key (Figure 5.16 - left); however, it can also happen in more distant keys, for instance when users are performing visual searches and simultaneously hovering their fingers over the keyboard (Figure 5.16 - right). In both cases, participants accidentally touched a key and entered an undesired character. Moreover, from both illustrations these touches are immediately before or after the intended key, and the time between keys is smaller than most inter-key presses. We believe that this type of error can be prevented by analyzing typing features, such as key press duration and inter-key intervals.

Overall, bounce errors and accidental touches account for a large number of insertion errors (Figure 5.15), thus dealing with them would significantly enhance input performance. Results suggest that by analyzing users' typing behaviors these errors could be prevented. On the other hand, the remaining insertion errors would be more difficult to automatically identify since they are intentional touches and occurred when participants tried to correctly input the last character again; that is, after failing the first attempt, some participants tried to correctly input the character in a second attempt. Since deletion was not available, this resulted in additional characters. We believe that in more naturalistic settings these errors will be reduced. Other alternative is using orthographic correction algorithms.

Tablet device. Results (Figure 5.15) showed an average of 2.25% Bouncing error rate (sd=3.5%). The relatively high standard deviation suggests that bounce errors are highly user dependent. However, we did not find any large correlation with either demographic or tremor measurements. Regarding accidental touches, participants obtained an average of 1.05% (sd=1.22%). As with bouncing error rate, we did not find any large correlations between accidental touches and tremor measures.

Although bouncing errors are hard to predict, they can be easily distinguished from correct (or at least intentional) key presses. Figure 5.17 illustrates the transcription of a word from participant #3. The required sentence included the word "receitado" (Portuguese for prescribed), however participant #3 transcribed "rreceiittadoo". Bounces are clearly highlighted due to its key press duration and between keys interval. We believe that a

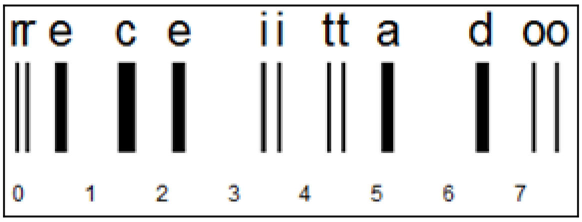


Figure 5.17: Individual typing pattern to illustrate bounce errors. Touch events are plotted on a timeline, 0 s marking the start of a touch down event. Black bars denote touch duration on screen, white represents inter-key interval. Top-bar characters are the inserted letters. Time in seconds is represented in the x axis. Bounce errors are characterized by short touch durations and small inter-key intervals.

filtering solution, similar to the one proposed by Trewin (Trewin, 2002), would be beneficial to some users. By adjusting some features of touch typing, these key presses could be ignored and therefore prevent errors.

Summary. Overall, bouncing errors and accidental touches account for the majority of insertion errors, therefore, preventing these gains higher relevance to significantly improve users' performance. Moreover, we believe that these types of errors could be automatically classified and dealt with by analyzing users' typing behaviors. However, further research is needed to confirm this hypothesis. Also, it would be interesting to observe if this solution would generalize to different user profiles and devices. From this user study, results suggest that individuals are consistent in their input behaviors; however, typing patterns may be both user- and device-dependent. For instance, participant #3 committed several bounce errors with the tablet device (14%) while with the mobile condition bounce errors were much lower (2%).

When comparing devices, we found a significant decrease of accidental touches in the tablet condition [Z=-2.292, p<0.05, eta squared statistic=.29]. Conversely, bouncing errors were not statistically different between device conditions [Z=-.314, p=0.754], although there is an increase of bounce errors with the tablet device (Figure 5.15). These results suggest that accidental touches are less common with larger and static targets (the tablet device was placed on a table).

5.3.7. Omission Errors

Omissions were the most common error type in this user study. In general, errors may occur for two main reasons: cognitive or motor (Kristensson, 2009). Cognitive errors are due to the user having an improper model of an intended word/sentence. For instance, a user may forget to insert a specific character or may not be able to spell an intended word correctly. Motor errors are due to mistakes that occur due to tremor or other behavioral causes. We believe that omission errors may be due to both reasons.

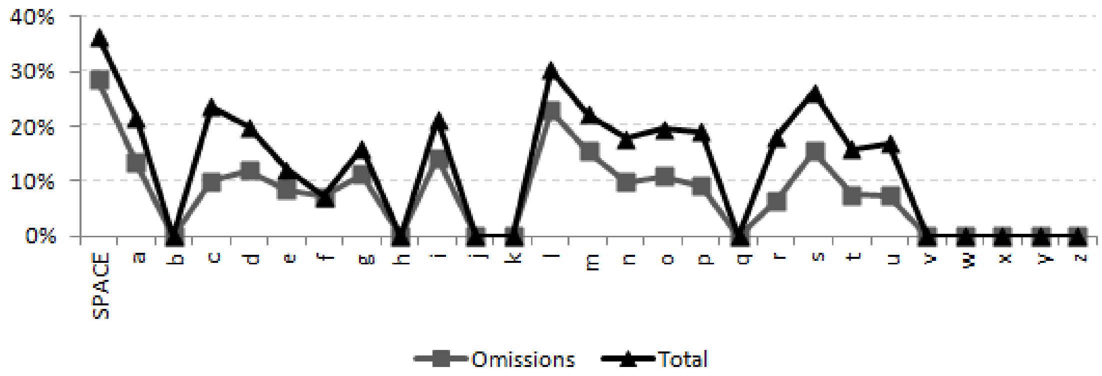


Figure 5.18: MSD and Omission error rate per key in mobile condition.

Understanding omission errors is particularly difficult since it is hard understand the reason why participants failed to enter the intended character/word. Was it because they forgot it or because the device was unable to recognize the users' touch? In order to answer this type of questions, we resorted to video recordings.

During our user study with older adults, 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, especially for those without QWERTY experience. From Figure 5.18, we can see that nearly 28% of errors associated with that key are omissions and that no other key achieves such a high error rate. We believe this effect may be due to the unfamilarity of participants with touchscreen technologies and text-entry tasks. Also, cognitive/learning skills may be playing an important role since participants were not able to retain the information presented in the practice session.

Other reason for high omission error rate was due to forgetting to transcribe some letters or words. For instance, participant #13 usually forgot to transcribe words at the middle and end of sentences, resulting in high error rates. Still, her performance was consistent across device conditions (see Figure 5.6 and Figure 5.7), which suggest that it was device independent. The lack of experience in copy tasks or an improper mental model of the required sentence are plausible reasons.

On the other hand, participants #8 and #10 experienced a significant decrease in omission error rates from mobile to tablet condition. From video recordings, we saw participants getting too frustrated and even swearing during trials, illustrating major difficulties within this device condition. A detailed inspection of touch points revealed that participants were in fact hitting the keyboard, however they were touching "empty spaces", near border keys (e.g 'l', see Figure 5.18). Thus, some omission errors were due to lack of keying accuracy.

Although participant #8 experienced a decrease in omission rate in the tablet condition, it was still relatively high (see Figure 5.7). Overall, the copy task seemed to be overwhelming for this participant as she could not manage and coordinate what she had transcribed and what was yet to be transcribed. She frequently asked things like: "where was I?", "have I written this?", resulting in skipped letters and words.

While omission errors may suggest participants' forgetfulness there were also some issues with touch interaction. Particularly, when holding or resting their non-dominant hand on the device, unintentional touches were common. These behaviors resulted in unrecognized key press, since our virtual keyboard only dealt with a single input point. When more than one finger was pressing the device, only the first one was recognized as valid touch. All the remaining were ignored, resulting in omitted characters.

Overall, there were three main reasons to the high level of omission errors: participants' forgetfulness (either due to lack of typing experience or inadequate mental models of required sentences), hitting "empty spaces", and accidental touches on the devices' borders.

5.3.8. Participants Preference, Comments, and Observations

At the end of this 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. 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 limit above the two-choice change expectation of 50%. The main reasons were due to key size and spacing, which gave them some margin of error. Moreover, some participants also referred that letters (i.e. visual feedback) were easier to see.

> *This one (the tablet) is easier than the phone because keys are larger.*
>
> *It is easier. It is easier to see (the letters). It is clearer.*
>
> *I like them both, but this one is easier because it has larger keys.*

When asked about their main difficulties, participant pointed some issues about: 1) key acquisition, particularly in the mobile condition, and 2) keyboard layout, mainly for those with little QWERTY experience.

> *I am always pressing the wrong keys. I am always hitting neighbor keys.*
>
> *My fingers are too large to use it (mobile device).*

> *The hardest thing is trying not to tremble while texting.*
>
> *The (mobile device) screen is too sensitive.*
>
> *The main difficult for me is in knowing where the letters are. I am not used to it.*
>
> *If letters were placed alphabetically it would be easier for me.*

5.4. Towards Inclusive Keyboards

The analysis presented above provide insight about older adults typing patterns and how keyboard features may be improved to better support text input on touch-based devices. We access the reliability of key press features to evaluate how accurate the modified designs could be. In this section, we perform a user-dependent and user-independent analysis. The goal is to demonstrate how much one can gain with simple key classification approaches; however, we believe that more efficient solutions can be found resorting to more sophisticated measures and machine learning algorithms.

5.4.1. Dealing with Insertion Errors

In order to deal with insertion errors, we calculated the optimal inter-key threshold that would allow reducing both insertions and overall MSD error rate. Threshold values higher than the optimal would most likely also reduce insertion errors, at the cost of omission errors and consequently increasing MSD error rate. In this analysis we used 40 values, which corresponded to the number of 25ms intervals from 0ms to 1000ms; that is, key presses that had an inter-key interval lower than the threshold being tested were considered insertions and, therefore, were discarded. We then computed both insertion and MSD error rate from all resulting sentences and compared it against the baseline condition (i.e. no filter).

For the user-dependent analysis, we calculated the optimal inter-key threshold for each participant, based on their insertion and MSD error rate. MSD error rate dropped, on average, 6.8% in the mobile condition, 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, which indicates that individual differences play an important role in this type of solutions.

For the user-independent analysis, we calculated the mean insertion 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 slightly higher, 150ms and resulted in a decrease of 1.1% of MSD error rate. It is noteworthy the decrease in performance from the user-depended classifier, especially in the mobile device, which corresponds to 8.5 times lower error rate. This result suggests that filtering solutions should take into account each user typing behaviors. Moreover, this simple solution removed nearly 30% and 50% of insertion errors in the mobile and tablet conditions, respectively, which is a good result, considering its simplicity.

5.4.2. Dealing with Substitution Errors

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. In this section, we present a user-dependent and user-independent classification for both devices.

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 based on the closest centroid. In the mobile condition, substitution and MSD error rate dropped, on average, 9.8% and 11.5%, respectively. In the tablet condition, substitution error rate was reduced by 1.1% and MSD error rate dropped 1.2%. Overall, participants were consistent within themselves, consistently hitting the same places for the same keys. Nevertheless, the mobile device condition seems to benefit more from adaptive solutions, which may be explained by the lower substitution rate of the tablet condition.

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 substitution and MSD error rates, on average, by 7.8% and 9.8% for the mobile condition, and 0.7% and 0.6% for the tablet condition, respectively. Again, the mobile gain is much higher than the tablet condition. However, the user-independent classification performed worst for all conditions, suggesting that personalization should be taken into account when designing touch-based solutions for older adults.

5.5. Discussion

Our goal was to investigate how older users inputted text in touch-based devices in order to improve their performance. In this section, we discuss the obtained results by: 1) answering the previously proposed research questions; 2) describing some lessons learned that could be useful for future works; and 3) identifying implications for design.

5.5.1. Answering Research Questions

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

1. *How do older adults perform on touchscreens regarding speed and accuracy?*

   Participants achieved a maximum of 11.5 wpm using a tablet device (mean of 4.7 and 5 wpm for mobile and tablet conditions, respectively). Also, input speed 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 for older adults? What are their main causes?*

   The most common error type among older adults was omissions (9-12.6%). This pattern occurred across device conditions, suggesting that most omission errors are device-independent. Cognitive abilities may be related with these errors as some participants showed difficulties performing copy tasks and coordinating the required-transcribed sentences. This resulted in users forgetting to enter blank spaces or entire words. 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 (i.e. users landed their finger on the wrong key).

3. *Do tablet devices allow lower error rates?*

   Overall, we found a decrease of 9% in MSD error rate from mobile to tablet devices. This finding suggests that tablet devices ease 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 between mobile and tablet devices on omission errors, suggesting that these errors are device-independent.

4. *Is users' text input performance correlated with hand tremor?*

   Although input speed was mainly related with QWERTY experience, errors were strongly correlated with participants' tremor. 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 the oscillation in the X axis (dominant hand). The non-dominant hand also played an important role in mobile errors: hand oscillation was correlated with both overall *MSD* error rate and accidental touches, while peak acceleration magnitude explained a wide range of slip errors. These findings suggest that future mobile interfaces should take into account the users' tremor profile in order to provide more suitable text-entry designs. Still, designers should consider different features of tremor.

5.5.2. Lessons Learned

In this section we describe a set of lessons learned derived from previous presented results and personal observations/experience.

Tapping is easily understood. Although participants had never used a touch-based device, the tapping technique was easily understood by all participants. They stated that it was very easy and enjoyable to use, even though they spent some time looking for the intended key.

Forgetfulness is common for older adults. This resulted in many omission errors. In fact, this was the most common error type, even though it is highly user-dependent. Participants often forgot to type individual characters or even entire words. The blank space was the most problematic character. We hypothesize that this effect may be due to the lack of experience in typing/copy tasks of our participants or due to an inappropriate mental model of the required sentences (cognitive errors).

Substitution of similar characters. A small amount of substitution errors were not related with motor abilities. Participants frequently replaced similar or symmetrical letters, such as i $\rightarrow$ l or p $\rightarrow$ q. We believe these behaviors to be related with either visual/perceptual or cognitive (mental confusion) errors.

Copy task was sometimes overwhelming. As seen before, omission errors were fairly common among this user population. Text-entry studies with younger populations did not report such magnitude of omission errors, even when situationally impaired (Nicolau and Jorge, 2012c), which suggests that this may be related with age or some collateral factor (e.g. familiarity). In fact, during the user study we have witnessed some users that were not able to coordinate the required-transcribed sentence and frequently asked for help. We decided to use copy tasks in order to minimize cognitive and memory demands, yet some participants felt overwhelmed and were not able to coordinate require-transcribed sentences, resulting in forgetfulness of entire words.

Accidental touches while scanning. Users that were unfamiliar with QWERTY keyboards spent more time searching for intended letters. As they did, they normally used their index finger to scan over the virtual keyboard. This behavior resulted in accidental touches and consequent insertion errors.

Lift-off typing strategy. To avoid multiple insertions or incorrect characters, mobile keyboards should implement lift-off typing strategy for all keys. Particularly, many commercial available keyboards allow multiple blank spaces to be inserted with a single land-on target action. We believe this to be an unsuitable solution for older adults, since key presses are usually long and would result in several blank spaces. Keyboards that use long key presses to provide alternative characters should also be aware of average key press durations in order to avoid unwanted characters.

5.5.3. Design Implications

In this section we describe implications for design that resulted from the user study.

Shift keyboard layout. 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 with hand dominance, thus further investigation should explore this hypothesis. Also, it is not clear if this change should be visible to the user. Indeed, previous work has shown that adaptation should not be visible to users, as they may be distractive and may influence users' behaviors (Findlater and Wobbrock, 2012). Nevertheless, future work should also explore how user behavior changes with different visual affordances.

Width rather than height. Whenever it is possible keys should be wider instead of taller. For both devices tested we found higher x-axis touch dispersion, suggesting that users are more favorable of wider keys. In fact, most 12-key (old) physical keypads respect this layout; however, this was lost in recent touch interfaces.

Narrower spacebar. The spacebar width is three times larger than the remaining keys. Thus, conversely to other keys, touch deviations suggest that spacebar should be narrower. Other authors also obtained similar results for tabletop touch typing (Findlater et al., 2011). Reducing the spacebar's size has the potential to diminish substitution errors, particularly with characters C and B. Thus, 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 modeling users' typing features 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. Indeed, we believe that many insertion and substitution errors can be automatically classified and handled. Participants seem to be consistent within each device regarding their typing behaviors; however, individual differences may play an important role.

Allow personalization. From our results we conclude that personalization will play a significant part in touch-based solutions for older adults. We observed several individual differences regarding typing behaviors, particularly hit point locations, key press duration, and inter-key interval. Future research should tackle these issues by providing user-dependent solutions and assessing their effect.

Deal with poor aiming rather than finger slips. Keyboard designers should focus on dealing 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. Non-motor errors were quite common among older adults. Simple language-based solutions can provide a suitable solution to these issues. For example, to deal with blank space omissions and substitution of very similar letters (e.g. p $\rightarrow$ q, m $\rightarrow$ w). However, one should take into account that traditional space-based correction approaches (Kane et al., 2008a) may be ineffective, since our participants often forgot to enter this character.

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

5.5.4. Limitations

The user study reported here does not contemplate error correction. While this was necessary to assess natural typing patterns, understanding users' correcting strategies is also needed. Further research should focus on reporting error correction effects on touch-based devices.

Our participants only included novice users in touch technologies, which are a common user group among older adults and are in need of more effective solutions. Although some of the reported results may not be generalized to expert users, the resulting design implications may still improve their performance. For instance, derived keyboard layouts may be reflective of users' motor abilities, which will not improve with age. Nevertheless, future work will need to see how typing behaviors change with experience.

Finally, in this user study we only controlled for hand tremor through a series of questions and tasks. Controlling for cognitive abilities may also provide some insights about error patterns. For instance, measuring both short-term memory and attention abilities, through various verbal tests (Oliveira et al., 2011a), may be useful to explain cognitive errors. If so, detecting these abilities in a configuration step would allow correction algorithms to adapt to each user.

5.6. Conclusion

We investigated text-entry performance of 15 older adults on touch-based devices. Our user study featured two device conditions (mobile and tablet) and assessed each user tremor profile. Results showed that error rates are still relatively high (20% on average) compared to younger users performance (Nicolau and Jorge, 2012c). Moreover, hand tremor was strongly correlated with input errors, indicating that this information can be used to enhance text-entry accuracy. Regarding types of errors, the most common were omissions (10.8%), followed by substitutions (5.8%), and insertions (4.6%). Nevertheless, we saw that tablet devices, due to their key size and static position, can improve accuracy by about 9%.

Next, we assessed the effect of new (simple) classification models to deal with insertion and substitution errors. These simple solutions were shown to improve substitution errors by 10-12%; while insertion rates dropped as much as 6.8%. Moreover, results show that personalization should be considered when designing touch-based input solutions for this target population, since user-dependent solutions allowed for higher accuracy rates. Lastly, we identify some design implications that should improve typing accuracy and encourage researchers to create more effective solutions for older adults.

Future research should apply the design implications described here and investigate their effect on text-entry performance. Moreover, we intend to improve input accuracy by analyzing typing features and tremor profiles. Recurring to probabilistic models and machine learning algorithms, it may be possible to improve key classification accuracy and reduce error rates. Time-based features will be analyzed as well in order to filter accidental touches and bounce errors. Using accelerometer data during text-entry tasks will be our main focus in order to deal with input errors in real-time (see Chapter 7).

6
The Disability Continuum

Over the years, solutions targeted at disabled people have demonstrably benefited all users. A good example is the T9 text-entry method (Pavlovych and Stuerzlinger, 2004). Still, accessibility largely remains a research area for minorities. The very distinction between accessibility and usability attests to this. The work presented in this chapter suggests that there are situations where the distinction between able-bodied and disabled users is not so clear. We thus argue that there is not a clear-cut line separating disabled from able-bodied people, which we regard as different points in a continuum of abilities.

After characterizing each user group in typing tasks (see Chapter 4 and 5), we are in condition to perform a comparative analysis. We aim to bridge the gap between SIID and HIID by answering two main research questions: First, what are the main similarities between situationally-impaired and older adults' regarding typing performance? Identifying common issues will allow us to provide the knowledge for informed design. Second, we seek to identify the main differences between user groups, so that future solutions can address them and provide inclusive designs.

One major challenge lies in fairly comparing abilities from two distinct user groups. How can we model and compare them? Although we acknowledge that older adults and situationally-impaired users are in many ways different, we use a cause-agnostic approach. We compare users at a functional level by modeling their abilities through observed input accuracy (Wobbrock et al., 2011). Moreover, we analyze participants' tremor profile and their relationship with input performance. In what follows, we discuss our experiment to measure text-entry performance, describing the experimental setup and drawing conclusions from observed results. Also, guidelines for inclusive design are presented as well as avenues for future work.

6.1. Revisiting Previous User Studies

Historically, accessibility and usability have been regarded as distinct research domains. However, recent advances in mobile technologies have the potential to unify these communities into Mobile Accessibility. However, it is not yet clear whether and how they overlap. Thus, we seek to describe the main differences and similarities between user groups and allow designers to build effective solutions for a wide range of abilities. To this end, we studied text-entry performance with both situational and health impaired users (Chapters 4 and 5). This section provides an overview of participants' profiles, apparati and experimental procedures.

6.1.1. Participants

Table 6.1 illustrates participants' profile. Twenty two participants, 3 females and 19 males, took part in the first user study. Their ages ranged from 23 to 40 with a mean of 26.5 years. They were recruited from our university. None of the participants had visual or motor impairments and all of the participants owned a mobile phone whereas only 15 of them used touch screen technology regularly. All participants were right-handed. This user group will be called the SIID group.

For the second study we recruited fifteen participants (11 females and 4 males) from a local social institution. Their ages ranged from 67 to 89 with a mean of 79 years. All participants were right-handed. None of the participants had severe visual impairments; with corrected vision all participants were able to read the screen content. The only exception was participant #9. This participant was unable to correctly perceive the mobile device's characters, and therefore did not perform the user study in this condition. We present her results for the tablet condition; however, she was removed from the within-subjects analysis. 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 never entered text. Regarding QWERTY familiarity, six participants had used this type of keyboard in typing machines or personal computers. Even though this group is composed by healthy aging older adults without extreme aging issues, they do have age-related impairments, particularly increased physiological hand tremor (health-induced impairment). Thus, the user group will be called the HIID group.

User group	Mean age	Gender	Mobile experience?	QWERTY experience?	Touchscreen experience?	Visual characteristics
SIID	26.5	3 females 19 males	22 yes 0 no	22 yes 0 no	15 yes 7 no	5 corrected vision 22 able to read mobile characters
HIID	79	11 females 4 males	12 yes 3 no	6 yes 9 no	1 yes 14 no	11 corrected vision 1 unable to read mobile characters

Table 6.1: Comparison between SIID and HIID participant groups. HIID group presents lower mobile, typing, and touchscreen experience.

6.1.2. Apparati

An HTC Desire and ASUS EEE Pad Transformer TF101 with a capacitive touchscreen were used in the user studies. A QWERTY virtual keyboard, similar to Android's SDK keyboard, was used in both devices; for the HTC Desire each key was 10x10mm on landscape mode and 7x10mm on portrait mode, while for the tablet each key was 20x10mm (landscape). A letter was entered when the user lifted his finger from the key. Neither word prediction nor correction was used. Acceleration data was captured through the mobile device's accelerometer for posterior analysis.

6.1.3. Procedures

SIID study. At the beginning of the experiment participants were told that the overall purpose of the study was to investigate how text-entry performance was affected by walking conditions. They were then informed about the experiment and how to use our evaluation application. We evaluated the participants' performance in three mobility settings: sitting, slow walking, and walking at average human pace (2 steps per second). In this chapter, we will only use results from seated and normal walking conditions. The experiment was conducted in an indoor test track at the university campus (without obstacles). In both walking conditions, we asked participants to follow a pacesetter while entering text. The experimenter instructed participants to stay within 2 meters of the pacesetter as he walked. If the participant fell behind by more than 4 meters, the experimenter logged a walking deviation for that trial. The pacesetter carried a mobile phone, which gave him feedback through vibration about the intended pace. Before each mobility condition participants had a 5 minute practice trial to get used to the pace and text-entry task. For each mobility setting, they were asked to enter text with 3 hand conditions (chosen randomly) using their thumbs: one-hand/ portrait, two-hand/ portrait, and two-hand/ landscape. For each condition participants copied seven different sentences (first two sentences were practice trials), resulting in 42 different sentences per participant. Both sentences and mobility conditions were chosen randomly to avoid bias associated with experience.

HIID study. At the beginning of the study, participants were told that the overall purpose was to investigate how text-entry performance is affected by the type of device. We then explained and exemplified to them how to use a virtual keyboard. Before the evaluation phase, we assessed the participants' capabilities regarding tremor (postural and action tremor) applying three different methods (Archimedes spiral test, capturing accelerometer data, UPDRS questionnaire). In this chapter, we will only use the Archimedes and accelerometer data, since the UPDRS results were not correlated with typing performance (see previous chapter). Participants 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 5 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; for the tablet device condition, it was placed on a table in front of them. For each evaluation condition, participants copied five different sentences (first sentence was a practice trial). The order of conditions was counter balanced to avoid bias associated with experience. Each subject entered a total of 10 different sentences.

In both studies, sentences were displayed one at a time, at the top of the screen. Copy typing was used to reduce the opportunity for spelling errors and to make error identification easier. Participants were instructed to type as quickly and accurately as possible. We wanted to elicit natural typing behaviors and did not want participants to be concerned with the accuracy of their input. Thus, we followed a similar approach to Gunawardana et al. (Gunawardana et al., 2010) and Goel et al. (Goel et al., 2012), and created a keyboard in such a way that error correction was not available. On the other hand, if the prototype had a delete key, it might introduce correcting strategies, which might vary across participants and upset the naturalness of the data.

All sentences, in both user studies, were extracted from the same written language corpus, each with 5 words, an average size of 4.48 characters per word, and a minimum correlation with the language of 0.97 (see Apendix A1).

6.1.4. Dependent Measures and Analysis

We measured performance during text-entry tasks by several quantitative variables: words per minute (WPM), minimum string distance (MSD) error rate, and character-level errors (substitutions, insertions, and omissions). We also gathered accelerometer data during each trial in order to characterize participants' level of impairment and applied Shapiro-Wilkinson tests to observed values in WPM, MSD error rate, and types of errors. We applied parametric statistical tests, such as repeated measures ANOVA and t-test, for normally-distributed dependent variables or non-parametric tests (Friedman

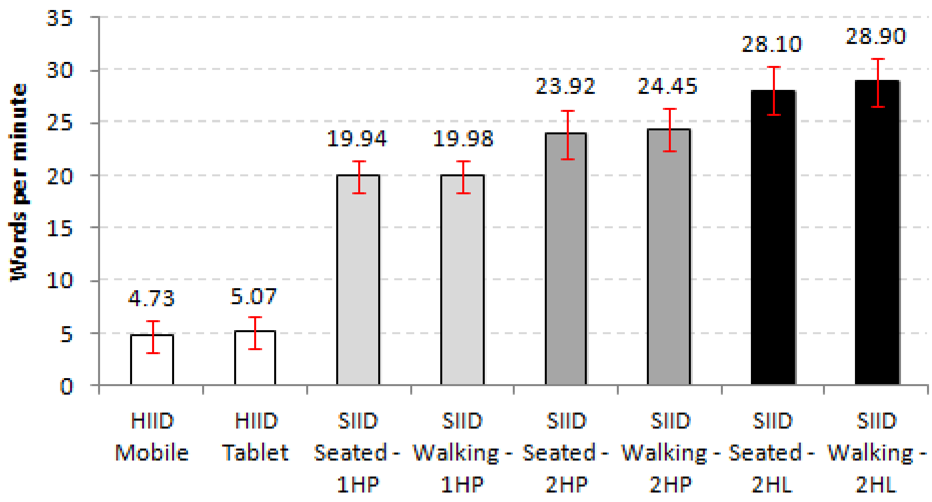


Figure 6.1: Words per minute for older adults and able-bodied participants. Error bars denote a 95% confidence interval.

and Wilcoxon) otherwise.

6.2. Investigating Differences and Similarities between HIID and SIID

Our goal in this chapter is to understand and bridge the existing gap between health- and situational-induced disabilities. Detailed results of each user study are reported in previous chapters. In this section, we focus our analysis in the comparison of both user groups, highlighting the main differences and similarities in text-entry performance. Additionally, we describe each user's tremor characteristics during typing tasks and relate them with input performance. This knowledge will enable designers to take into account users' abilities and provide effective solutions for a broader target population.

6.2.1. Input Speed

To assess speed, we used the words per minute (WPM) (MacKenzie and Soukoreff, 2002b) text input measure calculated as:

$$(transcribed\ text\ length - 1) \times (60\ seconds \div time\ in\ seconds) \div 5\ characters\ per\ word$$

Figure 6.1 shows the participants' average WPM for Mobile and Tablet conditions.

Words per minute. As expected, able-bodied users, even when situationally impaired, were significantly faster than older users [older adults/mobile - situationally impaired/one-hand portrait: t(34)=-12.932, p<.001]. Situationally impaired were (more than) 4 times faster than older adults. This result is easily explained as all younger participants were familiar with QWERTY keyboards and, therefore, did not need to scan the interface to enter the intended letters. On the other hand, all older participants were new to touch interfaces and most of them never entered text in electronic devices. Nevertheless, input speed is not our main concern when comparing both populations. We believe user groups will be more similar regarding accuracy and types of errors.

6.2.2. Input Accuracy

The quality of the transcribed sentences was measured using the Minimum String Distance (MSD) Error Rate (MacKenzie and Soukoreff, 2002b), calculated as:

$$MSD(required\,text, transcribed\,text) \div mean\ size\ of\ the\ alignments \times 100$$

We also present a fine grain analysis of errors and categorized by type: insertions - added characters; substitutions - incorrect characters; and omissions - omitted characters (MacKenzie and Soukoreff, 2002a).

MSD error rate. Figure 6.2 illustrates participants' MSD error rate across all conditions. The mobility effect can be clearly seen in the SIID group, from seated (m=5.1% sd=4%) to walking (m=7.5% sd=7%) conditions, resulting on a significant main effect [$F_{1.244,26.115}$=4.962, p<.05]. This result confirms that indeed users were situationally impaired by mobility. Nevertheless, SIID's MSD error rates were still lower than HIID's group. In fact, when comparing SIID upper bound (walking two-hand portrait: m=16.5%, sd=11.9%) and HIID lower bound conditions (tablet: m=8.7%, sd=8.3%), we found a statistically significant difference [Z=-2.598, p<.01], suggesting that older participants face additional difficulties compared to situationally impaired users.

Types of errors. In addition to overall MSD error rate, we performed a thorough analysis of errors in order to better understand the performance gap between HIID and SIID groups. In this subsection we present a fine grain analysis of errors, categorized by type: insertions, substitutions, and omissions. Also, we report the absolute and relative magnitude of each type of error for all conditions (Table 6.2). As the name suggests, the absolute magnitude consists in the mean error rate; while the relative magnitude corresponds to the mean error rate expressed as a comparison to the overall magnitude of MSD error rate (error rate / MSD error rate *100), which illustrates the importance of an error type for that user group.

Regarding insertion errors, the relative magnitude is very similar across user groups, sug-

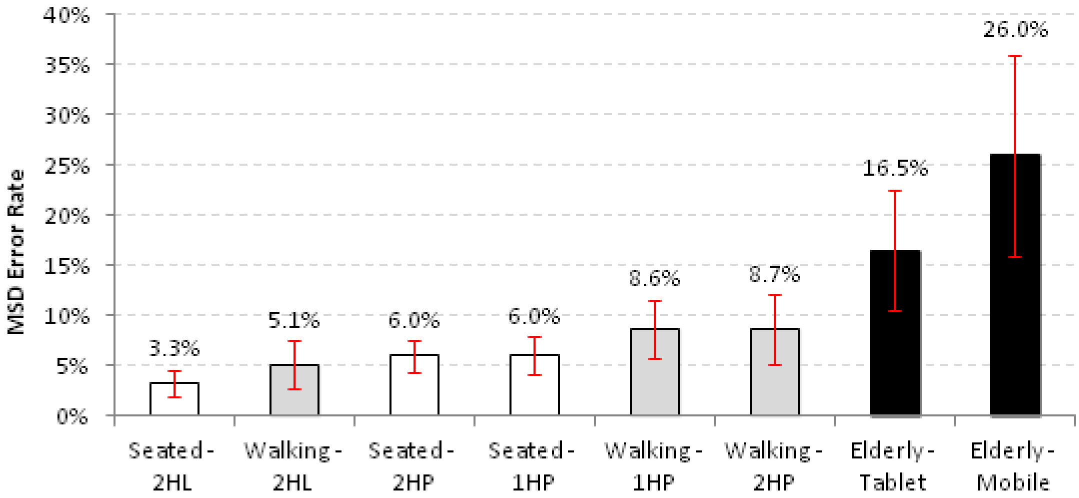


Figure 6.2: MSD Error rate across all conditions. Error bars denote a 95% confidence interval.

Condition	MSD	Insertions (absolute)	Insertions (relative)	Substitutions (absolute)	Substitutions (relative)	Omissions (absolute)	Omissions (relative)
HIID Mobile	26.0%	5.5%	21.2%	7.8%	30.1%	12.6%	48.7%
HIID Tablet	16.5%	3.8%	22.9%	3.7%	22.6%	9.0%	54.5%
Seated 1HP	6.0%	1.1%	18.6%	4.3%	71.4%	0.6%	9.8%
Seated 2HP	6.0%	1.1%	18.3%	3.8%	63.4%	1.1%	18.3%
Seated 2HL	3.3%	0.7%	20.7%	1.7%	50.6%	1.0%	28.4%
Walking 1HP	8.6%	1.0%	11.4%	7.0%	81.3%	0.6%	7.3%
Walking 2HP	8.7%	1.4%	16.5%	5.5%	62.9	1.8%	20.4%
Walking 2HL	5.1%	1.0%	18.8%	3.0%	59.1%	1.2%	22.3%

Table 6.2: MSD, insertion, substitution, and omission error rate for each condition.

gesting that insertion errors have a similar relative importance for both HIID and SIID participants. On the other hand, some types of errors are more relevant to one user group than another. While omissions are the most common error type of HIID group, substitutions account for the majority of errors of situationally impaired users. This result illustrates the differences in the relative importance of each error type.

Regarding absolute error rates, both insertion and omission errors follow the same pattern of MSD error rate; that is, HIID participants obtained higher error rates than SIID users. When comparing SIID upper bound (worst result) with HIID lower bound (best result), we still found significant differences on insertion [Z=-2.511, $p<.05$] and omission error rates [Z=-3.093, $p<.005$]; that is, health-induced disabilities introduce additional omission and insertion errors when compared to walking conditions.

Nevertheless, conversely to omission and insertion (absolute) error rates, differences between HIID and SIID groups concerning substitutions are more blurry (Figure 6.3). In fact, older users' performance with the tablet device is very similar to SIID seated conditions (no significant differences were found). This suggests able-bodied users are disabled by mobile devices. Moreover, whilst walking, the SIID group's performance was similar

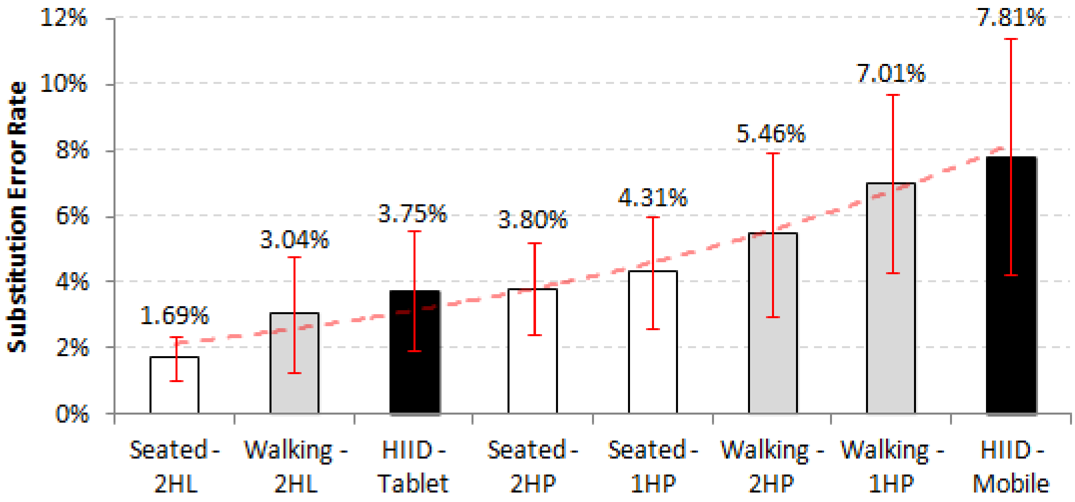


Figure 6.3: Substitution error rates across all conditions. Error bars denote 95% CI.

to older adults performance with a mobile device, showing that the increased demand of mobility results in a similar level of disability between user groups. We called this effect the “disability continuum”, since there is not a clear distinction between health- and situational-induced disabilities. Instead there is a continuum of disabilities that is revealed with the increased demand of each condition. Overall, HIID and SIID groups revealed similar substitution errors; however, older participants committed additional insertion and omission errors. Also, exception has to be made to two-hand landscape mode, which is able to compensate situationally impaired users' difficulties and provide a more accurate input [Z=-2.468, p<.05].

Summary. According to Table 6.2, particularly the relative magnitude columns, some types of errors are more relevant to one user group than another. While omissions are the most common error type on older adults performance, substitutions account for the majority of errors on situationally impaired. Nevertheless, dealing with substitutions will prove to be of great value to both populations, since its absolute magnitude is similar between health-induced and situational-induced disabilities.

6.2.3. Substitutions in Detail

In this section, we perform a more detailed analysis on substitution errors, since this was the type of error with most similarities between user groups. We identify the most problematic characters, substitution patterns and cause of errors. This analysis will provide valuable insights in order to draw design recommendations and provide knowledge to improve future designs of virtual keyboards for both target populations.

Most problematic characters. Overall, both user groups had similar difficulties across all keys and none stand out as the most problematic. Moreover, no row, column, or side

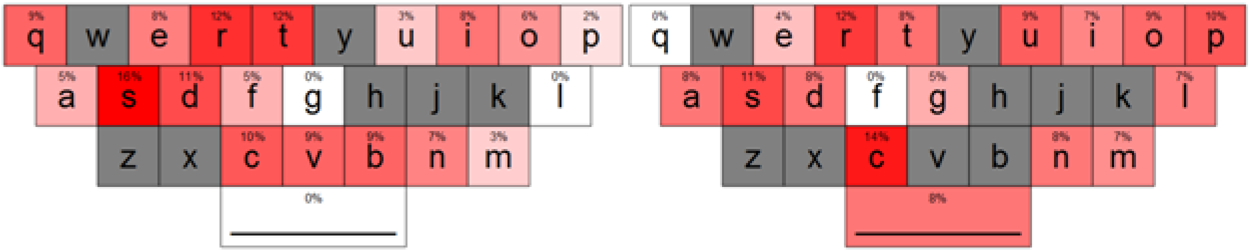


Figure 6.4: Substitution error rates for situational (left - one hand portrait condition) and health (right) disabled participants. Key's "redness" illustrates higher error rates. Grey cells represent keys that were not used.

emerged from data. Figure 6.4 illustrate substitutions error rates per key for situational and health disabled participants.

Substitution patterns. For situational impaired users, most frequent errors were at the right/left of intended keys. This finding may be related with key height in portrait mode; that is, keys are slightly higher than wider (7x10mm). Nevertheless, in landscape mode this pattern remained unchanged. Also, the non-dominant hand seems to be less accurate. For older participants we also found a consistent substitution pattern: right-bottom errors. Hit points were slightly deviated in right-bottom direction, which may be related to hand dominance and one-hand interaction mode. Moreover, most substitutions were at the distance of one-key. Overall, both user groups presented a consistent substitution pattern. Adjacent keys were commonly touched instead of intended keys. While same-row errors are common whilst walking, right and bottom substitutions are more frequent for older participants.

Cause of errors. Substitution errors have two main causes: poor aiming or finger slips. A finger slip consists in a correct landing on target and incorrect lift (i.e. users lift the finger on a nearby key), which originates a substitution. On the other hand, poor aiming errors occur when users land and lift their finger on wrong keys. Average slip error rate was below 10% and 14% for mobile users and older adults, respectively. Overall, most substitutions were due to poor aiming for both user groups, illustrating the importance of compensating hit points on touch typing.

One of the main differences between user groups regards substitution of similar letters. Some HIID participants consistently performed substitutions such as: p → q, m → w, i → j. We believe these errors to be either perceptual or cognitive; that is, users' visual abilities affected their perception of the required letter or they had an improper mental model of the character and confused it with a similar one. This finding suggests that substitutions are not due to motor errors alone. Still, adjacent substitutions were the most frequent.

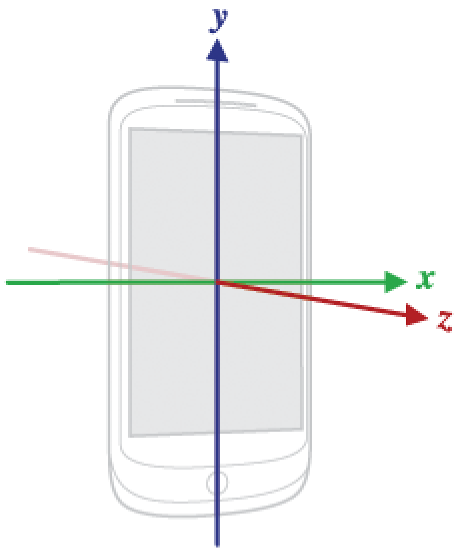


Figure 6.5: Mobile device coordinate system in portrait mode. We multiplied values on the Z axis by -1 so coordinates behind the screen have positive values.

6.2.4. Tremor Analysis

In this section we analyze the accelerometer data gathered during text-entry tasks. We will use data from older (mobile condition) and younger (both seated and walking conditions for one-hand posture) participants performance. Our goal is two-fold; firstly, we intend to characterize both user population tremor pattern whilst typing and identify their main differences and similarities regarding motor demands. Secondly, we will draw relationships between tremor features and text-entry performance and show the potential of each feature to be used to compensate typing errors.

Processing Accelerometer Data

The coordinate system of accelerometer readings is defined relative to the mobile device's screen in its portrait orientation (Figure 6.5). The X axis is horizontal and points to the right, the Y axis is vertical and points up and the Z axis points to the outside of the front face of the screen. Since older participants used the device in the landscape mode, X and Y axes are rotated 90º, i.e. the X axis is vertical and points up and the Y axis is horizontal and points to the left. In order to compare tremor profiles between both user groups, we translated gathered data to a common coordinate system defined as horizontal (positive - right; negative - left), vertical (positive - up; negative - down), and normal (positive - front; negative - back).

Finally, data from the mobile device's accelerometer was passed through a low-pass filter to remove noise (i.e. jitter). We used the 1€(one Euro) filter (Casiez et al., 2012), which is a first order low-pass filter with an adaptive cutoff frequency. We used 12Hz as the minimum cutoff frequency in order to capture high frequency physiological tremor (Elble and Koller, 1990).

Characterizing Participants' Tremor

In this sub-section, we will analyze users' hand motion during text-entry tasks and describe their tremor profile according to a set of features. This analysis will be performed for both user groups. We intend to answer questions such as: Do tremor characteristics follow a consistent pattern? Is it similar across populations? If not, what are their main differences and similarities? Answering these questions will allow us provide new insights on how situational- and health-induced tremor can be predicted and dealt with.

Dominant frequency. Analysis of this feature can provide new insights of how fast participants' hands shake. Additionally, it can also show the consistency of the tremor pattern within each user group. To calculate the dominant frequency, we took the Fast Fourier Transform (FFT) of the accelerometer signal and found the frequency with the maximum amplitude. The dominant frequency of walking conditions was 0.7 Hz (sd=.3), 1.67 Hz (sd=.4), and 1.68 Hz (sd=.4) for horizontal, vertical and normal axes, respectively. On the other hand, older adults obtained a dominant frequency of 0.04 Hz (sd=.07), 0.04 Hz (sd=.04), and 0.81 Hz (sd=2) for horizontal, vertical, and normal axis, respectively. From Figure 6.6, we can clearly see that situationally-impaired participants had higher and more consistent (less variance) dominant frequencies. The high standard deviations of seated and older users' suggest an irregular tremor pattern across participants. This is particularly visible on the normal axis: while SIID participants obtained a consistent pattern near the 2 Hz frequency, which corresponds to the imposed walking pattern (2 steps per second), the standard deviations of the remaining conditions is higher than the mean dominant frequency. Figure 6.7 (top left and bottom left) shows an example of the mean acceleration from SIID (whilst walking) and HIID participant. Note that the walking pattern is clearly visible, and previously reported by Crossan et al. (Crossan et al., 2005), while the older adults' motion is more irregular and occurs at lower frequencies.

Indeed, when comparing both participants' dominant frequencies (Figure 6.7, right images), the SIID participant has a clear spike near 1.77 Hz, resulted from a rhythmic motion, while the older participant as a set of frequencies with high amplitudes (1.64, 1.87, 2.15, and 6.7 Hz). This result illustrates the irregular nature of older adults tremor.

Amplitude of dominant frequency. The amplitude of the dominant frequency is generally used to characterize the amplitude of users' tremor (Salarian et al., 2007). Figure 6.8 shows the results obtained (in m/s^2) for both SIID and HIID participants across the three axes. Interestingly, mean acceleration values for older participants are between seated and walking results. For instance, in the normal axis results showed a mean amplitude of 8.93 m/s^2 (sd=5.8), 13.61 m/s^2 (sd=13.2), and 22.22 m/s^2 (sd=12) for seated, older adults and situational impaired participants. This suggests that older users indeed experience

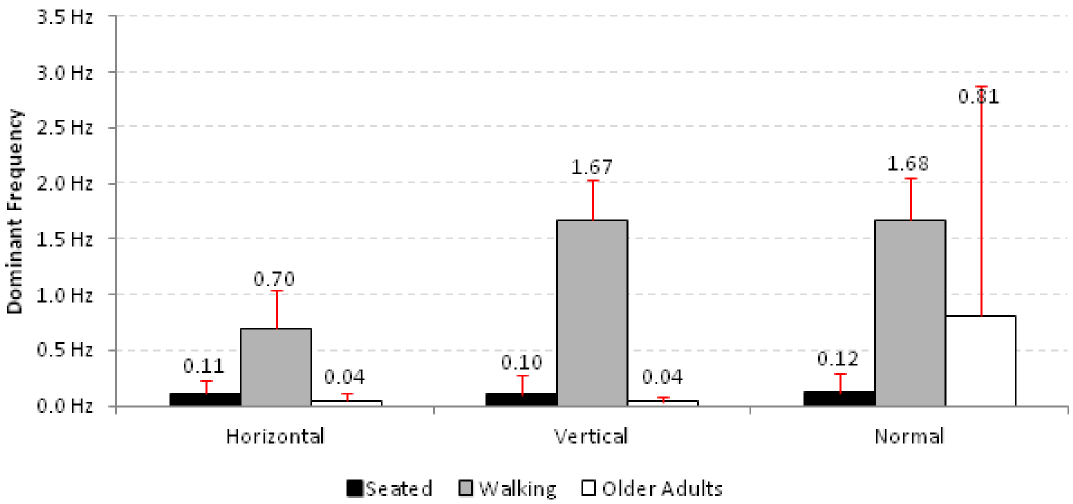


Figure 6.6: Dominant Frequency (Hz) across axes for younger (seated and walking) and older participants. Error bars denote standard deviations.

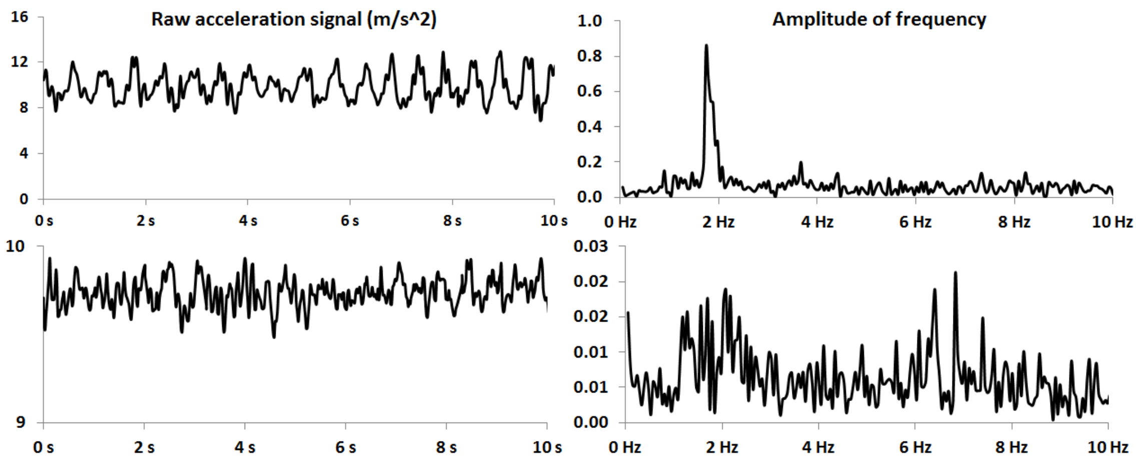


Figure 6.7: Top - SIID example: acceleration signal (left), and dominant frequencies (right); Bottom - Older adult example: acceleration signal (left), and dominant frequencies (right). As opposed to the older users (bottom-right), there is a clear dominant frequency for the situational impaired users (top-right).

increased physiological tremor, since the amplitude of the dominant frequency is higher than the one obtained by younger participants while seated; however, walking conditions seem to induce even higher tremor amplitudes.

Hand oscillation. We calculated hand oscillation as the standard deviations of acceleration (Bergstrom-Lehtovirta et al., 2011). Thus, if acceleration was fairly constant during trials, hand oscillation was near zero. On the other hand, if there were sudden accelerations or decelerations, hand oscillation would increase and reflect this behavior. Figure 6.9 shows hand oscillation across conditions for horizontal, vertical and normal axes in m/s^2.

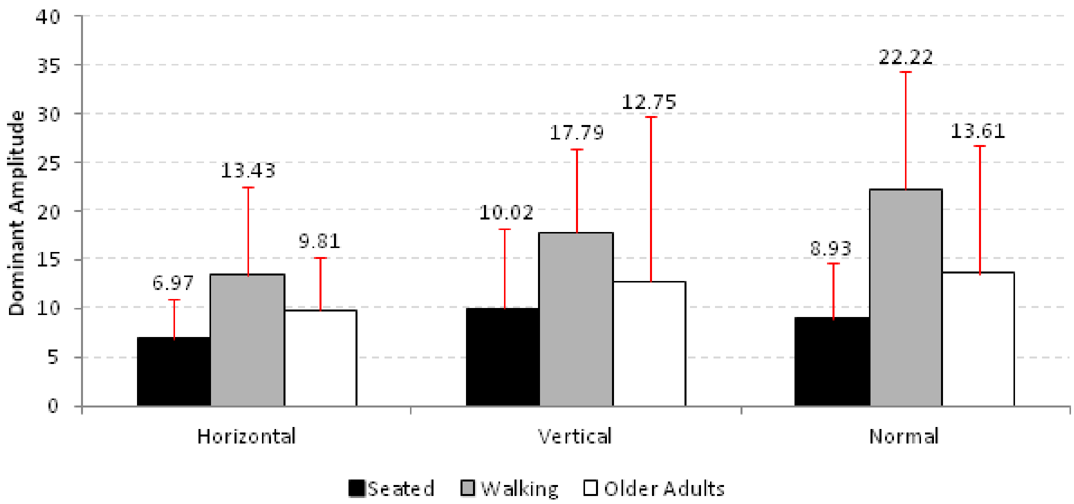


Figure 6.8: Amplitude of dominant frequency (m/s^2) across axes for younger (seated and walking) and older participants. Error bars denote standard deviations.

Overall, hand oscillation is higher for situationally-impaired users, reflecting the demand imposed by walking settings and confirming that, indeed, tremor amplitude is higher in these conditions. Nevertheless, although tremor amplitude and hand oscillations of older adults are lower than situational impaired users, they occur at lower frequencies (see Figure 6.6) and seem to be more irregular, which can be harder for users to compensate.

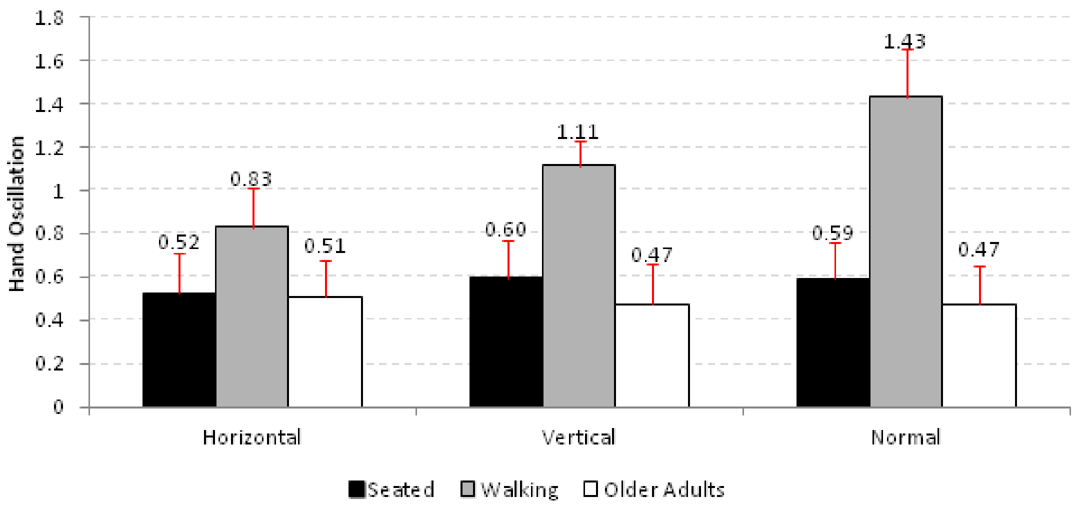


Figure 6.9: Hand oscillation (m/s^2) across axes for younger (seated and walking) and older participants. Error bars denote standard deviations.

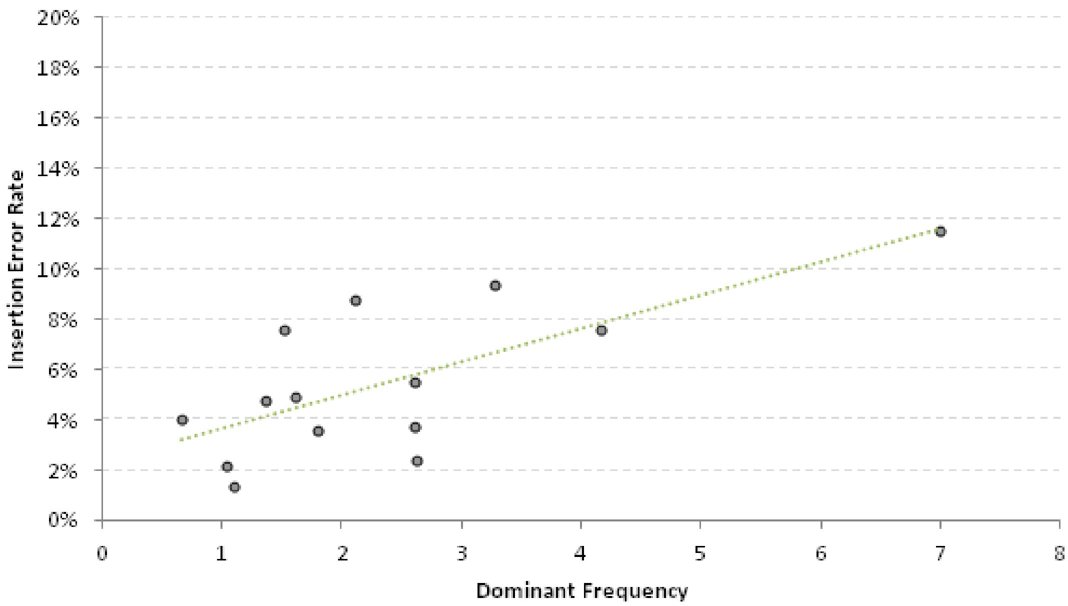


Figure 6.10: Scatter plot of amplitude (m/s^2) of XYZ dominant frequency vs. insertion error rate. There is a significant and strong positive relationship between the variables [Spearman rho=.574, n=14, p<.05].

Relating Accelerometer Data and Text-Entry Accuracy

The goal of this analysis is to provide insights on what tremor features may improve input performance and compensate for errors, particularly insertion and substitution errors. Results will allow designers to further explore these features and use tremor data to enhance users' performance. Additionally, we compare the results from SIID and HIID participants and derive some design implications.

Dominant Frequency. We did not find any strong or medium relationships between dominant frequency and text-entry performance for older users. However, there was a medium positive relationship between substitution errors and dominant frequency on the vertical (up/down) [Spearman rho=.324, n=44, p<.05] and normal (front/back) [Spearman rho=.345, n=44, p<.05] axis for situational impaired participants, illustrating the negative effect of walking conditions.

Amplitude of dominant frequency. There was a medium positive relationship between the amplitude of dominant frequency and substitution errors for HIID and SIID participants; that is, participants that experienced higher tremor amplitudes were more prone to substitution errors. For the HIID participants the relationship was between substitutions and the vertical axis [Spearman rho=.424, n=14, p=.131], while for the SIID user group we found a medium relationship with horizontal [Spearman rho=.336, n=44, p<.05] and normal [Spearman rho=.334, n=44, p<.05] axes. Regarding insertion errors, there was a strong positive relationship between mean amplitude of dominant frequency and XYZ axes [Spearman rho=.574, n=14, $p<.05$] for HIID group, accounting for 41.5% of shared variance (Figure 6.10).

Hand oscillation. Vertical axis hand oscillation was positively correlated with substitution errors in both HIID [Spearman rho=.349, n=14, $p=.221$] and SIID [Spearman rho=.335, n=44, $p<.05$] user groups. Additionally, we found a medium positive relationship between hand oscillation on the horizontal axis and substitutions for situational impaired users [Spearman rho=.471, n=44, $p<.005$]. Regarding insertion errors, there was a significant and strong relationship with the normal axis [Spearman rho=.556, n=14, $p<.05$] for older users; meaning that this axis is a good candidate to predict insertions. Regarding situational impaired users, we did not find any significant relationship with insertion errors, which was expected due to the low error rate.

Tapping likelihood. We hypothesized that one of the major reasons for inaccuracy in typing is the general movement of the device and its displacement from a stable position relatively to the user. Thus, we analyzed how the participants' movement direction affected their performance. Direction features were calculated for each axis. To do so, we compared the device's instantaneous acceleration with the mean acceleration of the device during that trial. For the horizontal axis, if the instantaneous acceleration was higher than the mean, we inferred that the device was moving rightwards; otherwise, it was moving leftwards. We performed a similar analysis for the vertical and normal axes. Although there is not a formal evaluation to this methodology results can indicate a general trend and provide new insights on how to use motion to compensate typing errors. Similarly to Crossan et al. (Crossan et al., 2005) that found that whilst walking users are more likely to perform target selections in specific phases of their movement, we found that tapping likelihood is also related to hand motion. Generally, both user populations were more likely to hit keys when the device was moving to the right. In walking conditions this corresponds to the right foot hitting the ground. Regarding the vertical axis, while situational impaired participants tapped more keys when the device was moving downwards (similar results are reported in (Crossan et al., 2005)), older adults achieved the opposite result. Finally, in the normal axis, differences between back and front directions were relatively small for HIID participants. On walking conditions this difference was more pronounced, which may be related to the nature of the gait cycle. All in all, these results suggest that participants were more likely to select keys on the deceleration phase of movement; that is, when a new step starts.

Error likelihood. Overall error (substitutions and insertions) likelihood follow the same pattern of tapping likelihood for both user groups; that is, the participants' higher likelihood to hit keys whilst moving their hands in specific directions was also reflected in higher error rates (Figure 6.11).

Direction-based analysis on substitution errors. First of all, the likelihood of substitution errors follows the same pattern of overall errors likelihood for both user groups. In this subsection, we analyze how hand movement direction influences key hits. As seen

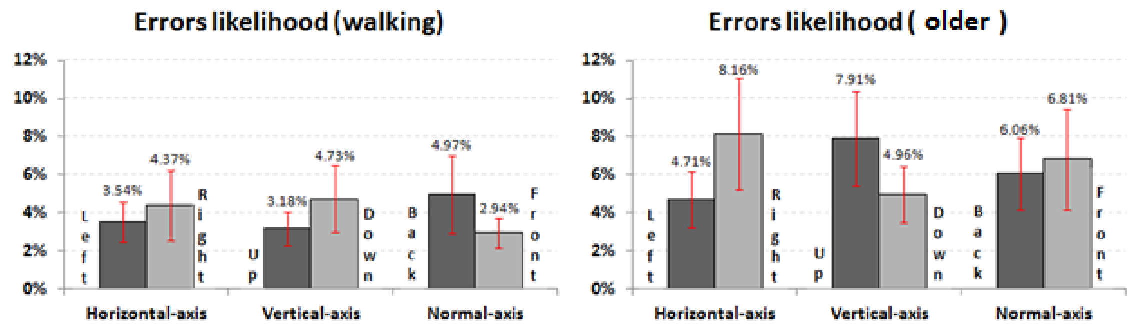


Figure 6.11: Errors (substitutions and insertions) likelihood across all axes for both user groups. Tapping likelihood follows a similar pattern. Error bars denote a 95% confidence interval.

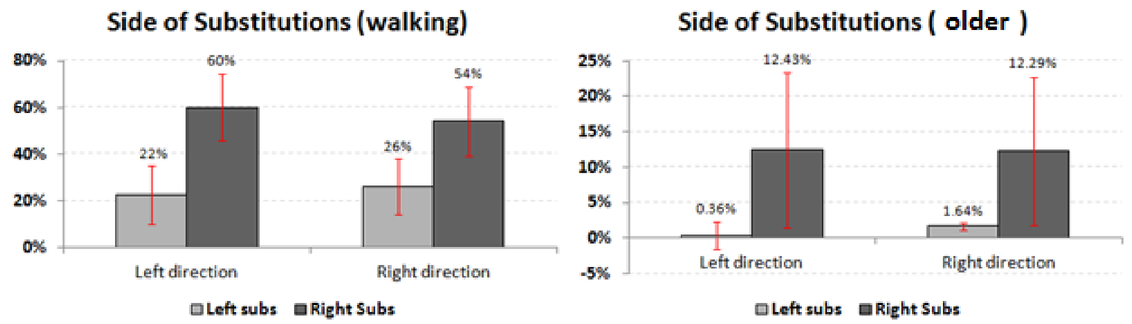


Figure 6.12: Relationship of substitution errors with horizontal directions for both user groups. Error bards denote a 95% confidence interval.

in previous chapters, hit points were slightly deviated to the right resulting in right-side substitutions for both SIID and HIID participants. Nevertheless, when performing a more thorough analysis, particularly on the horizontal axis, we observe that hand movements to the right tended to slightly increase substitutions to the left, and hand movements to the left shifted key hits to the right (Figure 6.12). Similar results were observed in (Goel et al., 2012) as users' taps tended to shift slightly to the center of the keyboard when their feet hit the ground.

Overall, several acceleration features were related with both substitution and insertion errors, showing potential to be used in modeling typing accuracy. Moreover, our findings suggest that some features can be used for both user groups, independently of impairment cause; for instance, the dominant frequency, as well as hand oscillations, were related with substitution errors in situationally- and health-induced disabilities. Nonetheless, some tremor features may be used specifically within each user group in order to accurately predict users' performance. For example, hand oscillations on the normal axis had a strong relationship with insertion errors for older users, while amplitude of the dominant frequency may be used to predict substitution errors in mobile conditions. Acceleration direction analysis also provided interesting insights: participants are more likely to commit errors in specific directions and hit points' deviations are directly affected by hand motion. These results suggest that both user groups' performance can be modeled through motion analysis, however minor adaptations may be needed to improve accuracy for each target population.

6.3. Discussion

After analyzing all collected data and describing user groups' performance, we are now able to answer the research questions proposed at the beginning and identify their main differences and similarities.

6.3.1. Typing Behaviors

Situationally-impaired users are faster typists. This behavior can be easily explained as situational impaired participants were acquainted with QWERTY keyboards, thus needed less scanning time, performed quicker touches and were consequently significantly faster than older participants.

Non-motor abilities take an important role on older adults' performance. One of the main differences between SIID and HIID groups were some of the substitution patterns. Namely confusion between similar letters were very common and consistent for some older participants. Moreover, forgetfulness in HIID was a recurring issue, originating high omission rates. The lack of experience in typing tasks or improper mental models of the required sentences may be related with this result.

Similar (relative) magnitude of insertion errors. Dealing with insertion errors shows to be equally important for both situational and health impaired users. We believe this type of error to be easily identifiable and automatically discarded by monitoring typing patterns. Nonetheless, personalization may play an important role in this solution as users' typing patterns may vary (see Chapter 5).

Substitutions continuum. We found substitution errors to be the most similar error type between user groups. In fact, SIID and HIID participants were equally affected by the increased demand of conditions, suggesting that designers will be able to reuse and leverage existing knowledge towards more inclusive solutions.

HIID participants have additional difficulties. Although we observed that HIID and SIID participants performed a "continuum" of substitution errors, the former experienced additional difficulties; namely, older participants are more prone to insertion and omission errors.

6.3.2. Tremor Profile

SIID profile is more consistent across individuals. Results show that SIID participants' acceleration pattern is more consistent across participants; the walking pattern resulted in similar dominant frequencies for all users, resulting in a similar hand motion profile. On the other hand, older participants' profile has higher variance and it seems to be user-dependent.

Tremor amplitude is lower for older adults. Situational impaired participants experienced higher tremor amplitude than older users. In fact, both measurements of tremor amplitude were higher for SIID participants; that is, mean amplitude of dominant frequency and hand oscillations whilst walking were always higher across all motion axes.

Older adults' tremor is more irregular. Although tremor amplitude is lower for HIID users their hand motion is characterized by ranging from low to high frequencies, particularly in the horizontal and vertical axes, suggesting a much more irregular tremor profile. This sometimes holds true for individual participants and may be the reason for older participants typing inaccuracy.

6.3.3. Relationship with Accelerometer Features

Similar motion features correlated with input performance for both user groups. Results show that the amplitude of the dominant frequency positively correlated with text-entry performance for both target populations. Moreover, hand oscillations also correlated with input errors, showing that tremor amplitude explains a high degree of variance across user groups.

Minor adjustments may be needed. Even though several similarities were found between target populations, some features may be more useful for one group than another. For instance, the experienced dominant frequency during text-entry trials only explains SIID participants' substitution error rate variance. Also, the amplitude of tremor has a different correlation with each user group depending on the axis: the horizontal axis may be useful to predict SIID performance while the vertical axis can be used for HIID participants. Thus, when modeling users' abilities and performance in text-entry tasks some adjustments may be needed according to the target population.

Motion direction showed to be a promising feature. Based on the premise that tapping accuracy is due to hand motion, we analyzed hand direction whilst typing and showed that SIID and HIID participants are more likely to perform errors in specific directions, which may be related to hand dominance. Also, both populations were equally affected by motion, particularly on the horizontal axis, slightly shifting their hit points to the center of the keyboard. Overall, these features showed promising results to be used to compensate touch locations.

6.4. Recommendations for Design

Enhance typing experience using accelerometer data. Several motion features, such as dominant frequency, tremor amplitude, hand oscillation, and motion direction showed potential to be used to compensate typing errors. Thus, future mobile text-entry systems should take an adaptive approach and use the devices' built-in accelerometer to compensate for hand movements.

Similar features with minor adjustments. Both older and situational impaired users seem to benefit from accelerometer data to accommodate hand tremor; however, minor adjustments regarding which features to include in classification models may be needed. Some features showed stronger correlations with a specific user group while others could equally benefit both populations.

Build data-driven models. This result is directly correlated with users' tremor profile: situational-induced tremor showed a clear motion pattern with high accelerometer amplitudes, while health-induced tremor featured small oscillations and irregular tremor. Classification models should be built with each user group's data. Also, personalization may play a crucial role for older users, since there was a high range of tremor profiles across participants.

6.5. Conclusion

We have investigated text-entry performance of both situational and health impaired users on touch-based devices and report their main differences and similarities. Our goal was to raise awareness and provide the knowledge to develop broader and effective solutions. Indeed, results suggest that with the advances of mobile technologies there is not a clear distinction between able-bodied and disabled users, but rather a "Disability Continuum". Older adults experience common substitution errors with mobile users; however, they are also more amenable to omission errors. Moreover, motion analysis revealed that although situational- and health- induced tremor is different in its nature, several motion features can be used to model and predict both target populations' typing performance.

Following this work, we intend to analyze hand tremor features that were captured through the device's sensors and develop motion-enhanced models that can be applied to both user groups.

7 Leveraging Motion Data as a Unifying Feature

Touchscreen devices have become a pervasive platform for mobile computing; yet, touch typing on these devices is a major challenge as no physical stability or tactile feedback are given by the small sized virtual keys. Moreover, when considering hand tremor, either induced by a situation (e.g. walking) or health condition (e.g. physiological tremor), inputting text is even more difficult (see previous chapters). Wrong keys are often tapped introducing substitution errors, and accidental touches result in insertion errors. In this chapter, we propose the use of a text-entry solution that can be applied to a wide range of abilities.

We faced this problem as a classification problem, where each key press is either correct or incorrect. The challenge is knowing how to classify key presses correctly. Moreover, since we wanted to be able to deal with insertion and substitution errors, we developed two different kinds of classifiers. Omission errors were not considered because they were not related to motor abilities (see Section 5.3.7).

We present several error-compensation classifiers that take advantage of both touchscreen and motion information. Our goal is to enhance text-entry accuracy using mainstream devices' inertial sensing capabilities to compensate typing errors. Several authors have worked on adaptive text-entry; however, their main focus has been on adapting the keyboard based on language models (Al Faraj et al., 2009)(Goodman et al., 2002)(Gunawardana et al., 2010)(Kane et al., 2008a). Also, to our knowledge, taking advantage of the device's motion as a unifying measure between situational- and health-induced impairments has not been explored in the past.

In order to design and develop our classifiers, we used the data collected from the user studies presented in Chapters 4 and 5. Based on data from both user groups, we built insertion and substitution classification models that include touch and motion data. We then compared their performance for both HIID and SIID participants. Moreover, we assessed the effect of personalization and transferring models between target populations.

7.1. Background

In this section, we analyze previous works that attempted to make interaction with devices easier, particularly in text-entry tasks. Mackenzie et al. (MacKenzie and Soukoreff, 2002b) provide a full review of the plethora of input methods developed in the last decades. Several approaches that tried to improve text-entry performance with QWERTY keyboards have been proposed. For instance, Kane et al. (Kane et al., 2008a) presented an auto-correct system to help motor impaired users during input tasks. TrueKeys combines models of word frequency, keyboard layout, and typing error patterns to automatically identify and correct typing mistakes. The Dynamic Keyboard (Trewin et al., 1997), also designed for people with motor impairments, continuously adjusts keyboard accessibility features by monitoring typing patterns and using bigram frequencies.

Other interaction techniques have also been used to input text, for example, using geometric pattern matching (Kristensson and Zhai, 2005) and gestures (Kristensson and Zhai, 2004). Other authors have explored the combination of both touch and language models (Goodman et al., 2002)(Gunawardana et al., 2010) to improve input performance. These models predict the next letter to be typed and adapt key size, without showing the adaptation to users (Apple's virtual keyboards use this approach). Gunawardana et al. (Gunawardana et al., 2010) proposed the use of key anchors when adaptations are invisible. This ensures that the center area of visible keys (anchor) always returns that letter, providing a predictable result for the user's actions. User studies revealed significant improvements in comparison with the more aggressive key-target resizing techniques.

Models of key-press distributions built on data collected from users are very common on adaptive keyboards. One of the most used approaches relies on bivariate Gaussian distributions of individual keys (Goodman et al., 2002)(Gunawardana et al., 2010)(Rudchenko et al., 2011). Rudchenko and colleagues (Rudchenko et al., 2011) performed key resizing, horizontal and vertical, based on touch positions using bivariate Gaussian distributions for each key. Also, the authors show that user-dependent models may improve overall accuracy.

Other approaches that have been used resort to adapting a key's location based on the centroid of users' previous finger touches (Himberg et al., 2003)(Go and Endo, 2007)(Al Faraj et al., 2009)(Findlater et al., 2011). Himberg and colleagues (Himberg et al., 2003) use this technique and form a Voronoi tesselation that represents the virtual keyboard. Findlater (Findlater et al., 2011) ran a simulation on larger devices, using the distance to the closest centroid as a proxy to the intended key, and showed an improvement in typing performance. However, others did not find this approach beneficial, especially when adaptations are visible (Go and Endo, 2007)(Al Faraj et al., 2009). In 2012, Findlater and Wobbrock (Findlater and Wobbrock, 2012) introduced a keyboard that adapts to suit each user's typing pattern using user-dependent data.

Others have also proposed solutions for mini-QWERTY physical keyboards. For example, Clawson et al. (Clawson et al., 2007) proposed a model to detect and ignore insertion errors on these devices, taking advantage of typing features. Afterward, the authors extended their algorithm to deal with substitution errors; however, language features had to be included: bi- and tri-letter frequencies (Clawson et al., 2008).

More recently, researchers have been focusing on solutions to be used "in the wild" and whilst mobile. Henze et al. (Henze et al., 2012) gathered typing data from everyday use and identified a systematic skew of finger-touches. Next, the authors derived a function that compensates this skew by shifting touch events, which resulted in an accuracy improvement. Nicolau et al. (Nicolau et al., 2012a) explored the usage of virtual keyboards whilst mobile and performed visual adaptations in order to deal with an increased visual demand. Kane et al. (Kane et al., 2008b) also investigated visual adaptations in walking conditions using accelerometer information to enlarge targets. However, this adaptive approach did not show improvements over the static interface. Similar results were reported when trying to compensate vibration during reading tasks by constantly moving the screen content in the opposite direction (Rahmati et al., 2009). Yamabe and Takahashi (Yamabe and Takahashi, 2007) have also used an inertial sensor to automatically adapt font size and images whilst walking. On the other hand, Goel et al. (Goel et al., 2012) showed that text-entry accuracy can be improved using accelerometer data whilst walking; however, to our knowledge, leveraging motion data as a unifying measure between health and situational impairments has not been investigated.

Overall, previously proposed models that aim to improve typing accuracy have focused on one type of input error, either substitutions or insertions. Those who address both types of errors make use of linguistic models and are, therefore, dependent of one specific language.

7.2. Machine Learning Methodology

In this chapter we use machine learning (or data mining) techniques to automatically classify users' key presses into specific categories. We use these techniques because they are usually able to produce classifiers of higher accuracy than those created using simple decision procedures or heuristics.

The main goal of machine learning techniques is to process large volumes of data and "learn" hidden patterns. It is about solving problems by analyzing data already collected. Similarly, knowing how to classify typing errors is about identifying particular patterns of interaction as belonging to a specific category.

A category, or class, is a collection of instances that share some common characteristics that identify them as members of that class. An instance is an individual, independent example of the "concept" or model to be learned. The input of every machine learning technique is a set of instances, called dataset. Each instance is characterized by its pre-defined set of attributes or features. For instance, a key press can be characterized by its duration and touch position. The values of these attributes are then used by the learning technique in order to generalize knowledge and formulate categorizations.

Machine learning techniques are used to train classifiers to make categorizations from example data (training instances). They look for patterns in the data that are able to accurately describe each class. Generally, a classifier is an algorithm that is able to identify whether an instance belongs to a given class. One example would be to identify whether a key press is an accidental touch by looking at the details of that action (i.e features). Statistical models, or classifiers, attempt to predict this.

A classifier works by finding the amount of correlations between features and the class it is trying to predict. Once a statistical model has been created (learned), it can then be used to classify or predict the class of new data (instances).

Most machine learning techniques are designed to learn which are the most appropriate attributes to use for making their decisions. Nonetheless, adding irrelevant attributes to a dataset often confuses machine learning schemes. To counteract this effect, we preceded learning with an attribute selection stage, using feature selection algorithms that aimed to provide an indication of the most relevant attributes for constructing classifiers. We used a wrapper-based approach, rather than filter methods. It is called wrapper-based because the learning algorithm is "wrapped" into the selection procedure; that is, the algorithm automatically and recursively performs an assessment of a subset of features using the machine learning technique that will be employed for learning.

To choose the machine learning techniques to be used in the building of our classifiers, we adopted a "simplicity-first" methodology where simple techniques, such as 1-rule classification, were tested first. Nevertheless, even though simple ideas often work very well, we also evaluated more complex techniques, recurring to linear and statistical models. The techniques were chosen to provide a wide range of approaches, namely: Rules, Statistical Modeling, Decision Trees, Linear Models and Instance-based Learning. For each of these approaches we chose one or more techniques, from the Weka data mining tool[1]: Rules (1-Rule), Statistical Modeling (Naïve Bayes), Decision Trees (C4.5), Linear Models (Simple Logistic, Voted Percepton), and Instance-based Learning (IB1).

These techniques were then evaluated by analyzing their classification accuracy. The most accurate statistical model was then used in our remaining analysis. In this evaluation process, we used cross-validation; a common technique that simulates the evaluation of a classifier performance with previously unseen data. In cross-validation, there is a fixed number of folds or partitions. Data is then split into this number of partitions and each partition in turn is used for testing and the remainder are used for training. The process is repeated so that in the end, every instance has been used exactly once for testing. In our analysis we used a tenfold cross-validation, since this has become the standard method. Moreover, we use stratified samples, where each class is properly represented in both training and test sets.

7.3. Building Instances

In order to create and evaluate our classifiers, we first needed to analyze all typing data collected from the user studies described in Chapters 4 and 5. However, we had to restrict our analysis. For the older adults study, we chose to analyze performance with the *mobile device*, since it was held during typing tasks, thus motion data is available. We also chose this dataset because this was the condition where users faced the biggest challenges. On the other hand, for the situational impaired users, we only considered the *normal walking* condition, where tremor played a significant role, and one hand posture - the *one-hand portrait*. This interaction mode was chosen due to: first, its similarities with how older users interacted with their mobile device (also with one hand); second, because one-hand interaction is often required in mobile contexts (e.g. hands-busy situations).

To analyze all data, we built a custom feature collection application, *Feature Collector* (Figure 7.1), that takes as input the touch and accelerometer log files, and produces a file that contains a set of instances. The application allows users to select the desired features, which will characterize each instance, as well as the intended class (insertion or substitution).

Regarding the input files, touch logs contained information about finger-down, finger-movement, and finger-up events; each of which comprised three attributes: touch position (x, y), timestamp, and current key. The accelerometer logs contained all readings from

[1]http://www.cs.waikato.ac.nz/ml/weka/

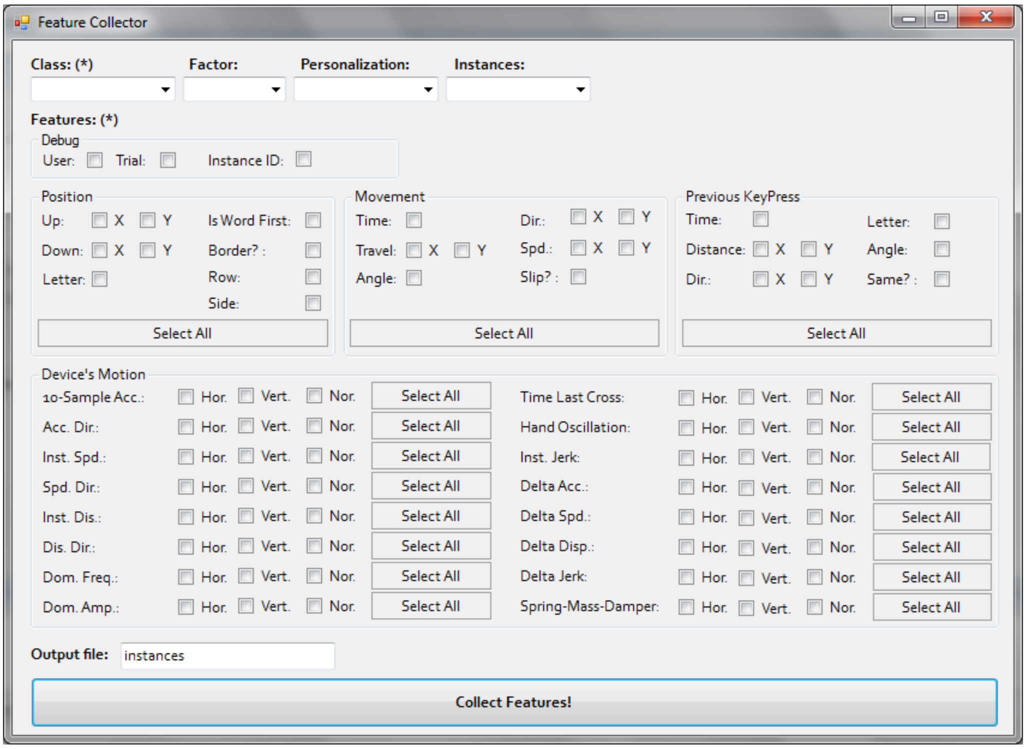


Figure 7.1: Screenshot of our Feature Collector application.

the device's built-in tri-axis inertial sensor with the respective timestamp. These files enabled us to compute all touch and motion features that will be described in the following sub-sections. Overall, 26 touch-related features and 48 motion-related features can be extracted with the *Feature Collector*.

7.3.1. Touch Features

The features described in this sub-section were calculated by analyzing the touch log files in isolation. We outline four main domains of features that can be derived from the mobile device's touch information: position, key, movement, and previous key press. These features can be computed for each key press.

Touch Position Features

Touch position features are based on finger-down and finger-up locations.

Down Event (X, Y). Integer values that correspond to the initial position (x, y) of key press events, also known as aiming position.

Up Event (X, Y). Absolute horizontal and vertical positions that correspond to finger lift action. Most times this position is different from the down event. Yet, it is the up event that originates the character insertion.

Key Related Features

These features are related to keyboard actions and layout. For instance, border keys may originate more errors or perhaps left-side keys may be more difficult to press.

Transcribed Letter. This feature corresponds to the character inserted by the user. Nevertheless, this may not be the intended or required character. There are 27 possible values for this feature (26 alphabet letters plus the space bar character).

Is Word First Character. Indicates whether the transcribed character is inserted after a blank space or is in the beginning of a sentence (boolean value).

Is Border Key. This is a boolean feature. It is set to true if the transcribed character is one of the following: Q, A, Z, X, N, M, L, P, Blank space; otherwise, it is set to false.

Row. This feature specifies the keyboard row of the transcribed character.

Side. Similarly to previous feature, it indicates the keyboard side of the transcribed character. Right side comprises the keys P, O, I, U, Y, L, K, J, H, M, N, and B; all remaining keys belong to left side. The blank space does not belong to either side since it is in the center of the keyboard.

Is Slip. This feature indicates whether the key initially touched by users (finger-down event) is the same when they lifted their finger (finger-up event). If so, there is no finger slip and the value is set to false; otherwise, it is set to true.

Movement Features

Movement features describe the movement of the finger during the key press action. Different keys may result in different movement behaviors due to their location in the keyboard.

Movement Time. Also known as key press time, it represents the time the user's finger is on the screen.

Movement Distance (X, Y). Overall traveled distance in both axes, calculated as the Euclidean distance between finger-down and finger-up positions.

Movement Direction (X, Y). Direction of finger movement in both horizontal (left/right) and vertical (up/down) axes.

Movement Angle. Angle between finger-down / finger-up positions and the horizontal axis, calculated as: atag (movement distance X, movement distance Y) * 180 / Π.

Movement Speed. Average movement speed in pixels per second.

Features Relative to Previous Key Presses

These features comprise knowledge about previous key presses and their relationship to the current action. For instance, tapping two consecutive keys on the right side of the keyboard may be different from pressing two keys in opposite positions.

Inter-Key Time. Elapsed time, in milliseconds, between key presses; that is, time between the last finger-up and current finger-down events.

Previous Transcribed Letter. This feature corresponds to the last character inserted.

Distance to Previous Key (X, Y). Euclidean distance between last finger-up and current finger-down positions.

Direction to Previous Key (X, Y). Overall x- and y- directions between last finger-up and current finger-down positions.

Angle to Previous Key. Angle between last finger-up / current finger-down positions and the horizontal axis.

Is Same Key. This feature indicates whether the previous entered character is the same as the current transcribed character.

7.3.2. Motion Features

In this sub-section, we describe how motion features were calculated through our accelerometer log files. We divided these features by type of motion: acceleration, speed, displacement, frequency domain, hand oscillation, jerk, and spring-mass-damper.

To calculate all motion features, the data from the mobile device's accelerometer was first passed through a low-pass filter to remove noise (Casiez et al., 2012) above 12Hz. This value was chosen based on current research on maximum frequency for physiological tremor (Elble and Koller, 1990). Also, all acceleration values from situational-induced impairments are well below the 12Hz threshold. Moreover, when dealing with accelerometer data it is often required to compensate for gravitational pull on the three axes. Preliminary tests showed this compensation to be unnecessary because device orientation remained approximately constant during typing tasks. Next, we describe all motion features available in our *Feature Collector* application.

Acceleration Features

10 Sample Acceleration (Horizontal, Vertical, Normal). This feature actually corresponds to 10 different features per axis. Acceleration data was re-sampled to 10 samples between two consecutive key presses. This sampling rate was selected as it gives a reasonable resolution and does not overly increase the number of acceleration features for classification.

Acceleration Direction (Horizontal, Vertical, Normal). Direction of acceleration, either positive (accelerating) or negative (decelerating), between last and current finger-up events. This feature gives us a general idea of how the device moved between key presses.

Delta Acceleration (Horizontal, Vertical, Normal). Acceleration difference (for all axes) between last and current finger-up events.

Speed Features

Instantaneous Speed (Horizontal, Vertical, Normal). To calculate speed features, we single-integrated filtered acceleration samples using the trapezoidal cumulative sum. The result corresponded to instantaneous speed. We repeated this process for each axis.

Speed Direction (Horizontal, Vertical, Normal). This feature corresponds to the overall direction of instantaneous speed, either positive or negative, between last and current finger-up events.

Displacement Features

Instantaneous Displacement (Horizontal, Vertical, Normal). In order to determine the instantaneous displacement between key presses, we double-integrated filtered acceleration samples using the trapezoidal cumulative sum technique. The value was calculated for each axis, giving a good proxy for magnitude of motion.

Displacement Direction (Horizontal, Vertical, Normal). The direction of the device's displacement was calculated by comparing the instantaneous acceleration with the moving mean acceleration of the device for each key press (Goel et al., 2012). If the instantaneous acceleration was less than the mean, we inferred that the device was moving left, down, and backwards for the horizontal, vertical, and normal axes, respectively. Otherwise, it was moving right, up, and forward for the horizontal, vertical, and normal axes, respectively.

Frequency Domain Features

Dominant Frequency (Horizontal, Vertical, Normal). To calculate the instantaneous dominant frequency of users' motion, we applied the Fast Fourier Transform (FFT) on the accelerometer signal and found the maximum amplitude. The result was the dominant frequency between two key presses.

Amplitude of Dominant Frequency (Horizontal, Vertical, Normal). This feature corresponds to the amplitude of the dominant frequency.

Time Elapsed Since Last Mean Cross (Horizontal, Vertical, Normal). This attribute indicates the time elapsed between last cross of the acceleration moving mean and finger-up event. This is a useful feature to pinpoint where in the acceleration signal a tap occurs, relative to the mean movement.

Hand Oscillation Features

Hand Oscillation (Horizontal, Vertical, Normal). Hand oscillation features were calculated as standard deviation of accelerometer values. This measure was found to be a good predictor of target selection accuracy whilst walking (Bergstrom-Lehtovirta et al., 2011). Also, it showed strong correlations with input errors in previous chapters. The higher the standard deviation, the greater the hand oscillation experienced by the user during that key press.

Jerk Features

Jerk (Horizontal, Vertical, Normal). In physics jerk is a measure of rate of change of acceleration. We calculated jerk as the first derivative of acceleration with respect to time, resulting in a scalar value. Jerk features can be seen as the change of acceleration "force" during a key press.

Spring-Mass-Damper Features

Spring-Mass-Damper (Horizontal, Vertical, Normal). Spring-Mass-Damper (SMD) features draw inspiration from the well known physical model where a mass is suspended in each direction by a critically-dampened spring and damper (see Figure 7.2). The springs allow the mass to remain relatively steady during shaking, while the dampers prevent oscillation of the mass. The SMD model aims at dynamically compensating for frequent and random movements by shifting the mass in the opposite direction of the shake. This seems to be a good approach to compensate for the device's motion, however it is yet to be proven its benefits (Rahmati et al., 2009).

For each axis, we calculated the displacement, Y(t), taking into account the acceleration readings, A(t), and the impulse response of the spring-mass-damper system, H(T):

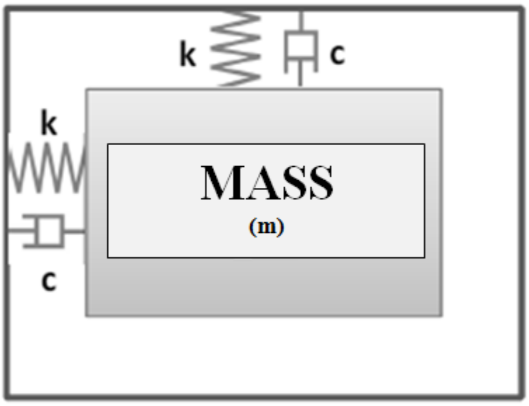


Figure 7.2: Spring-Mass-Damper physical model.

$$Y(t) = H(t) * -A(t)$$

Where H(t) is the impulse response from the equation:

$$H(t) = te^{-t\sqrt{k}}$$

Note that H(t) diminishes quickly as t increases. The acceleration readings were stored in a circular buffer of approximately 4 seconds, because the impulse response approaches zero in this time period. The constant k represents the spring rate. A larger k is analogous to firmer springs. We performed preliminary tests to assess the best value for constant k. We tested integer values from 1 to 20, and choose a spring rate of 7 N/m, since this was shown to yield the best classification results.

Concluding, the SMD features consist in the displacement compensation, Y(t), outputted by the physical model. This will allow classifiers to derive correlations between this displacement and typing errors.

7.3.3. Instances

The input to a machine learning scheme is a set of instances. These instances are the examples that are to be classified. This sub-section describes the instances used for our classification training and testing. These were built from typing data previously collected. Regarding the HIID instances, we used the typing data collected in the user study described in Chapter 5. For the SIID dataset we had to re-perform the user study described in Chapter 4, due to a logging bug that prevented us from computing all required touch features. We recruited 19 new participants from our university and asked them to type 10 sentences in two mobility conditions (seated and walking) and one hand posture (one-hand portrait). Results regarding input speed, text quality, hand motion, and mobility effect were very similar to the ones described in Chapters 4 and 6. Participants' profiles were also very similar and are illustrated in Table 7.1.

Participant	Age	Gender	Dominant hand	Mobile experience	QWERTY experience	Touchscreen experience
#1	25	male	Right	Yes	Yes	Yes
#2	27	male	Right	Yes	Yes	No
#3	25	female	Right	Yes	Yes	No
#4	28	male	Right	Yes	Yes	No
#5	31	male	Right	Yes	Yes	Yes
#6	25	female	Right	Yes	Yes	Yes
#7	27	male	Right	Yes	Yes	No
#8	24	male	Right	Yes	Yes	Yes
#9	24	male	Right	Yes	Yes	Yes
#10	30	male	Right	Yes	Yes	Yes
#11	25	male	Right	Yes	Yes	Yes
#12	27	male	Right	Yes	Yes	Yes
#13	28	male	Right	Yes	Yes	Yes
#14	28	male	Right	yes	Yes	Yes
#15	26	male	Right	Yes	Yes	Yes
#16	24	female	Right	Yes	Yes	No
#17	24	Female	Right	Yes	Yes	Yes
#18	24	Female	Right	Yes	Yes	Yes
#19	27	Male	Right	Yes	Yes	Yes

Table 7.1: SIID participants profile.

Our *Feature Collector* application, allowed the selection of 74 features and 4 classes: substitution, insertion, bounce, and accidental touch. Bounces and accidental touches instances are sub-classes of insertion errors.

Note that we collected data following a traditional text-entry task (MacKenzie and Soukoreff, 2003); we gave users sentences to copy so that we always knew what keys they were intending to hit. This procedure, besides giving a reliable training dataset, allows us to obtain a real distribution of data for training models. We are effectively learning that when users tried to type a given letter, they sometimes hit the right-adjacent key with some frequency and the left-adjacent with a different frequency. In fact, our laboratory testing conditions have proved in the past to be able to successfully model typing behaviors and error-compensation solutions (Gunawardana et al., 2010)(Findlater et al., 2011)(Goel et al., 2012).

Since we were eliciting natural typing patterns by not concerning users with correction tasks, the final sentences had mistakes. Some of those errors were indeed attempts to hit the intended key, others added noise to our data (e.g. confusion between similar letters). Therefore, we removed outliers by eliminating all key presses that landed outside the Euclidean bounds of the intended key or its immediate neighbors. This allowed us to accommodate transposition errors, for instance in the word "but", the character 'u' was entered before 'b'. This also enabled us to remove errors where users confused the letter to be entered. Overall, 2.24% and 0.56% key presses were filtered out in this process for HIID and SIID datasets, respectively.

Each dataset incorporates examples of correctly inserted characters, as well as examples of erroneous key presses. Classification schemes always optimize their accuracy, assuming that all instances have equal weight. In our insertion dataset, correct instances (no insertion instances) outnumbered error instances (yes instances) by a factor of 10. This resulted in a decision structure clearly biased towards avoiding errors on the no instances, because such errors are effectively penalized tenfold. Thus, the classifier maximized its accuracy, but was unable to identify insertion errors. In a two-class situation, such as identifying whether a key press is an insertion or not, a simple and general way to make any learning scheme cost sensitive is to duplicate instances in the dataset (Witten and Frank, 1999). Thus, we duplicated insertion instances (yes instances) by a factor of 10 in order to balance their weight. Accidental touches and bounce instances were also duplicated tenfold and threefold, respectively. Table 7.2 show the number of instances for each dataset.

	HIID (average per user)	**SIID** (average per user)
Substitution	1405 (100)	6537 (344)
Insertion	2958 (211)	10903 (574)
Bounce	2483 (177)	8793 (463)
Accidental Touch	181 (13)	703 (37)

Table 7.2: Number of instances in each dataset.

7.4. Research Questions

This chapter aims to answer the following research questions:

1. *What machine learning techniques are the most effective in predicting errors?*
2. *Do motion features enhance text-entry accuracy?*
3. *Does personalization affect the classifiers' accuracy?*
4. *Can we use data from HIID user group to build solutions to SIID users, and vice-versa?*
5. *Can we use similar classifiers and motion features for HIID and SIID?*

7.4.1. Evaluation Measures

After designing and building our models (see next section), we measured their performance in terms of accuracy and kappa statistic. Accuracy is simply the proportion of correctly classified instances over the whole dataset and it measures the overall performance of the model. Obviously, we applied different evaluation techniques to avoid bias associated with learning, such as stratified cross-validation. The other measure used to predict performance was the kappa statistic. The kappa statistic is used to measure agreement between the predicted and observed values of a dataset, while correcting for an agreement that occurs by change. This measure deduces the value of a random predictor from the model's accuracy and expresses the result as a proportion of the total for a perfect predictor (Witten and Frank, 1999). The maximum value of kappa is 100% and the expected value for a random predictor is 0%. Kappas over 75% are excellent, 40%-75% are fair to good, and below 40% are poor (Fleiss, 1973).

Regarding statistical analysis, Shapiro-Wilkinson tests were applied to accuracy results to assess normality. To appraise statistical significance, we applied parametric statistical tests, such as repeated measures ANOVA and t-test for normally distributes values, and non-parametric tests (Friedman and Wilcoxon) otherwise. For the repeated measures ANOVA, Greenhouse-Geisser's sphericity corrections were applied whenever Mauchly's test of sphericity showed a significant effect. For all pair-wise post-hoc tests, we applied Bonferroni corrections.

7.5. Dealing with Insertion Errors

As seen in Chapter 6, dealing with insertion errors is equally important for both situational and health impaired users. Our prior analysis also indicated that insertion errors could be easily identifiable by monitoring time-based patterns, such key press duration and inter-key interval. Thus, in this section we design and evaluate classification models aimed at identifying insertion errors.

7.5.1. Design of Insertion Classifiers

Knowing how to identify insertion errors is the most important stage of the classification process, since dealing with this type of error comes down to discarding key presses. Therefore, we built some classifiers that categorize each key press as either a unintentional or correct press (i.e yes/no). In the next sub-section, we present three touch-based classifiers (Simple Time-based, Complex Time-based, and Bounces & Accidental Touches) and two classifiers enhanced with motion data (Acceleration and Tremor Pattern). We describe them and explain the rationale for choosing their features. Since insertion errors mainly depend on timing features (although position can also be relevant), touch-based models will herein be referred to as time-based classifiers.

Simple Time-based Classifier

Description: Previous results suggest that insertion errors are usually characterized by short key press durations and small inter-key intervals (see Chapter 5). The classifier uses these two features to identify unintentional screen taps. While simple, it may be able to identify most insertion errors.

Features (2): Inter-key interval and Movement time.

Complex Time-based Classifier

Description: This classifier uses the same rationale as the previous one; however, it includes a new set of features that could be useful to discriminate small variances in data. For example, we included the inserted character, as a new feature, in order to identify individual key related issues. Moreover, since bouncing errors (i.e. unintentionally tapping the same key more than once) are a known issue, we added a feature indicating if the current and previous characters are the same. Finally, we hypothesized that unintentional touches are closely related to finger movement on the screen, therefore, we added six new attributes: movement distance (x, y), movement speed (x, y), and movement direction (x, y).

Features (10): Inter-key interval, Movement time, Transcribed character, Movement distance (X, Y), Movement speed (X, Y), Movement direction (X, Y), and Same key attribute.

Bounces & Accidental Touches Classifier

Description: This is a composite classifier that takes touch data and analyzes it through two classifiers: bounces and accidental touches. Our hypothesis is that analyzing both types of errors separately will increase the overall accuracy. Bounces classifier deals with repeated keys, while accidental touches classifier deals with the remaining insertion errors. Both classifiers use the same set of features of Complex Time-based classifier.

Features (9): Inter-key interval, Movement time, Transcribed character, Movement distance (X, Y), Movement speed (X, Y), Movement direction (X, Y).

Acceleration Classifier

Description: We believe that one of the reasons for insertion errors is the general sudden movements of the device. Thus, this classifier builds upon the Complex Time-based classifier by incorporating motion features in all three axes. Particularly, data from vertical and normal axes will probably be the most relevant to insertion errors, since they may indicate movements towards the users. We included the 10 acceleration samples described in Section 7.3.2, which make up 30 features for the classifier. Displacement features were also envisaged in the design of this classifier; however, preliminary tests did not show an improvement of performance, thus they were not included.

Features (40): Complex Time-based features, 10 Acceleration samples (Horizontal, Vertical, Normal).

Tremor Pattern Classifier

Description: When analyzing motion data, we observed strong correlations with hand oscillation and tremor amplitude features (see previous Chapter). Thus, in addition to the Complex Time-based set of features, the Tremor Pattern classifier incorporates 3 new features per axis. First, to make the model adaptive to different tremor behaviors, we included the instantaneous dominant frequency of users' hand motion. Second, we added the amplitude of the dominant frequency, since this is a commonly used measure to characterize tremor amplitude. Indeed, these two features give us an approximation of the users' hand motion frequency and intensity. Finally, in addition to frequency domain features, we also added the device's oscillation between finger touches.

Features (19): Complex Time-based features, Dominant frequency (Horizontal, Vertical, Normal), Amplitude of dominant frequency (Horizontal, Vertical, Normal), Hand oscillation (Horizontal, Vertical, Normal)

7.5.2. Evaluation of Insertion Models

The evaluation presented in this section is four-fold: comparison between machine learning techniques, comparison between models, personalization effect, and user group effect.

Comparison between Machine Learning Techniques

In this section, we assess the effect of different machine learning techniques on the previously proposed (insertion) classifiers. We used a 10 times stratified 10-fold cross validation,

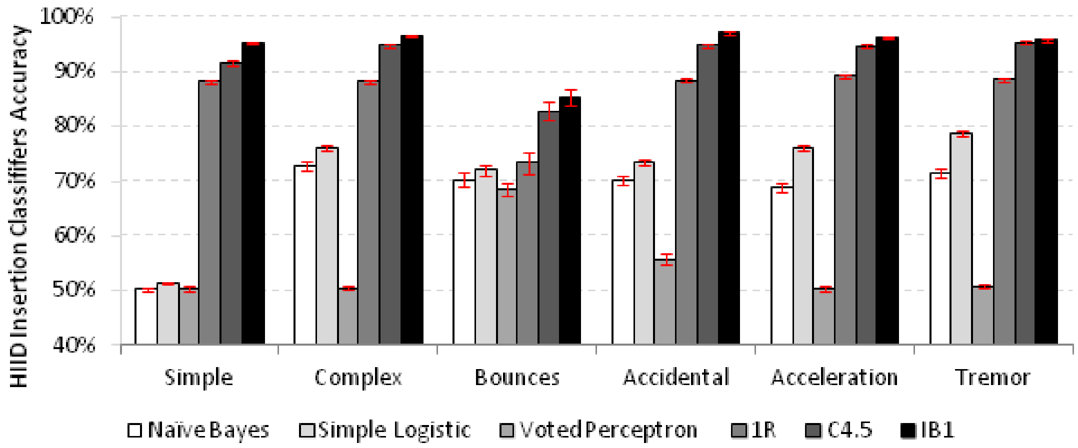


Figure 7.3: HIID insertion classifiers accuracy. Error bars denote 95% confidence interval.

which involved invoking each learning algorithm 100 times on datasets that are nine-tenths the size of the original. All models were built using Weka's default values.

In the remainder of this section we present results within each classifier for both user groups. Figures 7.3 and 7.4 illustrate classifiers' mean accuracies for HIID and SIID datasets, respectively.

	Accuracy (Kappa) %	Naïve Bayes	Simple Log.	Voted Per.	1-Rule	C4.5 Tree	IB1
HIID	Simple	50.3 (3)	51.39 (0)	50.46 (2)	88.21 (76)	91.64 (83)	95.31 (91)
	Complex	72.84 (46)	76.04 (52)	50.55 (1)	88.21 (76)	94.82 (90)	96.73 (93)
	Bounces	70.35 (12)	72.04 (24)	68.51 (0)	73.39 (30)	82.86 (55)	85.4 (59)
	Accidental	70.12 (37)	73.39 (46)	55.79 (6)	88.43 (77)	94.8 (90)	97.12 (94)
	Acceleration	68.72 (38)	76.6 (53)	50.38 (2)	89.18 (78)	94.88 (90)	96.32 (93)
	Tremor	71.47 (43)	78.79 (57)	50.87 (1)	88.53 (77)	95.39 (91)	95.8 (92)

Table 7.3: HIID insertion classifiers' mean accuracy and kappa statistics. For each line, different background colors illustrate significant differences on accuracy between machine learning techniques.

Simple time-based classifier. Regarding HIID data, a repeated measures ANOVA revealed significant differences on accuracy [$F_{3.297,326.371}$=20607.906, p<.001] (see first line of Table 7.3). Post-hoc tests using Bonferroni correction showed significant differences between IB1 and all remaining machine learning techniques. Concerning SIID data, as seen in Table's 7.4 first line, results showed statistically significant differences [$F_{3.235,320.254}$=13853.17, p<.001], namely between IB1 and the remaining conditions.

Complex time-based classifier. Using the HIID dataset and accordingly to the results presented in Table 7.3, there were significant differences between machine learning techniques [$F_{2.857,282.829}$=6096.819, p<.001]. The IB1 technique was the most accurate. For the SIID classifiers, results followed a similar trend (second line of Table 7.4) with significant differences between techniques [$F_{2.749,272.133}$=26930.426, p<.001]. Post-hoc tests revealed that IB1 is significantly more accurate than all remaining machine learning techniques.

Bounces classifier. For the bounces classifier, using the HIID dataset, we also found significant differences between conditions [$F_{4.453,440.808}$=91.55, p<.001]. Both C4.5 Tree and IB1 were significantly more accurate than all remaining techniques. Moreover, there were no significant differences between these two conditions [p<.001]. Regarding the SIID dataset, there were significant differences in the results [$F_{3.325,329.218}$=587.558, p<.001] (third line of Table 7.4), with both C4.5 Tree and IB1 outperforming the remaining machine learning techniques.

	Accuracy (Kappa) %	Naïve Bayes	Simple Log.	Voted Per.	1-Rule	C4.5 Tree	IB1
SIID	Simple	64.83 (20)	66.67 (25)	63.24 (20)	89.8 (44)	89.8 (79)	95.23 (90)
	Complex	86.32 (71)	88.29 (75)	64.05 (22)	87.2 (73)	97.64 (95)	98.27 (97)
	Bounces	90.43 (57)	94.56 (76)	85.77 (0)	94.6 (76)	96.87 (86)	97.27 (87)
	Accidental	84.77 (60)	86.7 (64)	70.56 (0)	86.4 (63)	97.33 (94)	98.11 (96)
	Acceleration	73.77 (48)	87.95 (75)	66.33 (25)	92.1 (84)	97.78 (95)	98.13 (96)
	Tremor	80.68 (61)	88.28 (75)	66.86 (26)	92.5 (85)	97.53 (95)	98.32 (97)

Table 7.4: SIID insertion classifiers' mean accuracy and kappa statistics. For each line, different background colors illustrate significant differences on accuracy between machine learning techniques.

Accidental touches classifier. The fifth line of Table 7.3 presents the results of accidental touches classifier using the HIID dataset. A repeated measures ANOVA revealed significant differences [$F_{2.573,254.754}$=2851.197, p<.001] between techniques, where IB1 was significantly more accurate than all remaining techniques. For the SIID dataset, accuracy results also showed significant differences [$F_{2.134,211.235}$=19801.749, p<.001]. Again, post-hoc tests revealed that IB1 outperformed their counterpart techniques.

Acceleration classifier. Regarding HIID training data, there was a significant effect of technique on accuracy [$F_{2.996,296.621}$=7172.682, p<.001], with IB1 outperforming the remaining machine learning techniques. Following a similar trend, SIID data showed statistically significant differences between techniques for the acceleration classifier [$F_{3.225,319.315}$=19822.745, p<.001], namely between IB1 and the remaining conditions (fifth line of Table 7.4).

Tremor pattern classifier. Using HIID dataset, results revealed a significant effect of technique on accuracy [$F_{2.760,273.225}$=6822.801, p<.001]. As show in the sixth line of Table 7.3, post-hoc tests showed significant differences between both C4.5 Tree and IB1, and the remaining techniques. No significant differences were found between C4.5 and IB1 techniques. For the SIID dataset, we also found significant differences between techniques [$F_{3.617,358.080}$=20825.283, p<.001]. Post-hoc tests revealed that IB1 is significantly more accurate than all remaining machine learning techniques.

Major results. IB1 (an instance-based learning approach) was consistently the most accurate technique for all classifiers and for both user groups. Indeed, all techniques showed similar accuracy trends within each dataset (i.e. HIID and SIID). The C4.5 Tree

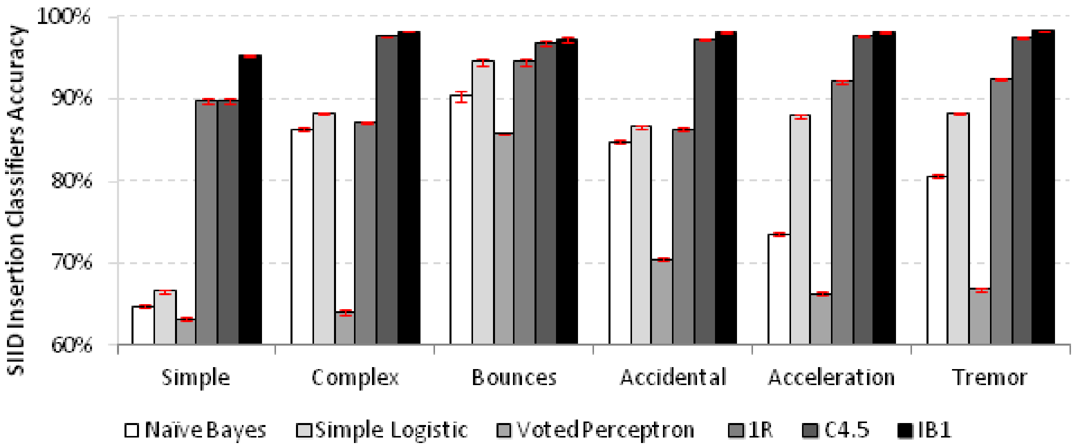


Figure 7.4: SIID insertion classifiers accuracy. Error bars denote 95% confidence interval.

was the second most accurate technique, which can prove to be useful, since instance-based learning approaches can become slow for large datasets or real-time learning (Witten and Frank, 1999). It is also interesting to mention that although the 1-Rule technique is a significantly simpler technique, it consistently outperformed more complex ones, which were representative of statistical and linear modeling approaches.

Comparison between Insertion Models

In this section, we compare the performance of the previously built models. In order to narrow our analysis, we chose the most accurate machine learning technique for each of the five proposed models (Simple Time-based, Complex Time-based, Bounces & Accidental, Acceleration and Displacement, and Tremor Pattern). Although Bounces & Accidental is presented as one model, it represents the combination of two classifiers: Bounces and Accidental Touches classifiers (see Section 7.5.1). Therefore, the reported accuracy was obtained through a combined classifier that uses the Bounces classification algorithm if previous and current letters are the same; otherwise the Accidental Touches classifier is used. Notice that both datasets are mutually exclusive; that is, there are no duplicated instances.

Additionally, we added a Baseline model, which reflects the original keyboard behavior (i.e percentage of taps which are not insertions). This allowed us to assess the effect of the new models, particularly the effect of motion-based models, over standard solutions. The performance of each model was measured for all participants. In order to avoid overfitting and accuracy bias, the participant being tested was excluded from the training set. Herein, we report the results for both user groups: HIID and SIID.

HIID dataset. We obtained a mean accuracy of 56.6% (kappa=11%), 59.64% (kappa=18%), 60% (kappa=17%), 59.24% (kappa=18%), and 59.79% (kappa=19%) for the Simple Time-

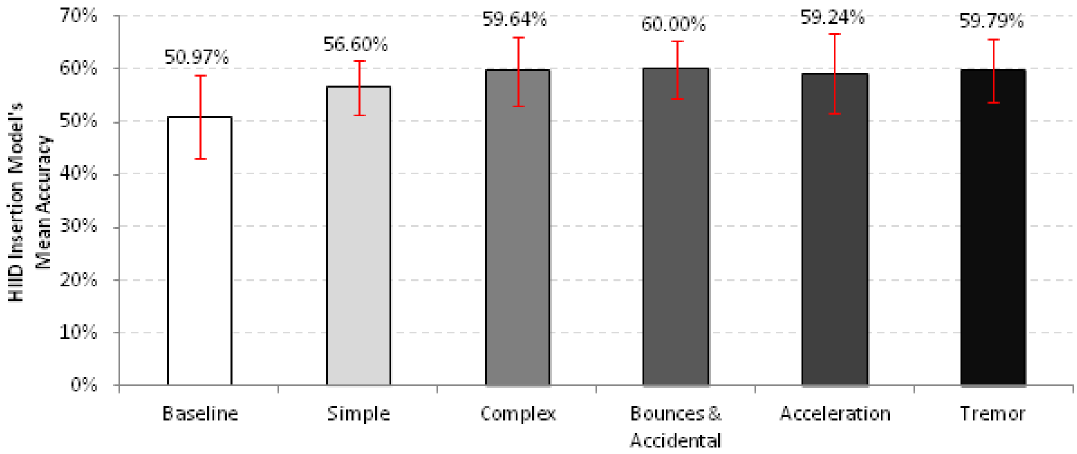


Figure 7.5: HIID insertion models' mean accuracy. Error bars denote 95% confidence interval.

based, Complex Time-based, Bounces & Accidental, Acceleration, and Tremor Pattern models (Figure 7.5). A repeated measures ANOVA revealed a significant effect on accuracy [$F_{5,65}$=3.552, p<.01], with differences between the baseline condition and all new models. No other significant differences were found. Particularly, motion-based models do not seem to enhance the accuracy of insertion errors identification. Both Acceleration and Tremor Pattern models did not significantly outperform touch-based models. In fact, all models' accuracy and kappa statistics are generally low. Although there was a statistically significant improvement of performance from the baseline condition, there was a small practical gain. This result can be easily explained by the users' variance in typing behaviors. Thus, building generic models (i.e. using data from all participants) will result in low classification performance. This issue will be further explored when accuracy of user-dependent models is assessed.

SIID dataset. Regarding the SIID dataset (Figure 7.6), results showed a mean accuracy of 63.35% for the Baseline, 64.42% (kappa=7%) for the Simple Time-based, 83.41% (kappa=56%) for the Complex Time-based, 83.51% (kappa=57%) for the Bounces & Accidental, 83.56% (kappa=51%) for the Acceleration, and 81.83% (kappa=46%) for the Tremor Pattern model. As in the HIID dataset there was a significant effect on accuracy [$\chi^2_{(19)}$=60.581, p<.001], with the Baseline and Simple models being outperformed by all the remaining. Moreover, there were no significant differences between touch-only and motion-enhanced models; that is, although motion and touch data can improve input performance in comparison to the baseline condition, there is not a significant gain of motion-based models by themselves. Also, it is worth noticing that unlike HIID data, kappa values for SIID models are between 41-60%, suggesting a moderate level of agreement.

Major Results. Overall, in both user groups, new proposed classifiers outperformed the Baseline condition. Nevertheless, generic models were shown to perform poorly (low accuracy and agreement) when considering HIID participants, which suggests that user-

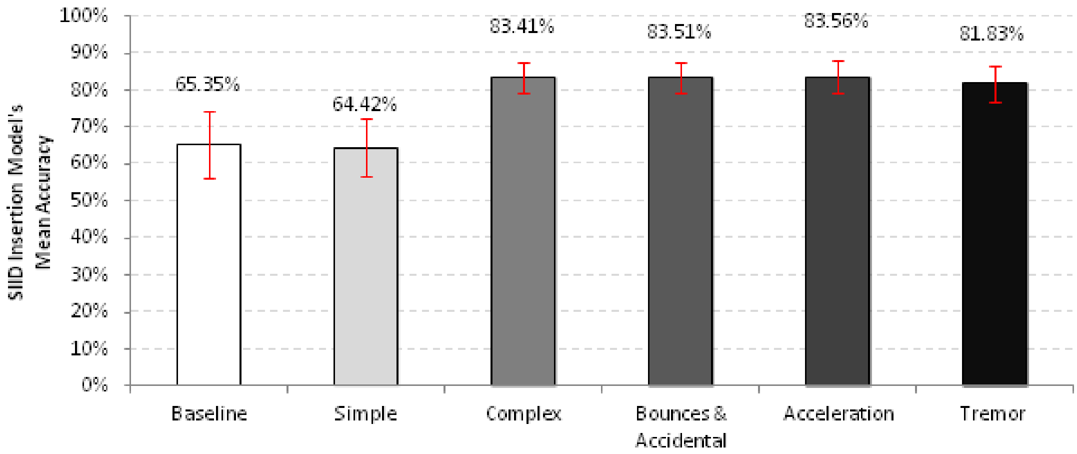


Figure 7.6: SIID insertion models' mean accuracy. Error bars denote 95% confidence interval.

dependent classifiers are needed in order to accurately identify insertion errors. On the other hand, SIID generic models may be satisfactory; however, there is still room for improvement. For instance, Table 7.5 shows the Complex model confusion matrix for participant #1. Notice that the percentage of false negatives is relatively high (20.3%), which means that nearly one fifth of instances are not being correctly classified as insertion errors.

A further major result from this analysis was that the composite model - Bounces & Accidental - did not demonstrate a significant improvement upon classification accuracy. Thus, we recommend using a single model to deal with bounces and accidental touches. Also, motion-based models were not able to enhance insertion errors classification in both user groups. Both Acceleration and Tremor Pattern models did not outperform the remaining models, which only featured touch attributes. Nevertheless, this result was expected since most correlations with motion data (reported in previous chapter) were related to substitution errors.

		Predicted	
		Insertion	Not Insertion
Actual	Insertion	18.5%	20.3%
	Not Insertion	1.5%	59.7%

Table 7.5: Complex Time-based model's confusion matrix for SIID participant #1.

Personalization Effect

In this section, we analyze the effect of personalization; that is, we compare the performance of generic user-independent models (built with data from all participants) against personalized user-dependent models. This will allow us to understand if classifiers have to be trained with data from individual users to achieve the best prediction performance. To mitigate bias caused by the training sample, we used a stratified 10-fold cross validation to compute the models' accuracy and kappa statistics for each participant.

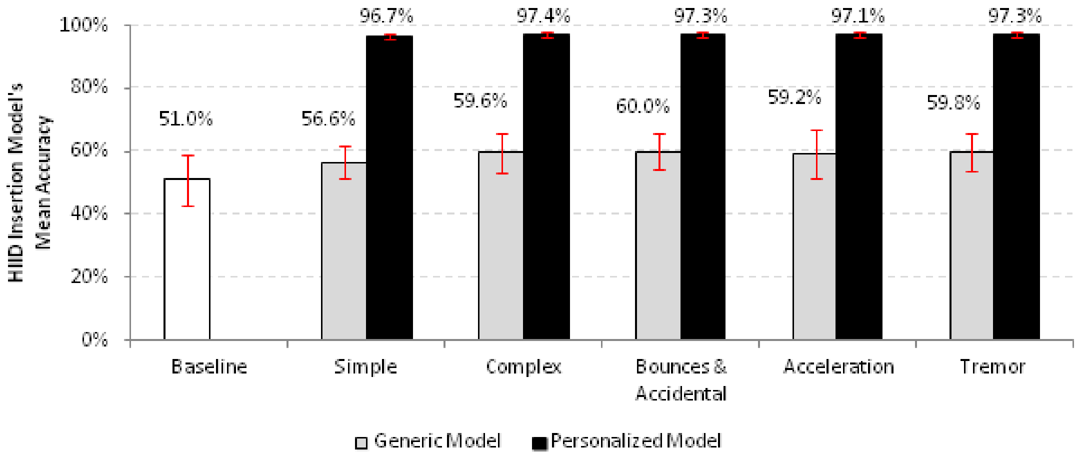


Figure 7.7: HIID insertion models' (generic and personalized) mean accuracy. Error bars denote 95% confidence interval.

HIID dataset. Personalized models achieved a mean accuracy of 96.7% (kappa=92%), 97.4% (kappa=94%), 97.3% (kappa=94%), 97.1% (kappa=94%), and 97.3% (kappa=94%) for the Simple Time-based, Complex Time-based, Bounces & Accidental, Acceleration and Tremor pattern models, respectively (Figure 7.7). A two-way repeated measures ANOVA revealed no significant main effect of model on accuracy [$F_{2.527,32.856}$=.793, p=.535], which means that all new models perform similarly (with exception to the baseline condition). Again, motion-based models did not outperform touch-only models. However, we found a significant main effect of personalization [$F_{1,13}$=203.940, p<.001], showing that indeed older adults greatly benefit from user-dependent classification models. Moreover, no significant Model*Personalization interaction was found, which means that all classifiers behave similarly when built with users' own data.

SIID dataset. Generally, personalized models' accuracy was higher than generic's models (Figure 7.8). We obtained a mean accuracy of 97% (kappa=93%) for Simple Time-based, 98.4% (kappa=96%) for Complex Time-based, 98.3% (kappa=96%) for Bounces & Accidental, 98.4% (kappa=96%) for Acceleration, and 98.4% (kappa=96%) for Tremor pattern model. Results showed a significant main effect of model on accuracy [$F_{4,72}$)=33.485, p<.001], with the Simple Time-based model being significantly worse than all other proposed classifiers. Again, motion-based models did not significantly improve classification performance. As for HIID participants, we also found a significant main effect of personalization [$F_{1,18}$=80.280, p<.001], which shows that situationally impaired users will benefit from user-dependent solutions to prevent insertion errors.

Major Results. For both user groups, personalization showed to play an important role in identifying insertion errors. Indeed, significant differences were found between generic

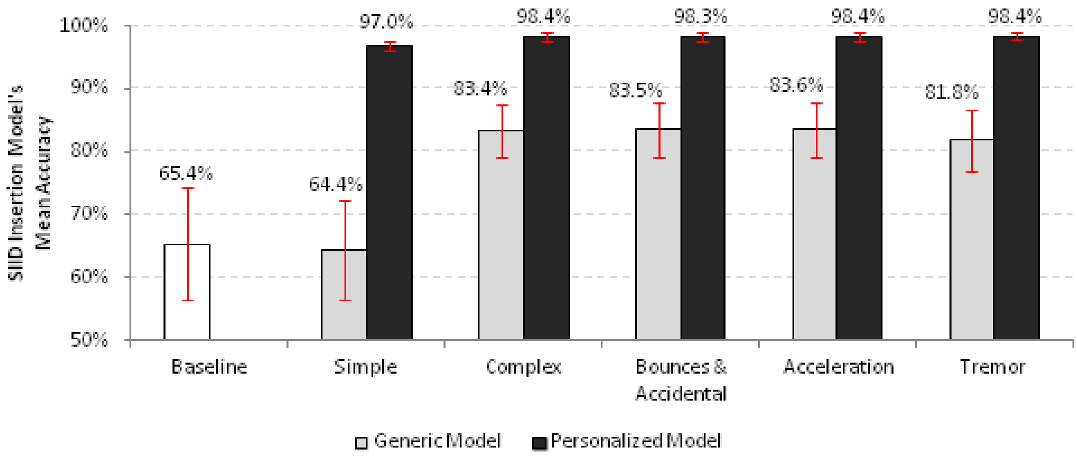


Figure 7.8: SIID insertion models' (generic and personalized) mean accuracy. Error bars denote 95% confidence interval.

and user-dependent models. On the other hand, motion data did not improve classification accuracy when compared to touch-based solutions.

User Group Effect

In the previous chapter we described the differences and similarities in how older adults and situational impaired people input text. In this section, we assess how predictive models built to identify insertion errors can be transfered between user groups. Overall, for each participant, we performed a recognition test using models built with data from three domains: 1) all other participants from the same user group, 2) users' own data, and 3) data from counterpart user group.

HIID dataset. HIID participants obtain a mean accuracy of 51% (kappa=10%), 64% (kappa=24%), 64% (kappa=24%), 62% (kappa=18%), and 66% (kappa=28%) for Simple Time-based, Complex time-based, Bounces & Accidental, Acceleration, and Tremor pattern models, respectively (Figure 7.9). A two-way repeated measures ANOVA showed a significant effect of target population on accuracy [$F_{2,26}$=182.399, p<.001], illustrating the differences between Personalized and both Generic and SIID data. In general, both accuracy and agreement is very similar between models built with Generic and SIID data. In fact, SIID trained models outperformed Generic models for most classifiers. However, no significant differences were found between these two conditions, suggesting that SIID models can be applied to HIID users without a significant loss of performance. Results show that most errors are false negatives, which corresponds to insertion errors misclassified as non-insertions. We believe this error type to be preferred to false positives, which would increase users' frustration as correct taps would be classified as insertion errors. Still, these models performed poorly when compared to personalized solutions.

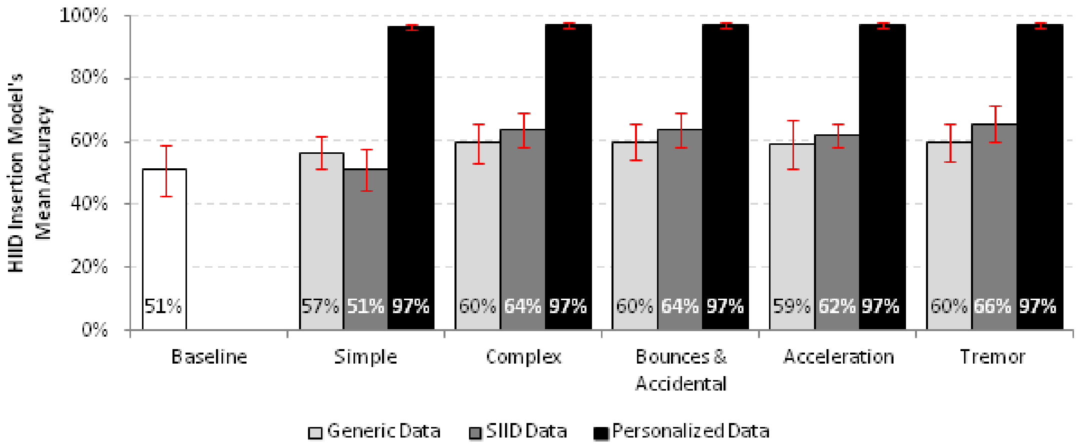


Figure 7.9: HIID insertion models' (generic, siid population, and personalized) mean accuracy. Error bars denote 95% confidence interval.

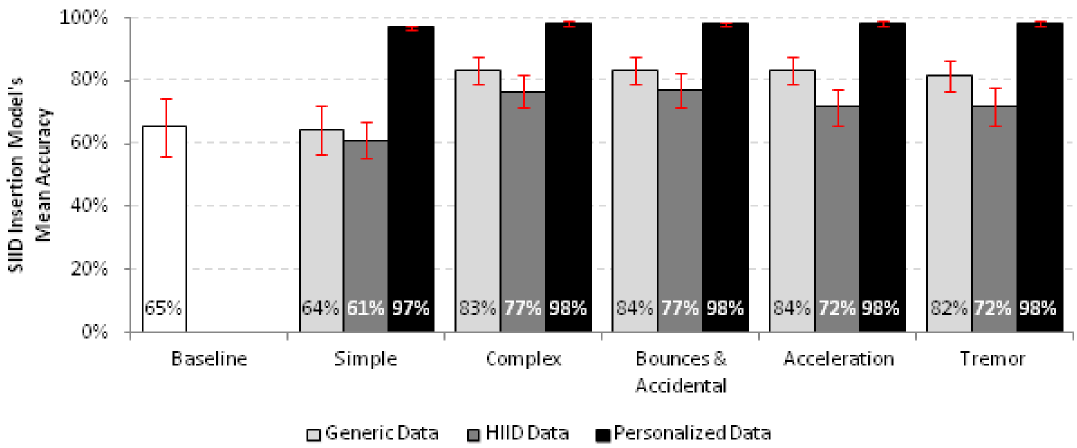


Figure 7.10: SIID insertion models' (generic, hiid population, and personalized) mean accuracy. Error bars denote 95% confidence interval.

SIID dataset. Figure 7.10 illustrates the models' accuracy with the three types of data. A repeated measures analysis revealed a significant main effect of target population on accuracy [$F_{1.237,22.266}$=99.163, p<.001]. Pair-wise comparison showed significant differences between all conditions. Models built with HIID data were significantly less accurate than generic (on average less 5.6% accuracy) and personalized models (on average less 24% accuracy). Moreover, applying HIID models to SIID data proved to be inefficient regarding agreement, as kappa statistic values were generally poor (<35%).

Major Results. If one desires to transfer solutions between user groups, it should be done from SIID to HIID; that is, classifiers can be trained with SIID data and then applied to HIID users, as no significant differences were found to models built with HIID data. However, we recommend these solutions to be avoided due to the low accuracy and agreement values. On the other hand, results show that models should be data-driven, particularly they should be user-dependent and reflect each user's typing patterns; that is, to achieve high accuracy, models should be trained with users' own data.

7.5.3. Major Results

In this subsection, we summarize the major results of insertion errors' predictive models.

Use instance-based learning. An instance-based learning approach (IB1) was the most accurate technique for both user groups. Nevertheless, the C4.5 Tree was consistently the second most accurate technique. In the presence of large datasets or real-time learning, this can be a good alternative, since instance-based learning can become significantly slow (Witten and Frank, 1999).

Touch-based models accurately predict insertion errors. Results showed that to predict insertion errors touch features are enough. Motion-enhanced models did not show an increase on accuracy when compared to Touch-only Time-based models. Overall, the Complex Time-based model achieved the best results for both user groups, which suggest that insertion errors are characterized by similar features for HIID and SIID participants.

Transferring models is a poor solution. Although the same classifier/features can be used for both user groups, results show that transferring the models is generally a bad solution. Doing so, results in low accuracy and agreement values.

The need for personalization on insertion errors. User-dependent models should be applied in order to deal with insertion errors. Both user groups benefit from a data-driven approach where models are built with users' own data. These models reflect their typing patterns and allow an effective error correction (or prevention).

7.6. Dealing with Substitution Errors

Previous results showed that substitution errors are the most similar error type between user groups (see Chapter 6). Although their tremor profile varies, several tremor measures, such as hand oscillation and displacement, were strongly correlated to typing performance. Thus, we are expecting that motion information can enhance error prediction.

7.6.1. Design of Substitution Classifiers

Conversely to insertion errors, which consisted in a two-class classification problem, dealing with substitutions is far more complex. In addition to correctly identifying a substitution error, one has to determine the appropriate correction, making it a 27-class problem.

In this subsection, we present three touch-based (Closest Centroid, Simple Touch-based, and Complex Touch-based) and four motion-enhanced (Acceleration & Displacement, Tremor Pattern, Hand Oscillation & Jerk, and Spring-Mass-Damper) classifiers. Moreover, we describe the features included in each classifier and the rationale to incorporate them.

Closest Centroid Classifier

Description: This classifier corresponds to one of the simplest adaptation mechanisms. Previous research using this approach has shown mixed results; Findlater and colleagues (Findlater et al., 2011) found significant improvements on ten-finger typing, while others (Go and Endo, 2007)(Al Faraj et al., 2009) did not find performance benefits. In essence, each key is characterized by a centroid, which is built based on users' previous key presses. Each centroid is calculated by averaging all finger-touches and then key positions are adapted accordingly. Then, in the classification stage, each new key press (x- and y-position) is classified based on the closest centroid (Euclidean distance).

Features (2): Position (X, Y).

Simple Touch-based Classifier

Description: We draw inspiration from Goel et al. (Goel et al., 2012) to build this classifier; it includes a set of touch features that yield the potential to increase substitutions classification accuracy: touch location, traveled distance, and inter-key interval. Traveled distance was included, since we noticed that, most times, touch-down and touch-up positions were not the same. The time elapsed between taps, represents the users' typing speed and their familiarity with QWERTY keyboards.

Features (5): Position (X, Y), Movement distance (X Y), Inter-key interval.

Complex Touch-based Classifier

Description: This classifier maximizes all information that can be derived from touch events; it includes a new set of features in addition to the Simple Touch-based classifier: transcribed letter, keyboard row, keyboard side (left or right), whether it is a border key, finger-down position, movement time, movement speed, and if a finger slip occurred. These new features intend to maximize the classifier's information about current touch position and movement, as well as its relationship with previous key presses. This will allow the model to adapt to key-related nuances, such as keyboard row and side, and improve prediction performance. Also, we included a new set of movement features, such as movement speed, since we notice that slips usually occurred when the users' fingers moved very fast on the screen. To complement, this feature we added a boolean attribute that indicates whether touch locations on down and up events belong to the same key bounds.

Features (15): Simple Touch-based features, Transcribed character, Keyboard row, Keyboard side, Border key, Finger-down position (X, Y), Movement time, Movement speed (X, Y), Slip action.

Acceleration & Displacement Classifier

Description: We believe that one of the reasons for users' to miss intended keys is the movement of the device and its displacement from a stable location relatively to the user. This classifier was built on this premise and includes acceleration and displacement features in all three axes. The acceleration features consist in 10 samples (per axis) collected from the device's inertial sensor. For the displacement attributes, we incorporated both magnitude and direction values. These were calculated by double integrating acceleration data using a cumulative sum approach. A full description on how acceleration and displacement features were calculated is available in Section 7.3.2. Finally, in order to have some information about the magnitude of variance between finger-touches, we also added two additional motion features: Delta acceleration and Delta displacement. These attributes provide additional information how the device behaved while users' transcribed the intended character.

Features (57): Complex Touch-based features, 10 Acceleration samples (Horizontal, Vertical, Normal), Displacement magnitude (Horizontal, Vertical, Normal), Displacement direction (Horizontal, Vertical, Normal), Delta acceleration (Horizontal, Vertical, Normal), Delta displacement (Horizontal, Vertical, Normal).

Tremor Pattern Classifier

Description: This classifier was designed to take advantage of users' hand movement patterns. Thus, in addition to the Complex Touch-based set of features, the Tremor Pattern classifier includes information in the frequency domain: the instantaneous dominant frequency of device's motion allows the model to adapt to different tremor patterns; and the amplitude of the dominant frequency gives us a proxy of tremor intensity (Salarian et al., 2007). These features were calculated applying a FFT (Fast Fourier Transform) on the accelerometer data and finding the frequency with the highest amplitude. Additionally, we added a time-based feature that enables the classifier to identify where in the pattern the tap occurred, given that errors can be more frequent in different phases of movement. As the last feature, we included the instantaneous direction of displacement.

Features (27): Complex Touch-base features, Dominant frequency (Horizontal, Vertical, Normal), Amplitude of dominant frequency (Horizontal, Vertical, Normal), Elapsed time since last mean crossing (Horizontal, Vertical, Normal), Displacement direction (Horizontal, Vertical, Normal).

Hand Oscillation Classifier

Description: This classifier enhances the Complex Touch-based classifier by analyzing motion variance during typing tasks. To achieve this, we included two main features. First, hand oscillations are represented by the standard deviations of accelerometer data during key presses (Bergstrom-Lehtovirta et al., 2011), thus the higher the standard deviations the higher hand oscillations were experienced by the user. Second, we added Jerk, which consists of the rate of change of accelerometer signal; in other words, it is the first derivative of acceleration with respect to time. Delta jerk was also included in order to measure the difference between the jerk experienced in the beginning and end of tapping action. Adding to these features, we also included information about the device's movement (displacement) direction.

Features (25): Complex Touch-based features, Displacement direction (Horizontal, Vertical, Normal), Hand oscillation (Horizontal, Vertical, Normal), Jerk (Horizontal, Vertical, Normal), Delta jerk.

Spring-Mass-Pattern Classifier

Description: This classifier draws inspiration from a physics model that aims to dynamically compensate for frequent and often random movement. The keyboard is modeled as a mass suspended in each direction by a critically-dampened spring and damper (see Section 7.3.2 for a full description). It counters the overall movement of the device by allowing the screen content to be shifted in the opposite direction. Thus, this classifier incorporates three features that represent the screen's displacement accordingly to the latest accelerometer readings and the impulse response of the Spring-Mass-Damper system. Although we do not directly shift the on-screen keyboard, we allow the classifier to learn when most substitution occur based on the displacement features.

Features (18): Complex Touch-based features, Spring-Mass-Damper (Horizontal, Vertical, Normal).

7.6.2. Evaluation of Substitution Models

The evaluation presented in this section is three-fold: comparison between machine learning techniques, comparison between models, and personalization effect. Conversely to insertion error models, we did not assess the effect of transferring models to deal with substitution errors. Since target sizes differed between HIID and SIID data this would not be a meaningful analysis. Nevertheless, we compare different machine learning techniques, the performance of multiple models, and assess the effect of personalization.

Comparison between Machine Learning Techniques

In this section, we assess the effect of different machine learning techniques on the previously proposed (substitution) classifiers. As with insertion errors, we adopted a "simplicity-first" methodology where simple techniques were tested first. However, we chose to discard Linear Models, particularly Simple Logistic and Voted Percepton techniques, due to the low accuracy rates obtained in preliminary evaluations and in Section 7.5.2.

In order to compute accuracy and agreement results, we used a 10 times stratified 10-fold cross validation, which involved invoking each learning algorithm 100 times on datasets that are nine-tenths the size of the original. All models were built using Weka's default values.

Accuracy (Kappa) %		1-Rule	IB1	Naïve Bayes	C4.5 Tree
HIID	Simple	81.63 (80)	87.98 (87)	66.73 (63)	91.07 (90)
	Complex	85.58 (84)	86.89 (86)	79.47 (78)	90.6 (90)
	Acc. & Disp.	90.87 (90)	88.32 (87)	94.07 (94)	97.15 (97)
	Tremor	90.87 (90)	89.34 (88)	95.72 (95)	97.3 (97)
	Hand Osc. & Jerk	90.87 (90)	90.77 (90)	96.75 (97)	97.14 (97)
	Spring-Mass-Damper	90.87 (90)	88.32 (87)	94.57 (94)	97.26 (97)

Table 7.6: HIID substitution classifiers' mean accuracy and kappa statistics. For each line, different background colors illustrate significant differences on accuracy between machine learning techniques.

In the remainder of this section we present results within each classifier for both groups. Figures 7.11 and 7.12 illustrate classifiers' mean accuracies for HIID and SIID datasets, respectively.

Simple touch-based classifier. Regarding HIID data, results for the simple touch-based classifier are presented in Table 7.6. A repeated measures ANOVA revealed significant differences on accuracy [$F_{2.346,232.299}$=2517.287, p<.001]. Post-hoc tests using Bonferroni correction showed significant differences between C4.5 Tree and all remaining machine learning techniques. Concerning SIID data, as seen in Table 7.7, results showed statistically significant differences [$F_{2.083,206.193}$=6234.684, **p**<.001], namely between C4.5 and the remaining conditions.

Complex touch-based classifier. As with the previous classifier, there were significant differences [$F_{2.669,264.224}$=560.988, p<.001] between conditions; C4.5 Tree showed to be the most accurate technique. For the SIID dataset, results followed a similar trend (see second line of Table 7.7) with significant differences between techniques [$F_{2.749,272.133}$=26930.426,

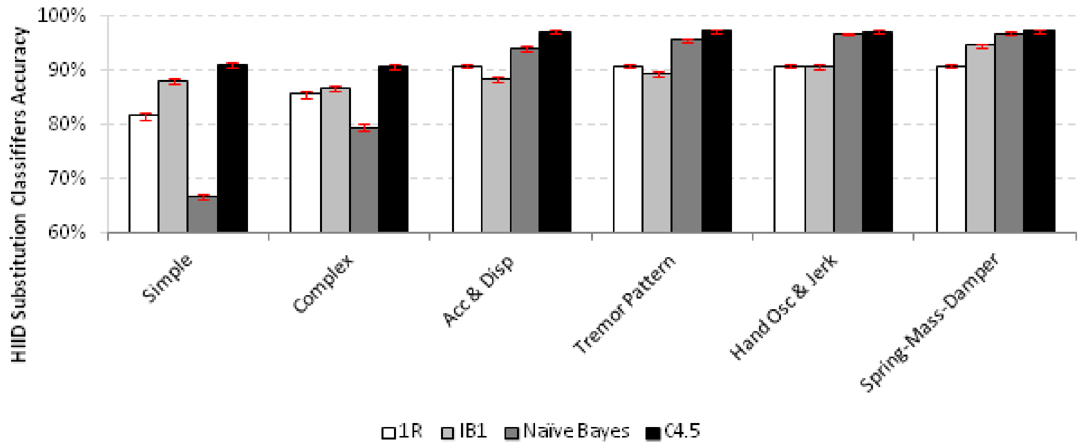


Figure 7.11: HIID substitution classifiers' mean accuracy. Error bars denote 95% confidence interval.

$p<.001$]. Post-hoc tests revealed that C4.5 Tree was significantly more accurate than all remaining machine learning techniques.

Acceleration & Displacement classifier. Concerning the HIID dataset, we found a significant effect of learning techniques on accuracy [$F_{2.638,261.125}=460.849$, $p<.001$], with the Tree classifier being significantly more accurate than all others. As seen in Table 7.7, a repeated measures ANOVA revealed significant differences between conditions [$F_{3,297}=1956.645$, $p<.001$] on SIID dataset. Post-hoc analysis showed that C4.5 classifier outperformed the alternatives.

Accuracy (Kappa) %		1-Rule	IB1	Naïve Bayes	C4.5 Tree
SIID	Simple	73.65 (71)	91.57 (91)	83.65 (82)	93.29 (93)
	Complex	88.08 (87)	91.21 (90)	86.66 (85)	93.45 (93)
	Acc. & Disp.	89.76 (89)	87.13 (86)	92.41 (92)	94.83 (94)
	Tremor	89.76 (89)	89.86 (89)	92.66 (92)	95.17 (95)
	Hand Osc. & Jerk	89.76 (89)	88.27 (87)	93.78 (93)	94.9 (94)
	Spring-Mass-Damper	89.76 (89)	93.13 (93)	93.92 (93)	95.07 (95)

Table 7.7: SIID substitution classifiers' mean accuracy and kappa statistics. For each line, different background colors illustrate significant differences on accuracy between machine learning techniques.

Tremor Pattern classifier. Using HIID data, we found a significant effect on accuracy [$F_{2.529,250.378}=498.442$, $p<.001$], for this motion-based classifier. Particularly, differences were found between C4.5 Tree and the remaining learning techniques (see fourth line of Table 7.6). Similar results were found for the SIID data as statistical analysis revealed a significant effect on accuracy [$F_{2.498,247.330}=1066.078$, $p<.001$] with significant differences between C4.5 Tree and remaining classifiers.

Hand Oscillation & Jerk classifier. For this motion-enhanced classifier, particularly when using the HIID dataset, the C4.5 machine learning technique outperformed all other

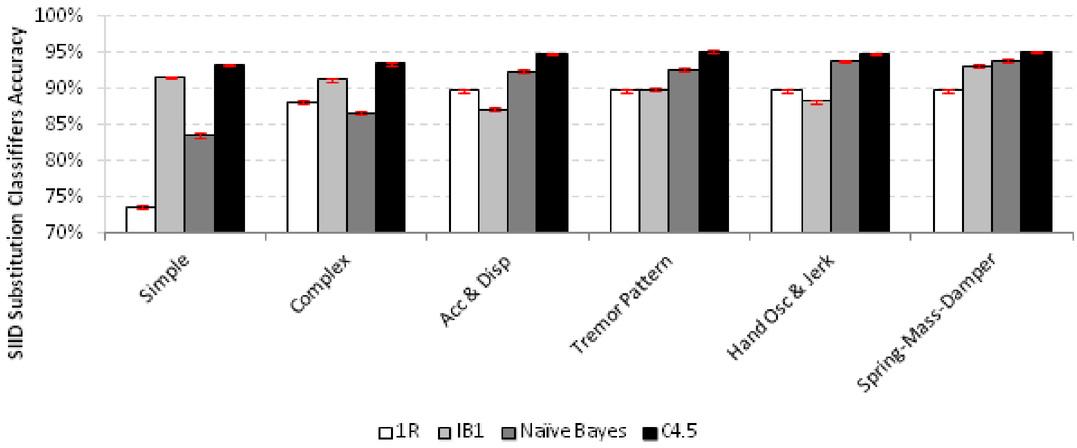


Figure 7.12: SIID substitution classifiers' mean accuracy. Error bars denote 95% confidence interval.

conditions [$F_{2.544,251.824}$=510.110, p<.001]. SIID results followed a similar trend. We found a significant effect of technique on accuracy [$F_{2.549,252.345}$=1829.114, **p**<.001] with the Tree classifier showing higher accuracy than its counterparts (Table 7.7).

Spring-Mass-Damper classifier. Regarding HIID data, results showed significant differences between techniques [$F_{2.443,241.845}$=420.968, **p**<.001]. Again, Post-hoc tests revealed that C4.5 was significantly more accurate than all remaining machine learning techniques. Concerning SIID data, once more, significant differences were found between techniques [$F_{2.743,271.594}$=1136.417, p<.001], with the C4.5 Tree outperforming all its counterparts.

Major results. Conversely to insertions errors results, the instance-based approach (IB1) did not show to be a viable solution for substitution errors. This classifier was not able to deal with the higher number of classes (one per character) of substitution instances. On the other hand, the C4.5 Tree classifier proved to be a good solution by consistently outperforming all the remaining techniques in both user groups. Thus, this classifier will be used in the following analysis.

Comparison of Substitution Models

In this section, we compare the performance of the previously proposed classifiers. Accuracy and kappa statistics values (i.e. agreement) are reported for each model. These measures were calculated by averaging all participants' performance with each model. In order to avoid overfitting, the participant being tested was always excluded from the training set.

In addition to the novel models, we also include a Baseline model, which reflects the participants' performance with the traditional keyboard; that is, it takes as input the

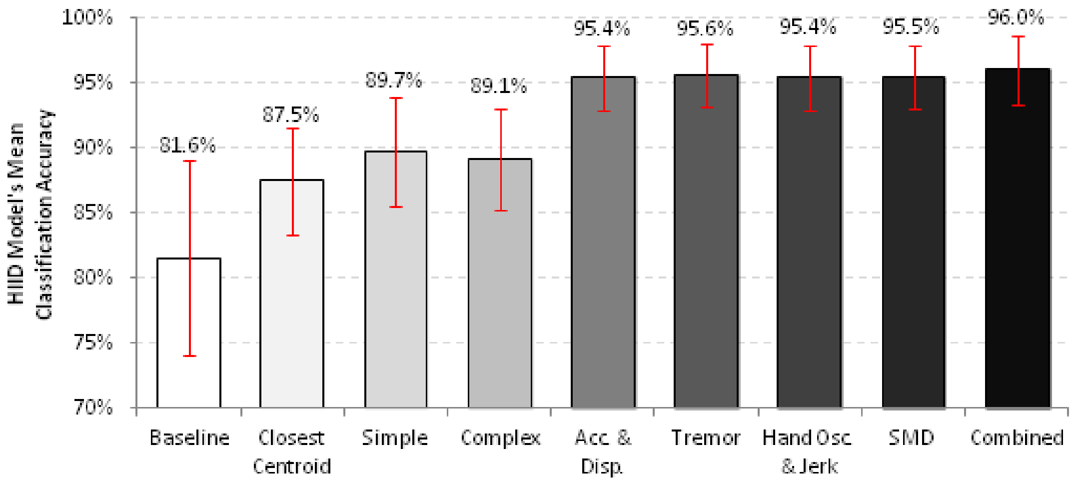


Figure 7.13: HIID substitution models' mean accuracy. Error bars denote 95% confidence interval.

finger-touch position (x, y) and returns the letter whose corresponding key's bounds contains those coordinates. Moreover, we included a combined model that is a composite of one touch-based model (Closest Centroid) and the four motion-based models (Acceleration & Displacement, Tremor Pattern, Hand Oscillation & Jerk, SMD). A majority voting approach is used to classify each key press. When a tie occurs, the most accurate sub-model prevails: for HIID data it corresponds to the Tremor Pattern model, while for SIID data is the SMD model. This decision was based on individual evaluations of each model.

HIID dataset. We obtained a mean accuracy of 81.56% (kappa=80%), 87.47% (kappa=86%), 89.71% (kappa=89%), 89.13% (kappa=18%), 95.40% (kappa=95%), 95.62% (kappa=95%), 95.42% (kappa=95%), 95.45% (kappa=95%), and 96.03% (kappa=96%) for the Baseline, Closest Centroid, Simple Touch-based, Complex Touch-based, Acceleration & Displacement, Tremor Pattern, Hand Oscillation & Jerk, SMD, and Combined models (Figure 7.13). A repeated measures analysis revealed a significant effect of model on accuracy [$\chi^2_{(8)}$=73.655, p<.001]. Post-hoc tests with a Bonferroni correction showed significant differences between Simple Touch-based model and both Baseline and Closest Centroid. This simple model already achieves a high level of accuracy, as well as agreement. Even so, motion data as shown to enhance classification accuracy for about 7%. In fact, significant differences were found between Acceleration & Displacement and Simple Touch-based models [Z=-2.919, p<.005]. However, no statistically significant differences were found between motion-enhaced models.

SIID dataset. Regarding the SIID dataset, we obtained a mean accuracy of 86.34% (kappa=85%), 88.26% (kappa=87%), 92.62% (kappa=92%), 92.88% (kappa=92%), 93.60% (kappa=93%), 92.9% (kappa=92%), 94.9% (kappa=94%), 94.63% (kappa=94%), and 94.92% (kappa=94%) for the Baseline, Closest Centroid, Simple Touch-based, Complex Touch-

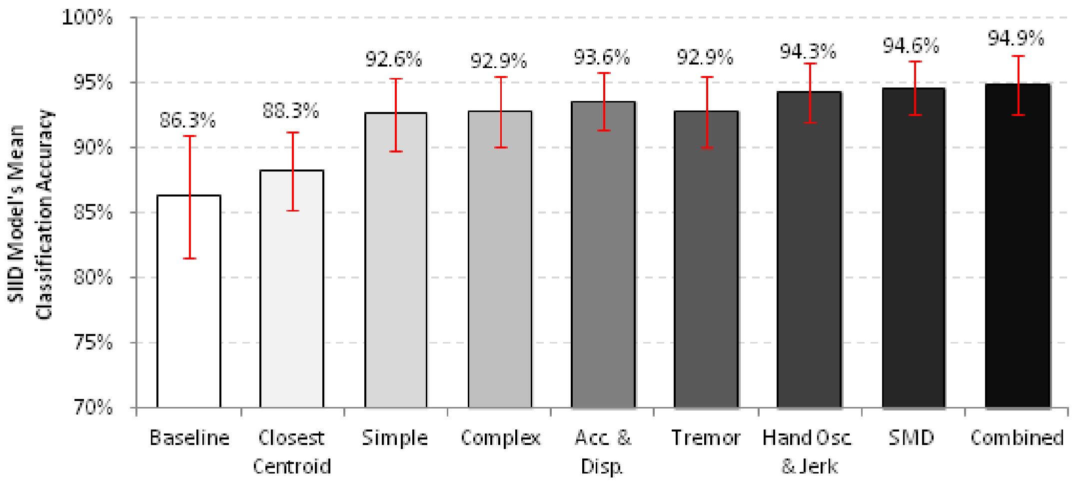


Figure 7.14: SIID substitution models' mean accuracy. Error bars denote 95% confidence interval.

based, Acceleration & Displacement, Tremor Pattern, Hand Oscillation & Jerk, SMD, and Combined models (Figure 7.14). As in the HIID dataset there was a significant effect on accuracy [$\chi^2_{(8)}$=91.918, p<.001], with the Baseline and Closest Centroid models being outperformed by all the remaining. More interestingly, there were significant differences between touch-only and motion-enhced models; for instance, the SMD model had a significantly higher accuracy than the Complex Touch-based model [Z=-3.071, p<.005]. However, no statistically significant differences between motion-based models were found.

Major results. For both user groups, the Baseline condition was outperformed by touch-based classifiers, either Simple or Complex Touch-based models. In fact, these models already presented a significantly higher accuracy and agreement. Still, motion data was able to leverage classification accuracy for both groups. While Tremor Pattern features seemed to be more effective for HIID conditions, SMD model achieved the highest accuracy for SIID data. Also, we found that combining multiple classifiers in a Combined model did not improve upon individual results.

Personalization Effect

In this section, we analyze the effect of user-dependent models; that is, we compare the performance of models built with data from all participants against personalized models. This analysis will enable us to understand if solutions have to be trained with each users' own data to achieve the highest accuracy scores. We used a stratified 10-fold cross validation to compute each model's accuracy and kappa statistics.

HIID dataset. As seen in Figure 7.15, personalized models achieved a mean accuracy of 83.29% (kappa=82%) for Closest Centroid, 80.28% (kappa=78%) for Simple Touch-based, 85.18% (kappa=84%) for Complex Touch-based, 90.64% (kappa=90%) for Acceleration &

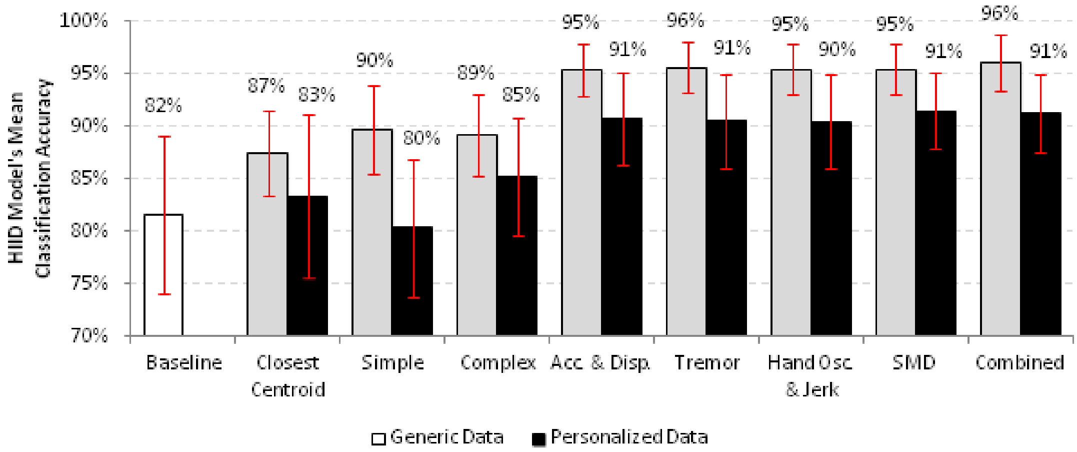


Figure 7.15: HIID substitution models' (generic and personalized) mean accuracy. Error bars denote 95% confidence interval.

Displacement, 90.50% (kappa=90%) for Tremor Pattern, 90.43% (kappa=90%) for Hand Oscillation & Jerk, 91.45% (kappa=91%) for SMD, and 91.27% (kappa=90%) for Combined model. A two-way repeated measures ANOVA revealed a significant main effect of model [$F_{1.452,18.875}$=16.495, p<.001] (described in previous section) and personalization [$F_{1,13}$=9.429, p<.01] on accuracy. Moreover, no significant Model*Personalization interaction was found, which means that all classifiers behave similarly when built with users' own data. Surprisingly, user-dependent models showed a decrease of performance for both accuracy and agreement in comparison to generic models. This decrease may be due to two main factors: 1) users' variance in typing patterns and 2) few samples per class.

SIID dataset. Generally, personalized models' accuracy was very similar to generics' models (Figure 7.16). Results showed a mean accuracy of 93.93% (kappa=93%) for Closest Centroid, 92.57% (kappa=92%) for Simple Touch-based, 92.39% (kappa=96%) for Complex Touch-based, 93.94% (kappa=93%) for Acceleration & Displacement, 94.15% (kappa=94%) for Tremor Pattern, 94.06% (kappa=94%) for Hand Oscillation & Jerk, 94.11% (kappa=94%) for SMD, and 94.32% (kappa=94%) for Combined model. Results showed no significant main effect of personalization on accuracy [$F_{1.,18}$=1.216, p=1.216], which suggests that in general there is no gain in building user-dependent models to deal with substitution errors. Nevertheless, we found a significant interaction between Model*Personalization [$F_{2.763,49.734}$=21.164, p<.001], showing that conversely to all remaining models, the Closest Centroid highly benefits from personalization. This suggests that situational impaired users are consistent within their own finger-touch locations. In fact, the Closest Centroid personalized model achieved similar accuracy rates to Combined [t(18)=-.838, p=.413] and SMD [t(18)=-1.140, p=.269] generic models.

Major results. In general, personalization did not improve models' accuracy for both user groups. Indeed, results show that HIID participants would suffer a significant decrease of typing accuracy. Nevertheless this result should be analyzed with caution since our

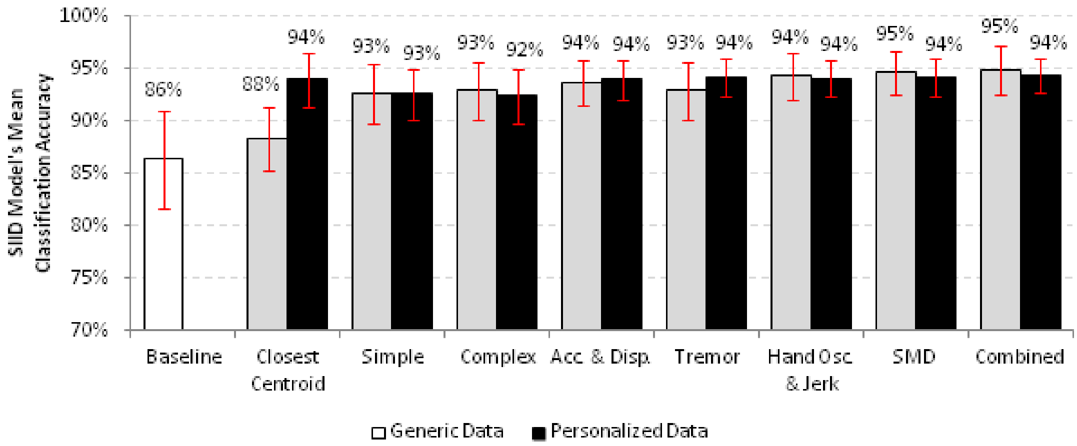


Figure 7.16: SIID substitution models' (generic and personalized) mean accuracy. Error bars denote 95% confidence interval.

dataset was relatively small taking into account the number of possible classes (total of 1405 instances, average of 100 instances per user, and 4 instances per characters). Regarding SIID data, there was no gain in building personalized models, which means that solutions can be trained with data from several users. In theory, manufactures can build and deploy those models without worrying about individual typing patterns. Nevertheless, exception has to be made for the Closest Centroid model; results show that this model benefits greatly from user-dependent data. In fact, it reaches the same level of accuracy as generic motion-based models.

7.6.3. Major Results

This section summarizes the major results of our analysis on substitution errors.

Tree classifier offered the highest accuracy. Although several machine learning techniques were used, the C4.5 Tree classifier consistently outperformed all the remaining. This result remained true for both target populations. Therefore, this was the technique used to build all newly proposed models.

Motion data leverages models' performance for both user groups. Although touch-based solutions already presented significantly higher accuracy and agreement levels than the Baseline condition, motion features were able to leverage their performance. Both HIID and SIID participants benefited from motion-enhanced models as more substitution errors were correctly compensated.

Personalization is not needed when dealing with substitution errors. Overall, user-dependent models did not improve over their generic counterparts. In fact, for HIID participants there was a decrease of performance: an effect previously seen in the literature when considering gesture recognizers for blind people (Kane et al., 2011). The only exception to this trend happened when applying the Closest Centroid model on SIID data. This suggests that manufactures can build their own generic-generated models and deploy them without a decrease in accuracy when compared to user-dependent solutions. However, this solution is restricted to users with SIID and does not address the needs of older adults.

7.7. Discussion

After our analysis of the proposed models for insertion and substitution errors, we are now able to answer the research questions proposed at the beginning of this chapter.

1. *What machine learning techniques are the most effective in predicting errors?*

   Regarding insertion errors predictive models, the Instance-based approach (IB1 algorithm) presented the highest accuracy across all models. Yet, the Tree classifier has shown to be a good alternative, particularly if in presence of large datasets or if efficient real-time processing is needed (Witten and Frank, 1999). Concerning substitution data, the C4.5 Tree classifier was consistently the most accurate technique outperforming Instance-based, Rules, and Statistical Modeling approaches.

2. *Do motion features enhance text-entry accuracy?*

   We obtained different results for insertion and substitution datasets. Regarding insertion errors, the proposed motion-enhanced models were not able to significantly improve upon touch-only models. Nonetheless, this result was expected since most correlations with motion features were with substitution errors (see Chapter 6). Indeed, motion-enhanced models were able to identify and deal with significantly more substitution errors. Both HIID and SIID user groups benefited from acceleration-related features, which allowed models to perform an effective discretization of errors and predict them.

3. *Does personalization affect the classifiers' accuracy?*

   Results showed that personalization has an important role when using time-based models and dealing with insertion errors. This result suggests that typing patterns vary across users; however, they are consistent within themselves. We observed this effect in both HIID and SIID participants, thus user-dependent models are highly recommended for use when dealing with insertion errors. Regarding substitution errors, results are less clear. Personalization only demonstrated improved accuracy in one user group - SIID - and one specific model - Closest Centroid. All remaining user-dependent models, for both user groups, did not show an improvement over their generic counterparts.

4. *Can we use data from HIID user group to build solutions to SIID users, and vice-versa?*

   Overall, building models with data from one user group and applying it to a different target population has shown to be a poor solution. Although some classifiers can be applied in both user groups, they need to be data-driven; that is, each classifier has to be built with data from that user group, especially if time-related features are being used.

5. *Can we use similar classifiers and motion features for HIID and SIID?*

   In our analysis we did not find significant differences between motion-enhanced models, which means that these classifiers can be expected to have similar performance. This result was true for both HIID and SIID user groups, thus similar classifiers can be applied to both target populations, provided that they are built with data from that group. Nevertheless, some motion features seem to be more appropriate for specific users. For older participants, frequency domain features, such as dominant frequency and amplitude of dominant frequency, provided the highest accuracy. On the other hand, for situational impaired users, the Spring-Mass-Damper model compensated more effectively for substitution errors.

7.8. Unifying Model

In previous sections we have designed, developed, and evaluated input models that are able to deal with insertion and substitution errors. Taking into account the obtained results, we are now in a position where we can build a Unifying Model that can be applied to HIID and SIID users and improve overall typing accuracy. This model comprises a composition of two sub-models that prevent insertion errors and deal with substitution errors.

The combined model takes as input touch and accelerometer events, which are then computed in order to gather the necessary features. Figure 7.17 shows a block diagram detailing our approach. Touch events are processed by an independent module, which retrieves position, movement, and previous key press features. The combined model uses a serial approach; that is, if a key is classified as an insertion error it is automatically discarded; otherwise, it continues in the processing stream and enters the substitution model. Accelerometer events are passed through a low-pass filter to remove noise and the classifier's features are then retrieved. The output corresponds to the intended key.

In order to deal with insertion errors, we use the personalized Complex Time-based classifier, since it showed to be the most accurate solution for both HIID and SIID (see Section 7.5.2). For substitution errors, several classifiers such as Acceleration & Displacement, Hand Oscillation & Jerk, SMD, or even a combination of those could be applied, as they performed similarly accuracy-wise (see Section 7.6.2). In this case, one should take into account the computing load of these solutions. Classifiers' usefulness is directly related to responsiveness time, since solutions must be able to run in real-time. We chose the generic Spring-Mass-Damper classifier due to two main factors: 1) simplicity - since it only requires three motion features that are easily computed, and 2) adequacy to both user groups - this classifier obtained the highest mean accuracy when considering both target populations.

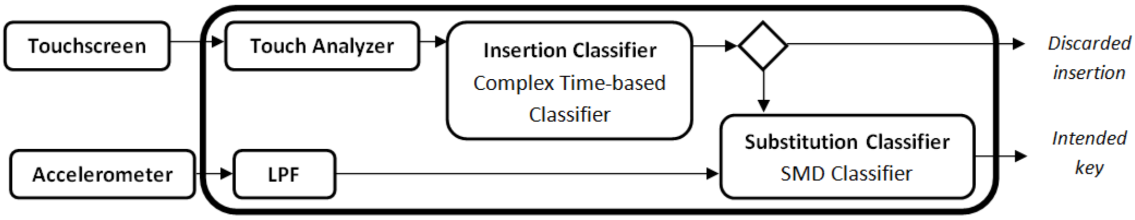


Figure 7.17: Block diagram of major components of Unifying Model.

7.8.1. Improvements on Typing Accuracy

This chapter presented an evaluation of several error-compensation classifiers that enabled us to propose a unifying model. While each of the classifiers that compose this solution were previously evaluated, we were also interested in investigating its effect, as a whole, on users' typing accuracy. In this section we describe the results obtained from a simulation of users' performance using the proposed text-input model. We aimed to answer research questions such as: Does the model improve overall typing accuracy? What is the effect on character-level errors (insertions, substitutions, and omissions)? Where should future work focus in order to improve input accuracy?

To achieve our goal we implemented and trained the Unifying Model and ran a simulation with previously collected typing data. Each of the users' transcribed sentences and key presses were passed through our model and classified accordingly. The resulting sentences were then analyzed and their quality was compared against the originally transcribed sentences.

Regarding the training datasets, and accordingly to the chosen classifiers, the insertions sub-classifier was trained with personalized data; that is, using each user typing dataset. In the evaluation stage, we removed all key presses from the sentence being tested from the training dataset to avoid overfitting. On the other hand, the substitutions classifier was built with generic data (i.e with all key presses). To run the simulation, typing data from the user being tested was not included in training dataset in order to avoid bias.

Typing accuracy was measured using the Minimum String Distance (MSD) Error Rate (MacKenzie and Soukoreff, 2002b), calculated as:

$$MSD(required\,text, transcribed\,text) \div mean\ size\ of\ the\ alignments \times 100$$

This value represents an overall measure of the quality of transcribed sentence relatively to a required sentence. Figures 7.18 and 7.20 illustrates the improvements of input accuracy if the Unifying Model had been used in previous user studies (Chapters 4 and 5).

HIID Typing Data

Older adults obtained an MSD error rate of 25.97% (sd=19.73) on the baseline condition and 19.9% (sd=16.6) with the proposed model (Figure 7.18). Overall, there was an average improvement of 6.08% (sd=2.95%) on MSD error rate per participant. We found a significant effect of *conditions* [Z=-3.008, p<.01], with the Unifying Model outperforming the baseline condition. This shows that the proposed solution can indeed improve typing accuracy significantly. Nevertheless, results show a relatively high standard deviation, illustrating the variability of older adults performance (see Chapter 5).

Figure 7.19 shows the MSD error rate for each participant for both baseline and Unifying Model conditions. Absolute performance improvements ranged from -3.33% (participant #11) to 25.62% (participant #1). Visual inspection of the data suggests that gains are related to consistent typing patterns. For example, considering participant #1, the consistent touch locations for selecting keys allowed the substitution classifier to deal with nearly 25% of errors. On the other hand, participant #11 experienced an increase of 3.3% on MSD error rate, due to incorrectly classified insertion errors, which originated new omissions. Results suggest that consistency in training data is of upmost importance to take advantage of the Unifying Model.

Regarding character-level errors, the Unifying Model was generally effective in dealing with substitution errors, with a drop of 5.3% on error rate from the baseline condition. This difference showed to be statistically significant [Z=-3.174, p<.005]. Similarly, the proposed model was able to prevent on average 1% of insertion errors, resulting in a statistically significant difference [t(13)=3.150, p<.01]. Concerning omission errors, we did not find a significant effect of *conditions* on *error rate* [Z=-.355, p>.7]. This result was expected, since the Unifying Model does not deal with this type of error. In fact, the predominance of omission errors clearly restricted performance improvements; note that omissions are the most common type of error (Figure 7.18).

Examples of required, transcribed (baseline), and simulation results are listed below. Overall, the Unifying Model deals well with substitution errors, replacing wrong characters by adjacent (correct) ones.

```
Required sentence:  muito para alem do necessario
```

```
Transcribed sentence:  muiyp para alrlm  dp mdvrsdaroip
Unifying Model:  muito para alemm  do necessaroio

Required sentence:  lugares de estacionamento mais procurados
Transcribed sentence:  lugared  deedtsciomeomdnto mqis iprocuradoz
Unifying Model:  lugares  deestaciomeomento mais iprocrados

Required sentence:  muito para alem do necessario
Transcribed sentence:  muiyo pqra alem do mnd essztlo
Unifying Model:  muito para alem do nnecessarlo

Required sentence:  de desorientacao e desanimo alastrou
Transcribed sentence:  de xe sorientacao e dsanmou alastrosu
Unifying Model:  de desorientacao e dsanmo alastrou

Required sentence:  algo de muito necessario para
Transcribed sentence:  algo de muitlo mmmd ezzarik paray
Unifying Model:  algo de muitoo nnndcessario parat
```

Regarding insertion errors, while some additional characters are successfully identified, others are not discarded (example #1 and #5). A detail inspection of the data revealed that these incorrect keystrokes did not follow the pattern of insertion errors; that is, their timing features were very similar to correct key presses. Particularly, the sequence of characters "mmm"" in the beginning of the fourth word in the fifth example appeared as three intentional key presses. Thus, the model did not classify them as insertion errors. A possible explanation to this effect may be that the user was repeatedly attempting

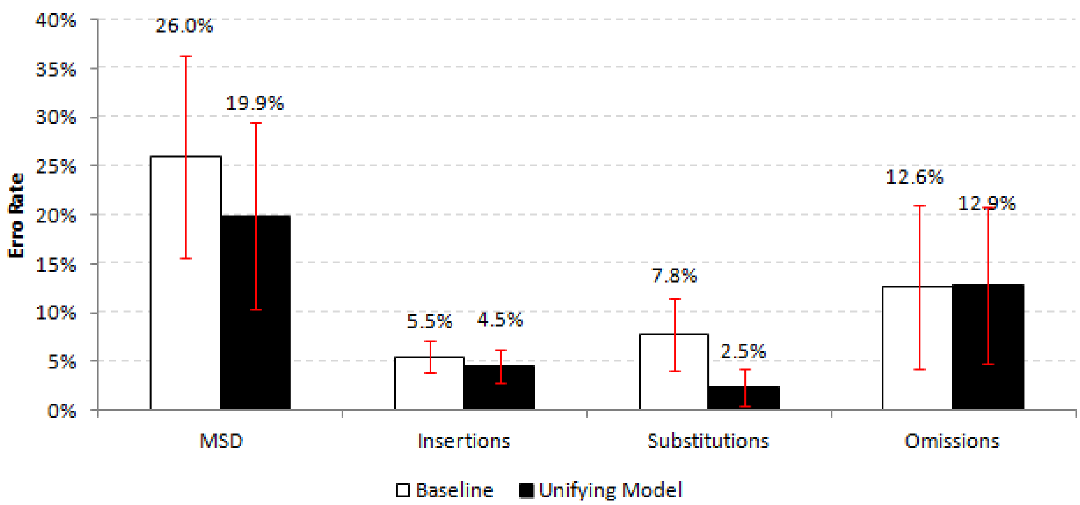


Figure 7.18: MSD, insertion, substitution, and omission error rate obtained through simulation with HIID typing data. Error bars denote 95% confidence interval.

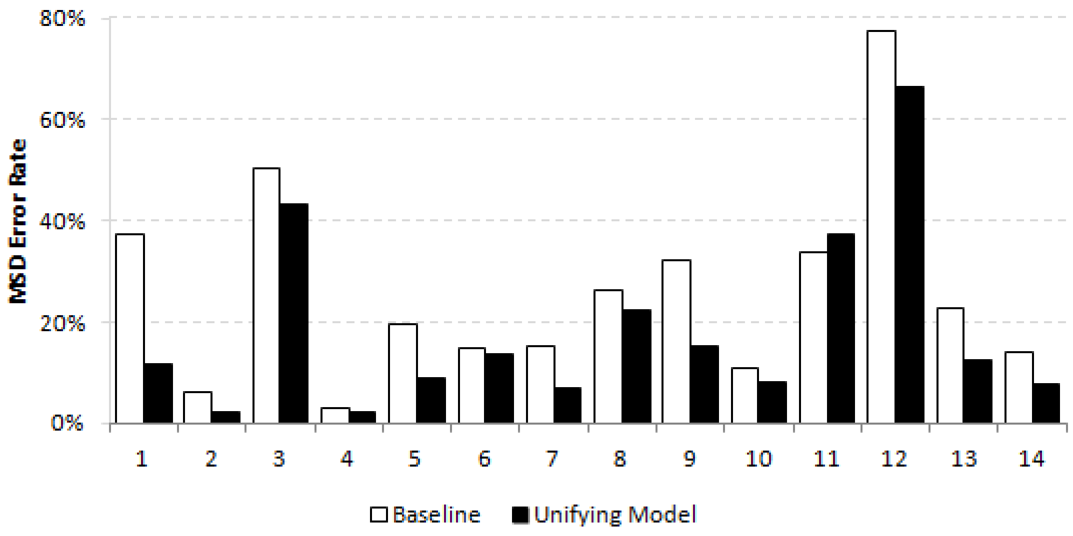


Figure 7.19: MSD error rate across conditions for each HIID participant.

to insert the correct character, originating insertion errors. In this case, if the model has been deployed beforehand, the user's first attempt to insert the letter "n" would have been successful, and further attempts would not be required. Still, we are not able to confirm this without running a complete user evaluation of our solution. This is a limitation of running a simulation.

SIID Typing Data

Regarding situationally impaired participants, results show an MSD error rate of 11.27% and 4.71% for the baseline and Unifying Model conditions, respectively. A Wilcoxon Signed Ranks Test showed a statistically significant improvement on *input accuracy* [Z=-3.725, p<.001]; an average improvement of 6.56% ± 5.9% from the baseline condition (Figure 7.20). Examples (in Portuguese) of these improvements are listed below:

```
Required sentence:  denuncia sobre o alegado separatismo
Transcribed sentence:  drnuncus sobre o alrgafo separatismo
Unifying Model:  denuncia sobre o alegado separatismo

Required sentence:  com anterioridade as nacoes europeias
Transcribed sentence:  com anterioridade as nacoed europrias
Unifying Model: com anterioridade as nacoes europeias

Required sentence:  nao tem aparecido as reunioes
Transcribed sentence: nao teem apateciddo as reunioes
Unifying Model:  nao tem aparecido as reunioes
```

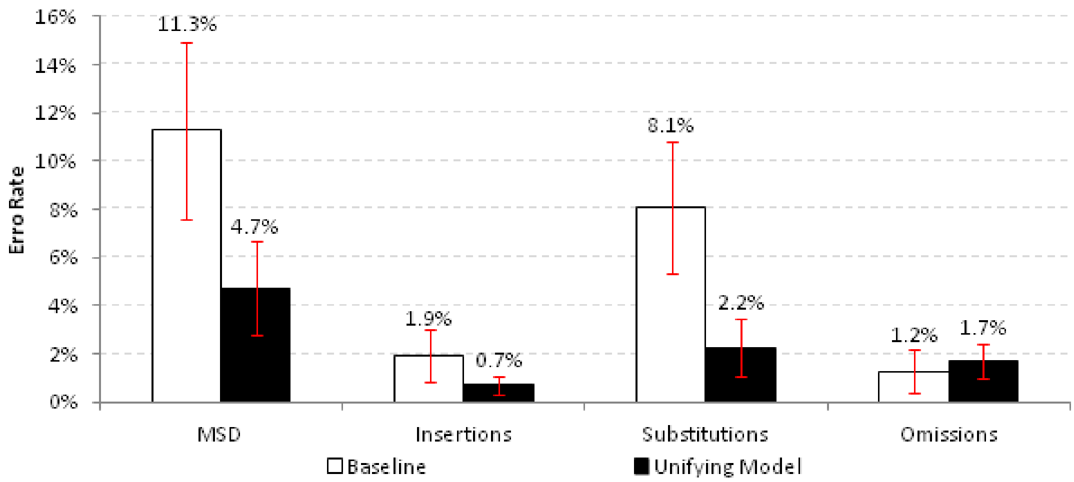


Figure 7.20: MSD, insertion, substitution, and omission error rate obtained through simulation with SIID typing data. Error bars denote 95% confidence interval.

```
Required sentence:  decisoes a tomar na reuniao
Transcribed sentence: decisoes a toomar na reuniao
Unifying Model: decisoes a tomr na reunio
```

Through our simulation, we found a decrease of 5.9% on substitution error rate, which corresponds to a drop of 73%. Statistical analysis revealed a significant difference between *conditions* [Z=-3725, p<.001]. The first and second examples listed above illustrate this effect.

Regarding insertion errors, we also found a drop on error rate from an average of 1.9% (sd=2.44) to 0.7% (sd=.86) [Z=-3.639, p<.001]. The third example show that the Unifying Model is able to deal with several bounce errors as well as substitutions.

Nevertheless, despite the significant increase of accuracy, the insertion classifier also introduced new omission errors. Indeed, the 0.5% increase in omission error rate, due to incorrectly classified key presses, resulted in a statistically significant difference [Z=-2.843, p<.005] (see last example of transcribed sentences). A possible explanation to this effect is the equal weight of insertion errors (yes instance) and correct key presses (no instance). Although we had intentionally given equal importance to both cases in the training stage, the resulting model seems to be "over sensitive" to insertion errors; that is, some key presses that are slightly different from the users' typing pattern are classified as an insertion error, without taking into account small variations in typing style. Still, the overall typing accuracy is significantly enhanced by the proposed model.

In fact, the simulation shows that all users gain from our solution (Figure 7.21). Exception has to be made for participant #17, which did not showed performance improvements. An analysis of this participant data revealed inconsistent typing patterns. On the other hand, participants #1, #3, and #16 experienced a drop in error rate of 78%, 76%, and

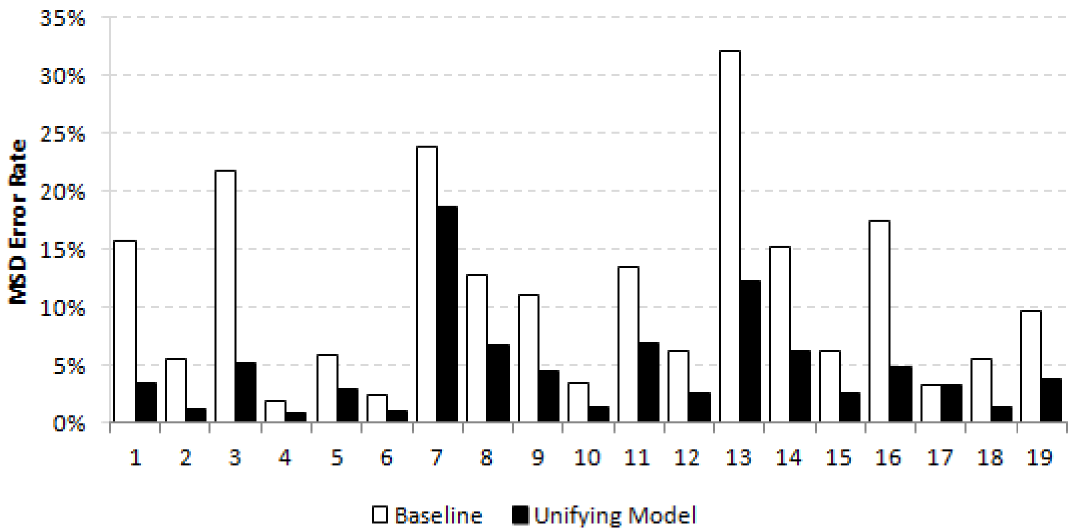


Figure 7.21: MSD error rate across conditions for each SIID participant.

72%, respectively. This was achieved due to the model's high effectiveness in dealing with substitution errors.

7.9. Conclusion

The work presented in this chapter investigated the use of prediction models to compensate insertion and substitution errors of HIID and SIID users. Particularly, we were interested in showing that motion data could be used as a unifying measure to leverage input performance. Moreover, we analyzed the effect of different machine learning techniques, prediction models, and user-dependent solutions. Results showed that, indeed, motion-enhanced models can significantly improve prediction accuracy for both user groups when dealing with substitution errors. Personalization also played an important role when considering insertion errors and classifiers featuring time-based attributes. On the other hand, transferring models between target populations showed to be an ineffective solution. Still, classifiers can be transferred as long as they are data-driven; that is, as long as they are built with data from the user group where they will be used.

At the end of this chapter, we discussed the major results from our analysis and propose a Unifying Model, which is able to adapt to users' abilities and deal with insertion and substitution errors. A simulation of users' performance revealed an overall absolute increase of text quality of 6.1% and 6.6% for HIID and SIID participants, respectively. The solution is particularly effective in dealing with substitution errors, using the motion-enhanced classifier, which resulted in a decrease of 68% - 73% of incorrect key presses.

8 Conclusion

In this chapter, we conclude this dissertation. We start by discussing both the major contributions and results. Next, we identify the benefits of our approach. As with all research projects, ours also have some limitations, which do not invalidate any of the research we carried out; however, are worth some attention as future work. We finish by suggesting future research avenues that can be developed from this dissertation.

8.1. Discussion

The goal of this thesis was to demonstrate that common solutions can be built for both situationally- and health-induced impairments and disabilities. Particularly, we focused on understanding users' abilities and then taking advantage of their similarities in order to deal with their differences in a unifying manner.

To fulfill this thesis goal, we conducted a series of user studies, each examining users' abilities when performing computer-oriented tasks. The first exploratory user study empirically measured the differences and similarities between motor-impaired and able-bodied across three touch-based interaction techniques. Next, we assessed the effect of visual demands and the viability of transferring solutions specifically designed for disabled users and applying them on mobility conditions. The findings from these preliminary studies, in combination with previous work (Chen, 2008), motivated us to run two subsequent laboratory studies to measure users' abilities when in presence of tremor-related impairments, either due to situationally- or health-induced impairments and disabilities. Results enabled us to propose and evaluate novel models that leverage accelerometer readings to improve typing performance. Finally, we demonstrated the benefits of looking at users as a set of abilities, independently of their cause, using motion data as a unifying measure.

In this section, we discuss the major contributions of our research work, as well as the benefits and limitations of this dissertation.

8.1.1. Contributions and Major Results

Beyond the demonstration of our thesis, this dissertation has presented several empirical studies, often analyzing text-entry performance of both situationally- and health-impaired users. We now summarize the major results of each stage of our work.

In the first preliminary user study we seek to **understand the differences and similarities between motor-impaired and able-bodied users** when interacting with a touchscreen mobile device. Results revealed that *traditional selection techniques could be used by both user groups*, opening an opportunity to delve in the development of common solutions. On the other hand, *the difference in the magnitude of errors*, especially in gesture-based approaches, was very significant between motor-impaired and able-bodied participants. This result motivated us to explore mobility conditions.

In our second user study, we focused on the visual demands of interaction whilst “on the go”. Moreover, we wanted to assess the **effect of applying accessibility solutions in mobile settings**. In line with previous research, we observed compensation strategies to deal with mobility, such as reducing walking speed (Barnard et al., 2007)(Bergstrom-Lehtovirta et al., 2011). Also, we learned that *transferring solutions should be executed with care* and a deep understanding of users’ abilities is needed. Still, based on our observations and previous work (Chen, 2008), *mobile users inevitably experience hand tremor* and, consequently, a decrease in target selection accuracy.

This preliminary research enabled us to focus and define our research topic: **bridge the gap between SIID and HIID**. Thus, in the following user studies, we attempted to thoroughly understand tremor-related impairments either due to contextual or health factors.

First, we aimed at **characterize users’ abilities in walking conditions**, analyzing types of errors and their correlation with hand oscillations. Results showed an effect of mobility, demonstrating that *users were indeed situationally-impaired*, with a significant *decrease of typing accuracy.* We found a *predominance of substitution errors*, showing these to be the most sensitive type of error to mobility conditions. Additionally, specific *error patterns* emerged from our data, particularly adjacent substitutions. Analysis of *accelerometer data showed potential to be used to automatically compensate these er-*

rors.

As a second stage of our approach, we **characterized older adults' abilities, which often show increased physiological hand tremor**. We found *similar patterns regarding substitution errors*, however conversely to users with SIID, there was a *high omission error rate*. We believe this to be a combined effect of lack of typing experience, perceptual, and cognitive errors. Additionally, we found a *high degree of variance across users. Hand tremor profile also showed a strong correlation with typing errors.*

Once both user groups' abilities were characterized, we performed a **comparative analysis** of typing behaviors, tremor profile, and relationship between typing performance and tremor measurements. Results showed that both situationally- and health-impaired users experience a *similar (relative) magnitude of insertion errors*. Nevertheless, non-motor abilities seem to play an important role in the performance of older adults. Regarding substitution errors, both populations' performance was interleaved resulting in the *"Disability Continuum" concept*. *Tremor profile was significantly different*: while users with SIID exhibit a consistent tremor pattern, older users show a more irregular and user-dependent pattern. Along with previous findings, *motion measures proved to be correlated with input performance for both user groups*; however, some motion features were more useful for one specific group.

Next, we **proposed and evaluated models of users' abilities**, designed to compensate typing errors. *Regarding insertion errors, we found that models need to be personalized* as timing features (typing speed) are highly user-dependent. On the other hand, concerning substitutions, results showed that *motion-enhanced solutions outperform touch-only models*. We end our analysis by proposing a **Unifying Model** that can be applied to users with either SIID or HIID.

Finally, we ran a simulation to **assess the effect of the Unifying Model on typing performance**, showing that *taking into account users' abilities, either experiencing SIID or HIID, we can develop common solutions*. Results showed a significant overall improvement on typing accuracy with a decrease of 58% and 23.5% error rates for SIID and HIID user groups, respectively. The Unifying Model was especially effective in dealing with substitution errors. The motion-enhanced classifier was able to reduce 68-73% of substitution error rates.

8.1.2. Benefits

Most existing approaches that attempt to deal with a wide range of abilities are limited to one cause of impairment, either due to situational or health conditions. Our work made important strides towards bridging this gap and *designing for both SIID and HIID in a unified manner*. There are several benefits in taking this approach: avoid the duplication of work; reuse knowledge between different research fields, namely between accessibility and mobile HCI domains; leverage research by coping with a wide range of abilities; reduce costs and increase availability; and remove the negative connotation of term "accessible computing", which is usually seen as a research field for minorities. We showed that there is a considerable overlap of problems and common solutions can be devised.

We look at users as a set of abilities, which vary across a limited range. Additionally, our approach is *agnostic regarding the cause of impairment or disability.* Instead of characterizing users' abilities through subjective (WHO, 2009) or clinical measures (Persad et al., 2007)(Oliveira et al., 2011a), which are usually ineffective in informing design, abilities were assessed through performance with the device itself.

This approach overcomes the problems of generalized demographic or clinical data, and enables an *objective and effective adjustment of interfaces in respect to each user.* Particularly, in text-entry performance, we looked at three main types of errors to characterize the users' abilities and issues with the interface: insertion errors - unintentionally added characters; omission errors - omitted characters; and substitution errors - incorrect characters.

Monitoring users typing behaviors allowed us to model and deal with both SIID and HIID users' abilities in the same way. Moreover, we went beyond typing performance and *leveraged the device's built-in sensors* (i.e. accelerometer) to better understand our audience's abilities. Indeed, motion-enhanced models showed significant improvements over traditional solutions, which resorted exclusively to touch information. Motion features acted as an additional source of knowledge that can be used to characterize both target populations.

Finally, our approach has the advantage of being completely *independent of language.* Unlike traditional error correction or prediction systems, our models do not use language-related features. Nevertheless, it can be easily integrated with these types of solutions. As with any keyboard, the result of each key press is a character, which can then be used to feed a prediction algorithm.

8.1.3. Limitations

Although our approach was able to deal with SIID and HIID users' abilities, it also has some limitations that are worth mentioning to be explored as future work.

In order to cope with the variability of users, both within and between each domain, we need to train our models with real data. This can be a tedious process, especially if we consider that thousands of data points are needed to build reliable solutions. This effect is even worse when dealing with numerous contexts and the dynamic diversity of users and mobile settings. Quick and effective assessment methods should be investigated and deployed "in the wild" in order to collect large amounts of data and require minor intervention from the user. For example, correction actions (deletes) and language corpora could be used to automatically classify unintended characters. This would allow users to start using our solution without requiring an explicit training stage. Sensing the surrounding context would still be a determinant role to predict users' performance and dynamically adjust the interface.

Another limitation is a technical one: to leverage motion data we need to use the device's sensor, which consume a significant amount of battery. Depending on the number and duration of text-entry tasks, users can experience a drop on battery life. This limitation can be mitigated by reducing the sampling rate. In fact, taking into account users' typing rate and from our observations, there is no need for sampling rates above 10Hz. Nevertheless, this value can be dynamically adjusted depending on users' typing behaviors.

Finally, our approach is clearly more useful for users that experience hand tremor and a decrease in keying accuracy. In situations where the device has large targets and is positioned in a static surface, our error-compensation mechanisms will probably provide little gain to end-users. Still, mobile devices are usually held in hand and are being increasingly used in mobile contexts (e.g. public transports, walking, during conversations, standing, and so on.), which only strengthens our work.

8.2. Future Work

The findings of this dissertation pave the way for further improvements of current work and new research avenues of future work.

Re-Evaluate Unifying Model. Although we showed that the proposed error-compensation model indeed improves typing accuracy, it would be interesting to re-evaluate it in real typing tasks. This will allow us to assess the effect of our solution on users' natural typing behaviors. Analysis of this re-feedback can provide useful insights for further improvements and adjustments of our classifiers.

Include Language Models. One of the main advantages of our approach consists in being able to identify and deal with typing errors independently of language. By modeling users' performance and taking advantage of acceleration data, we compensate insertion and substitution errors. Nevertheless, language models, particularly predictive language models, have been shown to increase typing speed (Song, 2010). Since our main focus was on improving input accuracy, language models can provide a good extension to current work, resulting in more effective and efficient keyboards for wider-audiences. Additionally, this can be a valid solution to deal with non-motor errors (e.g. forgetting to enter some letters).

Explore Sensor Fusion. Additional sensors may be useful in improving our classification models. Our work only leveraged accelerometer data; however, other sensors, such as the gyroscope would provide relevant data for inferring rotations. Moreover, sensor fusion approaches can be applied to discriminate hand tremor, gravitational acceleration, and tapping actions. We believe that the identification and quantification of these events separately, holds potential to further improve classification accuracy. This is certainly an interesting research avenue.

Provide Graceful Adaptation. Mobile devices are used in increasingly heterogeneous and dynamic contexts. Although proposed models are able to deal with a wide range of abilities and conditions, we did not explore the effect of constant adaptation within different environments. It is expected that users' abilities vary in such contexts and providing dynamic interfaces will be a challenge. Still, tremor measures can be used as predictors of the challenges users will face in those dynamic environments, even before touch has taken place on the screen. Again, the ability to reliably and quickly measure users' abilities is of extreme importance.

Assess other Abilities. Our work focused on tremor-related impairments. However, our approach might easily translate to other domains, such as dexterity- or vision-based abilities. We only explored one component of human abilities that can be affected either by situational and health conditions. For instance, cold temperatures may affect users' finger dexterity in similar ways of chronicle arthritis. Similarly, there may be an overlap of interaction problems between speech-impaired people and mobile users in noisy environments, when trying to use a speech recognizer. Continuing to bridge this gap between accessibility and mobile research domains is of the utmost importance to raise awareness in the field, and to bring problems that are often ignored by "mainstream" designers/researchers into the discussion.

"System-Wide" Understanding of Users' Abilities. Typing improvements obtained by leveraging motion data suggest that a similar approach may be applied to other tasks. Stabilizing input resorting to similar techniques to those proposed in this dissertation, and making the operating system aware of users' hand movement patterns, may be useful to create general touch compensating techniques. Moreover, performance measures of tapping or gestural actions (e.g. swipe) should be available application-wide, allowing the operating system to reliably assess users' abilities and adjust the interface accordingly. Only then will we be able to create interfaces for each user.

8.3. Final Remarks

This dissertation demonstrated the following thesis:

> *We can improve typing performance of both health- and situational-impaired users by leveraging motion data on touch-based mobile devices.*

In fact, motion data proved to improve error classification accuracy. Moreover, our Unifying Model compensated typing errors of end-users, whether situationally- or health-impaired. Motion measures have shown to be able to characterize users' abilities in an objective and unifying manner. Empirical results demonstrated that this thesis indeed holds.

Moving forward it is important that both designers and researchers of mobile solutions consider their users' abilities as a dynamic entity. This *shift in thinking* is crucial to design more accessible devices and interfaces, which ultimately will be used by wider audiences. This will become extremely important as these tools take a preponderant role in our lives and are expected to effectively work in everyday situations, as well as for a target population with a wide range of abilities, expectations, values, beliefs, cultures, and so on. Looking through this perspective, the Mobile Accessibility research field is still in its infancy and the time to come is extraordinarily enthralling.

Bibliography

Abowd, G., Dey, A., Brown, P., Davies, N., Smith, M., and Steggles, P. (1999). Towards a Better Understanding of Context and Context-Awareness. In Gellersen, H.-W., editor, *Handheld and Ubiquitous Computing*, volume 1707 of *Lecture Notes in Computer Science*, pages 304–307. Springer Berlin / Heidelberg.

Al Faraj, K., Mojahid, M., and Vigouroux, N. (2009). BigKey: a virtual keyboard for mobile devices. *Human-Computer Interaction. Ambient, Ubiquitous and Intelligent Interaction*, pages 3–10.

Anouti, A. and Koller, W. C. (1995). Tremor disorders. Diagnosis and management. *Western journal of medicine*, 162(6):510.

Azenkot S., Wobbrock J.O., Prasain, S., and Ladner, R. E. (2012). Input Finger Detection for nonvisual touch screen text entry in Perkinput. In *Proceedings of Graphics Interface (GI '12).*

Bacher, M., Scholz, E., and Diener, H. C. (1989). 24 Hour continuous tremor quantification based on EMG recording. *Electroencephalography and clinical neurophysiology*, 72(2):176–183.

Bain, P. (1993). A combined clinical and neurophysiological approach to the study of patients with tremor. *Journal of Neurology, Neurosurgery & Psychiatry*, 56(8):839–844.

Bain, P. G., Mally, J., Gresty, M., and Findley, L. J. (1993). Assessing the impact of essential tremor on upper limb function. *Journal of neurology*, 241(1):54–61.

Baldwin, T. and Chai, J. (2012). Towards online adaptation and personalization of key-target resizing for mobile devices. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, IUI '12, pages 11–20, New York, NY, USA. ACM.

Barnard, L., Yi, J. S., Jacko, J. A., and Sears, A. (2005). An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices. *International Journal of Human-Computer Studies*, 62(4):487–520.

Barnard, L., Yi, J. S., Jacko, J. A., and Sears, A. (2007). Capturing the effects of context on human performance in mobile computing systems. *Personal and Ubiquitous Computing*, 11(2):81–96.

Beaton, R. J. and Weiman, N. (1984). Effects of touch key size and separation on menu-selection accuracy. *Proceedings of the Society of Photographic Scientists and Engineers*, pages 8–10.

Bender, G. T. (1999). *Touch screen performance as a function of the duration of auditory feedback and target size*. PhD thesis, Wichita State University.

Benito-Leon, J., Bermejo-Pareja, F., Morales, J. M., Vega, S., and Molina, J. A. (2003). Prevalence of essential tremor in three elderly populations of central Spain. *Movement disorders*, 18(4):389–394.

Bergstrom-Lehtovirta, J., Oulasvirta, A., and Brewster, S. (2011). The effects of walking speed on target acquisition on a touchscreen interface. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, pages 143–146. ACM.

Beringer, D. B. (1990). Target size, location, sampling point and instructional set: more effects on touch panel operation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 34, pages 375–379. SAGE Publications.

Birmingham, A. T., Wharrad, H. J., and Williams, E. J. (1985). The variation of finger tremor with age in man. *Journal of Neurology, Neurosurgery & Psychiatry*, 48(8):788–798.

Bonner, M., Brudvik, J., Abowd, G., and Edwards, W. (2010). No-Look Notes: Accessible eyes-free multi-touch text entry. *Pervasive Computing*, pages 409–426.

Brewster, S. (2002). Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing*, 6(3):188–205.

Brewster, S., Chohan, F., and Brown, L. (2007). Tactile feedback for mobile interactions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, page 162. ACM.

Brewster, S., Lumsden, J., Bell, M., Hall, M., and Tasker, S. (2003). Multimodal 'eyes-free' interaction techniques for wearable devices. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 473–480, New York, NY, USA. ACM.

Budde, P. (2009). Global Mobile Communications - Statistics, Trends and Forecasts. Technical report, BuddeComm.

Budde, P. (2010). Global Mobile Communications - Key Trends and Growth in a Challenging Environment. Technical report, BuddeComm.

Bunt, A., McGrenere, J., and Conati, C. (2007). Understanding the utility of rationale in a mixed-initiative system for GUI customization. *User Modeling 2007*, pages 147–156.

Burkhard, P. R., Langston, J. W., and Tetrud, J. W. (2002). Voluntarily simulated tremor in normal subjects. *Neurophysiologie Clinique/Clinical Neurophysiology*, 32(2):119–126.

Burkhard, P. R., Shale, H., Langston, J. W., and Tetrud, J. W. (1999). Quantification of dyskinesia in Parkinson's disease: validation of a novel instrumental method. *Movement disorders*, 14(5):754–763.

Carboncini, M. C., Manzoni, D., Strambi, S., Bonuccelli, U., Pavese, N., Andre, P., and Rossi, B. (2001). The relation between EMG activity and kinematic parameters strongly supports a role of the action tremor in parkinsonian bradykinesia. *Movement disorders*, 16(1):47–57.

Casiez, G., Roussel, N., and Vogel, D. (2012). 1Euro filter: a simple speed-based low-pass filter for noisy input in interactive systems. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, pages 2527–2530, New York, NY, USA. ACM.

Chen, T. (2008). Input to the mobile web is situationally-impaired. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*, pages 303–304. ACM New York, NY, USA.

Choudhury, T., Consolvo, S., Harrison, B., Hightower, J., LaMarca, A., LeGrand, L., Rahimi, A., Rea, A., Bordello, G., Hemingway, B., and Others (2008). The mobile sensing platform: An embedded activity recognition system. *Pervasive Computing, IEEE*, 7(2):32–41.

Clawson, J., Lyons, K., Rudnick, A., Iannucci Jr, R. A., and Starner, T. (2008). Automatic whiteout++: correcting mini-QWERTY typing errors using keypress timing. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 573–582. ACM.

Clawson, J., Rudnick, A., Lyons, K., and Starner, T. (2007). Automatic whiteout: Discovery and correction of typographical errors in mobile text input. In *MobileHCI07: Proceedings of the 9th conference on Human-computer interaction with mobile devices and services*.

Colle, H. A. and Hiszem, K. J. (2004). Standing at a kiosk: Effects of key size and spacing on touch screen numeric keypad performance and user preference. *Ergonomics*, 47(13):1406–1423.

Cook, A. M. and Hussey, S. (2001). *Assistive Technologies: Principles and Practice (2nd Edition)*. Mosby, 2 edition.

Crossan, A., Murray-Smith, R., Brewster, S., Kelly, J., and Musizza, B. (2005). Gait phase effects in mobile interaction. In *CHI'05 extended abstracts on Human factors in computing systems*, pages 1312–1315. ACM.

Dey, A. K. (2001). Understanding and using context. *Personal and ubiquitous computing*, 5(1):4–7.

Dunlop, M. and Levine, J. (2012). Multidimensional pareto optimization of touchscreen keyboards for speed, familiarity and improved spell checking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2669–2678, New York, NY, USA. ACM.

Elble, R. J. (1998). Tremor in ostensibly normal elderly people. *Movement disorders*, 13(3):457–464.

Elble, R. J. (2003). Characteristics of physiologic tremor in young and elderly adults. *Clinical neurophysiology*, 114(4):624–635.

Elble, R. J., Brilliant, M., Leffler, K., and Higgins, C. (1996). Quantification of essential tremor in writing and drawing. *Movement disorders*, 11(1):70–78.

Elble, R. J. and Koller, W. C. (1990). *Tremor*. Johns Hopkins University Press Baltimore:.

Elble, R. J., Sinha, R., and Higgins, C. (1990). Quantification of tremor with a digitizing tablet. *Journal of neuroscience methods*, 32(3):193–198.

Findlater, L. and Wobbrock, J. O. (2012). Personalized Input: Improving Ten-Finger Touchscreen Typing through Automatic Adaptation. *Proceedings of the 2012 annual conference on human factors in computing systems (CHI'12)*.

Findlater, L., Wobbrock, J. O., and Wigdor, D. (2011). Typing on flat glass: examining ten-finger expert typing patterns on touch surfaces. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, pages 2453–2462. ACM.

Fleiss, J. L. (1973). *Statistical methods for rates and proportions*. John Wiley & Sons.

Frey, B., Southern, C., and Romero, M. (2011). Brailletouch: mobile texting for the visually impaired. *Universal Access in Human-Computer Interaction. Context Diversity*, pages 19–25.

Froehlich, J., Wobbrock, J. O., and Kane, S. K. (2007). Barrier pointing: using physical edges to assist target acquisition on mobile device touch screens. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, Assets '07, pages 19–26, New York, NY, USA. ACM.

Gajos, K. and Weld, D. S. (2004). SUPPLE: automatically generating user interfaces. In *Proceedings of the 9th international conference on Intelligent user interfaces*, IUI '04, pages 93–100, New York, NY, USA. ACM.

Gajos, K. Z., Wobbrock, J. O., and Weld, D. S. (2007). Automatically generating user interfaces adapted to users' motor and vision capabilities. In *UIST '07: Proceedings of the 20th annual ACM symposium on User interface software and technology*, pages 231–240, New York, NY, USA.

Gerr, F., Letz, R., and Green, R. C. (2000). Relationships between quantitative measures and neurologist's clinical rating of tremor and standing steadiness in two epidemiological studies. *Neurotoxicology*, 21(5):753.

Go, K. and Endo, Y. (2007). CATKey: customizable and adaptable touchscreen keyboard with bubble cursor-like visual feedback. In *Proceedings of the 11th IFIP TC 13 international conference on Human-computer interaction*, INTERACT'07, pages 493–496, Berlin, Heidelberg. Springer-Verlag.

Goel, M., Findlater, L., and Wobbrock, J. O. (2012). WalkType: Using Accelerometer Data to Accommodate Situational Impairments in Mobile Touch Screen Text Entry. *Proceedings of the 2012 annual conference on human factors in computing systems (CHI'12).*

Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M. B., Dodel, R., and Others (2008). 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.

Goodman, J., Venolia, G., Steury, K., and Parker, C. (2002). Language modeling for soft keyboards. In *Proceedings of the 7th international conference on Intelligent user interfaces*, pages 194–195. ACM.

Gregor, P., Newell, A. F., and Zajicek, M. (2002). Designing for dynamic diversity: interfaces for older people. In *Assets '02: Proceedings of the fifth international ACM conference on Assistive technologies*, pages 151–156, New York, NY, USA. ACM.

Guerreiro, T., Lagoá, P., Nicolau, H., Gonçalves, D., and Jorge, J. A. (2008). From tapping to touching: Making touch screens accessible to blind users. *IEEE MultiMedia*, pages 48–50.

Guerreiro, T., Nicolau, H., Jorge, J., and Gonçalves, D. (2009). NavTap: a long term study with excluded blind users. In *Proceeding of the eleventh international ACM SIGACCESS conference on Computers and accessibility*, pages 99–106. ACM.

Guerreiro, T., Nicolau, H., Jorge, J., and Gonçalves, D. (2010). Towards accessible touch interfaces. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*, ASSETS '10, pages 19–26, New York, NY, USA. ACM.

Gunawardana, A., Paek, T., and Meek, C. (2010). Usability guided key-target resizing for soft keyboards. In *Proceedings of the 15th international conference on Intelligent user interfaces*, pages 111–118. ACM.

Hallett, M. (1991). Classification and treatment of tremor. *JAMA: the journal of the American Medical Association*, 266(8):1115–1117.

Harper, S. (2007). Is there design-for-all? *Universal Access in the Information Society*, 6(1):111–113.

Harrin, E. (2008). Universal usability: Designing computer interfaces for diverse users. *The Computer Journal*.

Henze, N., Rukzio, E., and Boll, S. (2011). 100,000,000 taps: analysis and improvement of touch performance in the large. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, pages 133–142, New York, NY, USA. ACM.

Henze, N., Rukzio, E., and Boll, S. (2012). Observational and experimental investigation of typing behaviour using virtual keyboards for mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2659–2668, New York, NY, USA. ACM.

Himberg, J., Häkkilä, J., Kangas, P., and Mäntyjärvi, J. (2003). On-line personalization of a touch screen based keyboard. In *Proceedings of the 8th international conference on Intelligent user interfaces*, pages 77–84. ACM.

Hinckley, K., Pierce, J., Sinclair, M., and Horvitz, E. (2000). Sensing techniques for mobile interaction. In *Symposium on User Interface Software and Technology: Proceedings of the 13 th annual ACM symposium on User interface software and technology*, volume 6, pages 91–100. Citeseer.

Hoff, J. I., Wagemans, E. A., and van Hilten, B. J. (2001). Ambulatory objective assessment of tremor in Parkinson's disease. *Clinical neuropharmacology*, 24(5):280.

Hoggan, E., Brewster, S. A., and Johnston, J. (2008). Investigating the effectiveness of tactile feedback for mobile touchscreens. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 1573–1582, New York, NY, USA.

Holz, C. and Baudisch, P. (2011). Understanding touch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2501–2510, New York, NY, USA. ACM.

Höök, K. (2000). Steps to take before intelligent user interfaces become real. *Interacting with computers*, 12(4):409–426.

Hurst, A., Hudson, S. E., and Mankoff, J. (2007). Dynamic detection of novice vs. skilled use without a task model. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 271–280, New York, NY, USA. ACM.

Hurst, A., HUDSON, S. E., Mankoff, J., and Trewin, S. (2008). Automatically detecting pointing performance. In *IUI '08: Proceedings of the 13th international conference on Intelligent user interfaces*, pages 11–19, New York, NY, USA.

Hwang, F., Keates, S., Langdon, P., and Clarkson, J. (2004). Mouse movements of motion-impaired users: a submovement analysis. In *Proceedings of the 6th international ACM SIGACCESS conference on Computers and accessibility*, Assets '04, pages 102–109, New York, NY, USA. ACM.

Jin, Z., Plocher, T., and Kiff, L. (2007). Touch screen user interfaces for older adults: button size and spacing. *Universal Acess in Human Computer Interaction. Coping with Diversity*, pages 933–941.

Kane, S. K., Wobbrock, J. O., Harniss, M., and Johnson, K. L. (2008a). TrueKeys: identifying and correcting typing errors for people with motor impairments. In *IUI '08: Proceedings of the 13th international conference on Intelligent user interfaces*, pages 349–352, New York, NY, USA.

Kane, S. K., Wobbrock, J. O., and Ladner, R. E. (2011). Usable gestures for blind people: understanding preference and performance. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, pages 413–422. ACM.

Kane, S. K., Wobbrock, J. O., and Smith, I. E. (2008b). Getting off the treadmill: evaluating walking user interfaces for mobile devices in public spaces. In *MobileHCI '08: Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*, pages 109–118, New York, NY, USA.

Keates, S., Trewin, S., and Paradise, J. (2005). Using pointing devices: Quantifying differences across user groups. *Proceedings of UAHCI.*

Kelly, J., Taggart, H. M., and McCullagh, P. (1994). Normal and abnormal tremor in the elderly. *Neurological Disease and Therapy*, 30:351–370.

Koester, H. H., LoPresti, E., and Simpson, R. C. (2005). Toward Goldilocks' pointing device: determining a "just right" gain setting for users with physical impairments. In *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, Assets '05, pages 84–89, New York, NY, USA. ACM.

Koester, H. H., Lopresti, E., and Simpson, R. C. (2007a). Toward automatic adjustment of keyboard settings for people with physical impairments. *Disability & Rehabilitation: Assistive Technology*, 2(5):261–274.

Koester, H. H., LoPresti, E. F., and Simpson, R. C. (2006). Measurement validity for Compass assessment software. In *Proceedings of the RESNA 29th Annual Conference (RESNA'06).*

Koester, H. H., Simpson, R. C., Spaeth, D., and LoPresti, E. F. (2007b). Reliability and validity of Compass software for access assessment. In *Proceedings of the RESNA 30th Annual Conference (RESNA'07).*

Kondraske, G. V. (1995). A working model for human system-task interfaces. *The biomedical engineering handbook*, 2:147.

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

Kristensson, P. O. and Zhai, S. (2004). SHARK 2: a large vocabulary shorthand writing system for pen-based computers. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 43–52. ACM.

Kristensson, P. O. and Zhai, S. (2005). Relaxing stylus typing precision by geometric pattern matching. In *Proceedings of the 10th international conference on Intelligent user interfaces*, pages 151–158. ACM.

Kristoffersen, S. and Ljungberg, F. (1999). Making place to make IT work: empirical explorations of HCI for mobile CSCW. In *Proceedings of the international ACM SIGGROUP conference on Supporting group work*, pages 276–285. ACM New York, NY, USA.

Lee, S. and Zhai, S. (2009). The performance of touch screen soft buttons. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 309–318, New York, NY, USA. ACM.

Li, K. A., Baudisch, P., and Hinckley, K. (2008). Blindsight: eyes-free access to mobile phones. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 1389–1398, New York, NY, USA.

Lin, M., Goldman, R., Price, K. J., Sears, A., and Jacko, J. (2007). How do people tap when walking? An empirical investigation of nomadic data entry. *International Journal of Human-Computer Studies*, 65(9):759–769.

Louis, E. D. and Ferreira, J. J. (2010). How common is the most common adult movement disorder? Update on the worldwide prevalence of essential tremor. *Movement Disorders*, 25(5):534–541.

Lukhanina, E. P., Kapoustina, M. T., and Karaban, I. N. (2000). A quantitative surface electromyogram analysis for diagnosis and therapy control in Parkinson's disease. *Parkinsonism & Related Disorders*, 6(2):77–86.

Lumsden, J. and Brewster, S. (2003). A paradigm shift: alternative interaction techniques for use with mobile & wearable devices. In *Proceedings of the 2003 conference of the Centre for Advanced Studies on Collaborative research*, page 210. IBM Press.

Mace, R. L., Hardie, G. J., and Place, J. P. (1990). *Accessible environments: Toward universal design*. Center for Accessible Housing, North Carolina State University.

MacKenzie, I. S. and Soukoreff, R. W. (2002a). A character-level error analysis technique for evaluating text entry methods. In *Proceedings of the second Nordic conference on Human-computer interaction*, pages 243–246. ACM.

MacKenzie, I. S. and Soukoreff, R. W. (2002b). Text entry for mobile computing: Models and methods, theory and practice. *Human–Computer Interaction*, 17(2):147–198.

MacKenzie, I. S. and Soukoreff, R. W. (2003). Phrase sets for evaluating text entry techniques. In *CHI'03 extended abstracts on Human factors in computing systems*, pages 754–755. ACM.

MacKenzie, I. S. and Zhang, S. X. (1999). The design and evaluation of a high-performance soft keyboard. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, pages 25–31. ACM.

Marsden, C. D., Meadows, J. C., Lange, G. W., and Watson, R. S. (1969). Variations in human physiological finger tremor, with particular reference to changes with age. *Electroencephalography and clinical neurophysiology*, 27(2):169–178.

Martin, G. L. (1988). Configuring a numeric keypad for a touch screen. *Ergonomics*, 31(6):945–953.

Martinez-Martin, P., Gil-Nagel, A., Gracia, L. M., Gomez, J. B., Martinez-Sarries, J., and Bermejo, F. (1994). Unified Parkinson's disease rating scale characteristics and structure. *Movement Disorders*, 9(1):76–83.

Mertens, A., Jochems, N., Schlick, C. M., Dünnebacke, D., and Dornberg, J. H. (2010). Design pattern TRABING: touchscreen-based input technique for people affected by intention tremor. In *Proceedings of the 2nd ACM SIGCHI symposium on Engineering interactive computing systems*, EICS '10, pages 267–272, New York, NY, USA. ACM.

Mizobuchi, S., Chignell, M., and Newton, D. (2005). Mobile text entry: relationship between walking speed and text input task difficulty. In *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*, page 128. ACM.

Moffatt, K. and McGrenere, J. (2010). Steadied-bubbles: combining techniques to address pen-based pointing errors for younger and older adults. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pages 1125–1134, New York, NY, USA. ACM.

Montague, K., Hanson, V. L., and Cobley, A. (2012). Designing for individuals: usable touch-screen interaction through shared user models. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*, ASSETS '12, pages 151–158, New York, NY, USA. ACM.

Mustonen, T., Olkkonen, M., and Hakkinen, J. (2004). Examining mobile phone text legibility while walking. In *Conference on Human Factors in Computing Systems*, pages 1243–1246. ACM New York, NY, USA.

Myers, B. A., Wobbrock, J. O., Yang, S., Yeung, B., Nichols, J., and Miller, R. (2002). Using handhelds to help people with motor impairments. In *Proceedings of the fifth international ACM conference on Assistive technologies*, page 96. ACM.

Newell, A. F. (1995). *Extra-ordinary human-computer interaction*, pages 3–18. Cambridge University Press, New York, NY, USA.

Newell, A. F. and Gregor, P. (2000). User sensitive inclusive design - in search of a new paradigm. In *Proceedings on the 2000 conference on Universal Usability*, pages 39–44. ACM.

Nicolau, H., Guerreiro, T., Jorge, J., and Gonçalves, D. (2012a). Mobile Text-Entry and Visual Demands: Reusing and Optimizing Current Solutions. *Universal Access in the Information Society*, (Special Issue on Mobile Accessibility).

Nicolau, H., Guerreiro, T., Lucas, D., and Jorge, J. (2012b). Mobile Touch Screen User Interfaces: Bridging the Gap between Motor-Impaired and Able-Bodied Users. *Universal Access in the Information Society (UAIS)*, (Special Issue on Mobile Accessibility).

Nicolau, H. and Jorge, J. (2012a). Elderly Text-Entry Performance on Touchscreens. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*.

Nicolau, H. and Jorge, J. (2012b). The Disability Continuum: Investigating Health and Situational Induced Impairments and Disabilities. In *In 14th International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI'12 Workshop on Mobile Accessibility*, San Francisco, USA.

Nicolau, H. and Jorge, J. (2012c). Touch typing using thumbs: understanding the effect of mobility and hand posture. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, pages 2683–2686, New York, NY, USA. ACM.

Nielsen, J. (1993). *Usability Engineering*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Norman, D. A. and Draper, S. W. (1986). *User Centered System Design; New Perspectives on Human-Computer Interaction*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA.

Norman, K. E., Edwards, R., and Beuter, A. (1999). The measurement of tremor using a velocity transducer: comparison to simultaneous recordings using transducers of displacement, acceleration and muscle activity. *Journal of neuroscience methods*, 92(1):41–54.

Novak, M., Schauer, J., Hinkens, J., and Vanderheiden, G. (1991). Providing computer access features under DOS. In *Resna*, volume 91, pages 163–165.

Oliveira, J., Guerreiro, T., Nicolau, H., Jorge, J., and Gonçalves, D. (2011a). Blind people and mobile touch-based text-entry: acknowledging the need for different flavors. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*, pages 179–186. ACM.

Oliveira, J., Guerreiro, T., Nicolau, H., Jorge, J., and Gonçalves, D. (2011b). BrailleType: unleashing braille over touch screen mobile phones. *Human-Computer Interaction–INTERACT 2011*, pages 100–107.

O'Suilleabhain, P. E. and Dewey Jr, R. B. (2001). Validation for tremor quantification of an electromagnetic tracking device. *Movement disorders*, 16(2):265–271.

Oulasvirta, A., Tamminen, S., Roto, V., and Kuorelahti, J. (2005). Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 919–928, New York, NY, USA. ACM.

Parhi, P., Karlson, A. K., and Bederson, B. B. (2006). Target size study for one-handed thumb use on small touchscreen devices. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*, MobileHCI '06, pages 203–210, New York, NY, USA. ACM.

Park, Y. S., Han, S. H., Park, J., and Cho, Y. (2008). Touch key design for target selection on a mobile phone. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*, pages 423–426. ACM.

Pascoe, J., Ryan, N., and Morse, D. (2000). Using while moving: HCI issues in fieldwork environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(3):417–437.

Pascoe, M. J. (1998). Adding Generic Contextual Capabilities to Wearable Computers. In *Proceedings of the 2nd IEEE International Symposium on Wearable Computers*, ISWC '98, pages 92—-, Washington, DC, USA. IEEE Computer Society.

Pavlovych, A. and Stuerzlinger, W. (2004). Model for non-expert text entry speed on 12-button phone keypads. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 351–358. ACM.

Paymans, T. F., Lindenberg, J., and Neerincx, M. (2004). Usability trade-offs for adaptive user interfaces: ease of use and learnability. In *Proceedings of the 9th international conference on Intelligent user interfaces*, IUI '04, pages 301–303, New York, NY, USA. ACM.

Persad, U., Langdon, P., and Clarkson, J. (2007). Characterising user capabilities to support inclusive design evaluation. *Universal Access in the Information Society*, 6(2):119–135.

Pielot, M., Kazakova, A., Hesselmann, T., Heuten, W., and Boll, S. (2012). PocketMenu: non-visual menus for touch screen devices. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, MobileHCI '12, pages 327–330, New York, NY, USA. ACM.

Pirhonen, A., Brewster, S., and Holguin, C. (2002). Gestural and audio metaphors as a means of control for mobile devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, page 298. ACM.

Powers, S. and Brown, V. (2012). Design Mobile Apps from the Outside In. Technical report, Policy and Procedures: The Mobile App Development Playbook.

Raethjen, J., Pawlas, F., Lindemann, M., Wenzelburger, R., and Deuschl, G. (2000). Determinants of physiologic tremor in a large normal population. *Clinical neurophysiology*, 111(10):1825–1837.

Rahmati, A., Shepard, C., and Zhong, L. (2009). NoShake: Content stabilization for shaking screens of mobile devices. In *Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on*, pages 1–6. IEEE.

Rajaraman, V., Jack, D., Adamovich, S. V., Hening, W., Sage, J., and Poizner, H. (2000). A novel quantitative method for 3D measurement of Parkinsonian tremor. *Clinical neurophysiology*, 111(2):338–343.

Richtel, M. (2010). Forget gum. Walking and using phone is risky. *The New York Times*.

Riviere, C. N., Reich, S. G., and Thakor, N. V. (1997). Adaptive Fourier modeling for quantification of tremor1. *Journal of neuroscience methods*, 74(1):77–87.

Rudchenko, D., Paek, T., and Badger, E. (2011). Text text revolution: a game that improves text entry on mobile touchscreen keyboards. *Pervasive Computing*, pages 206–213.

Salarian, A., Russmann, H., Wider, C., Burkhard, P. R., Vingerhoets, F. J. G., and Aminian, K. (2007). Quantification of tremor and bradykinesia in Parkinson's disease using a novel ambulatory monitoring system. *Biomedical Engineering, IEEE Transactions on*, 54(2):313–322.

Salber, D., Dey, A. K., and Abowd, G. D. (1999). The context toolkit: aiding the development of context-enabled applications. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, CHI '99, pages 434–441, New York, NY, USA. ACM.

Sawhney, N. and Schmandt, C. (2000). Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(3):383.

Schedlbauer, M. and Heines, J. (2007). Selecting While Walking: An Investigation of Aiming Performance in a Mobile Work Context. *AMCIS 2007 Proceedings*, page 433.

Schildbach, B. and Rukzio, E. (2010). Investigating selection and reading performance on a mobile phone while walking. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, MobileHCI '10, pages 93–102, New York, NY, USA. ACM.

Schmidt, A., Aidoo, K., Takaluoma, A., Tuomela, U., Van Laerhoven, K., and de Velde, W. (1999). Advanced interaction in context. In *HandHeld and Ubiquitous Computing*, pages 89–101. Springer.

Scott, B. and Conzola, V. (1997). Designing Touch Screen Numeric Keypads Effects of Finger Size, Key Size, and Key Spacing. In *Human Factors and Ergonomics Society Annual Meeting Proceedings*, volume 41, pages 360–364. Human Factors and Ergonomics Society.

Sears, A., Lin, M., Jacko, J., and Xiao, Y. (2003). When computers fade: Pervasive computing and situationallyinduced impairments and disabilities. In *Proc. HCI Int*, volume 2, pages 1298–1302.

Sears, A. and Young, M. (2003). Physical disabilities and computing technology: An analysis of impairments. *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications. Lawrence Erlbaum, New Jersey, USA*, pages 482–503.

Selker, T., Collins, P., and Dayton, W. (2011). Psychosocial indicators via hand tremor. *Human-Computer Interaction–INTERACT 2011*, pages 596–599.

Shneiderman, B. (1987). *Designing the user interface (videotape).* University of Maryland at College Park, College Park, MD, USA.

Shneiderman, B. (2000). Universal usability. *Communications of the ACM*, 43(5):84–91.

Song, Y. C. (2010). Joystick text entry with word prediction for people with motor impairments. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*, ASSETS '10, pages 321–322, New York, NY, USA. ACM.

Spieker, S., Boose, A., Breit, S., and Dichgans, J. (1998). Long-Term Measurement of Tremor. *Movement disorders*, 13(S3):81–84.

Steinfeld, E. (1994). The concept of universal design. In *Proceedings of the Sixth Ibero-American Conference on Accessibility.*

Stephanidis, C. (1995). Towards user interfaces for all: some critical issues. In Yuichiro Anzai, K. O. and Mori, H., editors, *Symbiosis of Human and Artifact Future Computing and Design for Human-Computer Interaction Proceedings of the Sixth International Conference on Human-Computer Interaction, (HCI International '95)*, volume 20 of *Advances in Human Factors/Ergonomics*, pages 137–142. Elsevier.

Stephanidis, C. (2000). *User interfaces for all: Concepts, methods, and tools.* CRC.

Stephanidis, C. (2001). User interfaces for all: New perspectives into human-computer interaction. *User Interfaces for All-Concepts, Methods, and Tools*, 1:3–17.

Stephanidis, C. and Salvendy, G. (1998). Toward an information society for all: An international research and development agenda. *International Journal of Human-Computer Interaction*, 10(2):107–134.

Stephanidis, C., Salvendy, G., Akoumianakis, D., Bevan, N., Brewer, J., Emiliani, P. L., Galetsas, A., Haataja, S., Iakovidis, I., Jacko, J., and Others (1998). Toward an information society for all: an international R&D agenda. *International Journal of Human-Computer Interaction*, 10(2):107–134.

Story, M. F. (1998). Maximizing usability: the principles of universal design. *Assistive technology*, 10(1):4–12.

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

Sturman, M. M., Vaillancourt, D. E., and Corcos, D. M. (2005). Effects of aging on the regularity of physiological tremor. *Journal of neurophysiology*, 93(6):3064–3074.

Sun, X., Plocher, T., and Qu, W. (2007). An empirical study on the smallest comfortable button/icon size on touch screen. *Usability and Internationalization. HCI and Culture*, pages 615–621.

Tinwala, H. and MacKenzie, I. S. (2010). Eyes-free text entry with error correction on touchscreen mobile devices. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, pages 511–520. ACM.

Trewin, S. (2000). Configuration agents, control and privacy. In *Proceedings on the 2000 conference on Universal Usability*, CUU '00, pages 9–16, New York, NY, USA. ACM.

Trewin, S. (2002). An invisible keyguard. In *Proceedings of the fifth international ACM conference on Assistive technologies*, pages 143–149. ACM.

Trewin, S. (2003). Automating accessibility: the dynamic keyboard. *ACM SIGACCESS Accessibility and Computing*, (77-78):71–78.

Trewin, S., Keates, S., and Moffatt, K. (2006). Developing steady clicks:: a method of cursor assistance for people with motor impairments. In *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*, Assets '06, pages 26–33, New York, NY, USA. ACM.

Trewin, S. and Pain, H. (1999). Keyboard and mouse errors due to motor disabilities. *International Journal of Human-Computer Studies*, 50(2):109–144.

Trewin, S., Pain, H., and of Edinburgh. Dept. of Artificial Intelligence, U. (1997). Dynamic Modelling of Keyboard Skills Supporting Users with Motor Disabilities. *Courses and lectures - International Centre for Mechanical Sciences*, pages 135–146.

Van Den Eeden, S. K., Tanner, C. M., Bernstein, A. L., Fross, R. D., Leimpeter, A., Bloch, D. A., and Nelson, L. M. (2003). Incidence of Parkinsons disease: variation by age, gender, and race/ethnicity. *American Journal of Epidemiology*, 157(11):1015–1022.

Van Someren, E. J. W. (1997). Actigraphic monitoring of movement and rest-activity rhythms in aging, Alzheimer's disease, and Parkinson's disease. *Rehabilitation Engineering, IEEE Transactions on*, 5(4):394–398.

Vanderheiden, G. (2000). Fundamental principles and priority setting for universal usability. In *Proceedings on the 2000 conference on Universal Usability*, CUU '00, pages 32–37, New York, NY, USA. ACM.

Vanderheiden, G. C. (1998). Universal design and assistive technology in communication and information technologies: alternatives or complements? *Assistive technology: the official journal of RESNA*, 10(1):29.

Wacharamanotham, C., Hurtmanns, J., Mertens, A., Kronenbuerger, M., Schlick, C., and Borchers, J. (2011). Evaluating swabbing: a touchscreen input method for elderly users with tremor. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, pages 623–626. ACM.

Wade, P., Gresty, M. A., and Findley, L. J. (1982). A normative study of postural tremor of the hand. *Archives of neurology*, 39(6):358.

Waite, L. M., Broe, G. A., Creasey, H., Grayson, D., Edelbrock, D., and O'Toole, B. (1996). Neurological signs, aging, and the neurodegenerative syndromes. *Archives of neurology*, 53(6):498.

Watkins, K., Fu, H., Fuentes, R., Ghosh, A., Giamberardini, C., Johansson, C., Kuonqui, C., Montes, A., Stewart, D., Ugaz, C., and Others (2005). Human development report 2005. *New York: United Nations Development Programme*, page 212.

Weiser, M. (1993). Ubiquitous computing. *IEEE Computer*, 26(10):71–72.

WHO (2009). International Classification of Functioning, Disability and Health (ICF). Technical report, WHO Press, Geneva.

Wickens, C. D. (2005). Multiple resources and performance prediction. *Ergonomics: Psychological mechanisms and models in ergonomics*, page 83.

Wilson, K. S. and Liu, M. I. S. (1995). A comparison of five user interface devices designed for point-of-sale in the retail industry. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 39, pages 273–277. SAGE Publications.

Witten, I. H. and Frank, E. (1999). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (The Morgan Kaufmann Series in Data Management Systems)*. Morgan Kaufmann, 1st edition.

Wobbrock, J. O. (2006). The future of mobile device research in HCI. In *CHI 2006 Workshop Proceedings: What is the Next Generation of Human-Computer Interaction*, pages 131–134. Citeseer.

Wobbrock, J. O., Aung, H. H., Rothrock, B., and Myers, B. A. (2005). Maximizing the guessability of symbolic input. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, pages 1869–1872, New York, NY, USA. ACM.

Wobbrock, J. O., Chau, D. H., and Myers, B. A. (2007). An alternative to push, press, and tap-tap-tap: gesturing on an isometric joystick for mobile phone text entry. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, page 676. ACM.

Wobbrock, J. O., Fogarty, J., Liu, S.-Y. S., Kimuro, S., and Harada, S. (2009). The angle mouse: target-agnostic dynamic gain adjustment based on angular deviation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1401–1410, New York, NY, USA. ACM.

Wobbrock, J. O. and Gajos, K. Z. (2007). A comparison of area pointing and goal crossing for people with and without motor impairments. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, Assets '07, pages 3–10, New York, NY, USA. ACM.

Wobbrock, J. O. and Gajos, K. Z. (2008). Goal Crossing with Mice and Trackballs for People with Motor Impairments: Performance, Submovements, and Design Directions. *ACM Trans. Access. Comput.*, 1(1):4:1—-4:37.

Wobbrock, J. O., Kane, S. K., Gajos, K. Z., Harada, S., and Froehlich, J. (2011). Ability-based design: Concept, principles and examples. *ACM Transactions on Accessible Computing (TACCESS)*, 3(3):9.

Wobbrock, J. O. and Myers, B. A. (2006). Analyzing the input stream for character- level errors in unconstrained text entry evaluations. *ACM Trans. Comput.-Hum. Interact.*, 13(4):458–489.

Wobbrock, J. O., Myers, B. A., and Kembel, J. A. (2003). EdgeWrite: a stylus-based text entry method designed for high accuracy and stability of motion. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, page 70. ACM.

Wyne, K. T. (2005). A comprehensive review of tremor. *Jaapa*, 18(12):43–50.

Yamabe, T. and Takahashi, K. (2007). Experiments in mobile user interface adaptation for walking users. In *Intelligent Pervasive Computing, 2007. IPC. The 2007 International Conference on*, pages 280–284. IEEE.

Yatani, K. and Truong, K. N. (2009). An evaluation of stylus-based text entry methods on handheld devices studied in different user mobility states. *Pervasive and Mobile Computing*, 5(5):496–508.

Yesilada, Y., Chen, T., and Harper, S. (2010a). A simple solution: solution migration from disabled to small device context. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)*, W4A '10, pages 27:1—-27:2, New York, NY, USA. ACM.

Yesilada, Y., Harper, S., Chen, T., and Trewin, S. (2010b). Small-device users situationally impaired by input. *Computers in Human Behavior*.

Yfantidis, G. and Evreinov, G. (2006). Adaptive blind interaction technique for touchscreens. *Universal Access in the Information Society*, 4(4):328–337.

Zhai, S., Hunter, M., and Smith, B. A. (2000). The metropolis keyboard-an exploration of quantitative techniques for virtual keyboard design. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*, pages 119–128. ACM.

Zhai, S., Hunter, M., and Smith, B. A. (2002). Performance optimization of virtual keyboards. *Human–Computer Interaction*, 17(2-3):229–269.

Zhai, S. and Kristensson, P. O. (2003). Shorthand writing on stylus keyboard. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 97–104. ACM.

Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R., and Baudisch, P. (2007). Earpod: eyes-free menu selection using touch input and reactive audio feedback. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1395–1404, New York, NY, USA. ACM.

A1 Evaluation Sentences

This appendix contains:

- Set of sentences used in text-entry evaluations.

To select this set of sentences we followed a similar approach to MacKenzie et al. (MacKenzie and Soukoreff, 2003). We extracted sentences with an average of 5 words and 4.48 characters per word from a Portuguese written language *corpus*. Moreover, each sentence had a minimum correlation with the language of 0.97.

The following sentences were used in all text-entry evaluations presented in this dissertation:

- uma poderosa corrente de insatisfacao
- lugares de estacionamento mais procurados
- assembleia de cooperantes autorizou ainda
- decisoes a tomar na reuniao
- muito para alem do necessario
- a restauracao do estomago presidencial
- autor em casos de interoperacionalidade
- sera equipada com modernas tecnologias
- algo de muito necessario para
- da posicao anteriormente assumida pelo
- indeterminado para questoes a colocar
- consensual desde a primeira votacao
- de apoio a causa timorense
- uma dimensao de representacao social
- luciano patrao disse nao terem
- de incompreensao toldaria a sua
- um espaco tradicionalmente reservado aos
- reunioes da comissao parlamentar de
- dao entrada nas competicoes europeias
- de desorientacao e desanimo alastrou
- receitado alguns medicamentos para os
- da posicao anteriormente assumida pelo
- alimentares ou a doencas por
- relacionamento com as respectivas autoridades
- de apoio a causa timorense
- reuniao sectorial da comissao permanente
- comunicada as autoridades por elementos
- a impressao de desorientacao numa
- aumentando o risco de represalias

- varios aquartelamentos conforme a especialidade
- ocasionar o desaparecimento de algumas
- e do repoliciamento das ruas
- queda das compartimentacoes regionais pelo
- ou ainda em editoras escolares
- as interpelacoes da vereadora comunista
- castro almeida nao precisou de
- muito para alem do necessario
- instalado entre as democracias europeias
- estava sobretudo relacionada com espionagem
- de respeitar a decisao autonoma
- duas prestacoes ao alegado intermediario
- da retoma das economias europeia
- seleccionador maturana e disse ao
- instrucao do processo teria alegadamente
- nao tem aparecido as reunioes
- material necessario ao desaparecimento dos
- se pode considerar eutanasia como
- producao de alimentos e materias
- dado ministerio ou apenas alteracoes
- seu lado considera importante a
- os indicadores apresentarem uma evolucao
- o seu metodo ainda precisara
- acudir aos problemas de estacionamento
- responder a uma solicitacao de
- americano aos produtos alimentares de
- o desespero da minoria culta
- mediante a instauracao de processo
- sua posicao de maior entre
- responder a uma solicitacao de
- as operacoes militares da onu
- metropolitano das imediacoes serao encerradas
- cosmologia ou a eternidade necessaria
- ainda acentuados pelo desaparecimento misterioso

- decisao relativa ao prosseguimento da
- oposicao da inglaterra durante meses
- sobretudo pela apetencia dos americanos
- uma investigacao ordenada pelo secretario
- presidente da maior associacao levou
- reaccao aos insultos alegadamente proferidos
- de acelerar os pagamentos comunitarios
- espacial ate aos pormenores cuidadosamente
- agradou especialmente ao ministro das
- citados pela reuter nao escondiam
- aos investidores a regulamentacao do
- conversacoes eram a ultima oportunidade
- na comissao europeia terao de
- sera uma troca de opinioes
- decisoes a tomar na reuniao
- comissao instaladora da empresa entregou
- a europa de casais monteiro
- estalou a crise monetaria do
- lado esquerdo pertencia aos americanos
- os indicadores apresentarem uma evolucao
- as operacoes militares da onu
- comensalidade nos temos a prefiguracao
- pode conceder amnistia aos autores
- os centrais paulo madeira e
- acreditamos nele ou nao acreditamos
- denuncia sobre o alegado separatismo
- treinador escoces apostou ainda em
- resultado era de somenos importancia
- reuniao da comissao de arte
- eventual ostracismo na propria sociedade
- com anterioridade as nacoes europeias
- pecas fundamentais do relatorio e

A2
Motor Demands Study Materials

This appendix contains:

- Background questionnaire
- Functional assessment evaluation
- Evaluation monitor checklist
- Debriefing questionnaire
- Motor-impaired participants profile
- Motor-impaired funcional assessment results

Background Questionnaire

Interview #____

Participant

1) Name: ____________________

2) E-mail: ____________________

3) Telephone: ____________________

4) Sex:
Male ☐
Female ☐

5) Age: ____________________

6) Habilitations:

4th Grade ☐
9th Grade ☐
12th Grade ☐
BsC ☐
MsC ☐
PhD ☐

Clinical History

7) Level of cervical lesion: ______________________

8) Complete or incomplete: ______________________

9) How many years have you become tetraplegic? __________

10) What was the cause for the impairment? ______________________

11) Point out the controlled areas:

Area	**Left**	**Right**
Eyes	☐	☐
Face	☐	☐
Neck	☐	☐
Shoulder	☐	☐
Arm	☐	☐
Forearm	☐	☐
Hand	☐	☐
Fingers	☐	☐

Clinical History

7) Level of cervical lesion: ______________________

8) Complete or incomplete: ______________________

9) How many years have you become tetraplegic? __________

10) What was the cause for the impairment? ______________________

11) Point out the controlled areas:

Area	Left	Right
Eyes	☐	☐
Face	☐	☐
Neck	☐	☐
Shoulder	☐	☐
Arm	☐	☐
Forearm	☐	☐
Hand	☐	☐
Fingers	☐	☐

Mobile Device Interaction

21) Can you interact with the device? ______________________________

22) If yes, how? ______________________________

23) Which tasks can you accomplish? ______________________________

24) Which tasks you often require help? ______________________________

25) In which conditions can you operate the device?

Position: ______________________________

Interaction: ______________________________

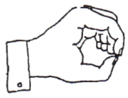
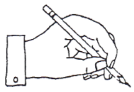
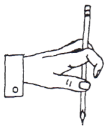
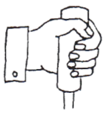
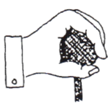
Motion

26) Do you have an electric wheelchair? ______________________________

27) Check all that apply:

- ☐ Sits independently in regular chair
- ☐ Sits in wheelchair without support
- ☐ Sits in wheelchair with support
- ☐ Unable to sit upright
- ☐ Walks independently
- ☐ Walks with assistance
- ☐ Depends on wheelchair for mobility

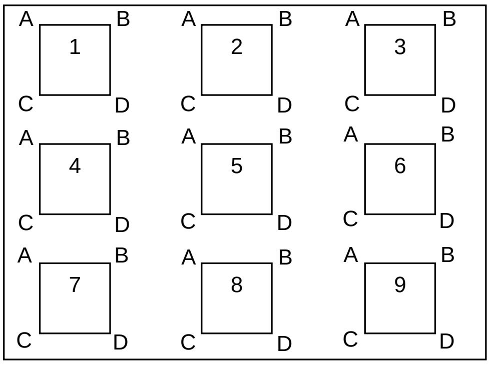
Functional Assessment Evaluation

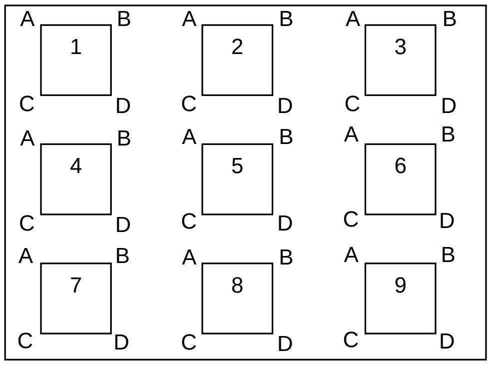
A. Grasps			R	L
28) Tip		Grasp		
		Release		
29) Palmar		Grasp		
		Release		
30) Two-finger		Grasp		
		Release		
31) Lateral		Grasp		
		Release		
32) Cylindrical		Grasp		
		Release		
33) Tee		Grasp		
		Release		
34) Spherical		Grasp		
		Release		
35) Press		Grasp		
		Release		

0 – Unable; 1 - Poor; 2 - Fair; 3 – Good

B. Range

The targets are the corners. Use the sequence: touch the square then 1A/1B/1C/1D and repeat for all squares within the person's range.





	Left Hand	Right Hand
Furthest reach:	______________	______________
Closest reach:	______________	______________
Max. left reach:	______________	______________
Max. right reach:	______________	______________
Preference for location	______________	______________

Interview #____

Consent Form

Thank you for participating in our research studies. These studies aim to understand how people interact with mobile touchscreens. You will be asked to fulfill predetermined tasks commonly performed with a mobile device. We will log the interactions performed with the system as well as their timings. In addition, we will be videotaping your session to allow further analysis. All the results collected from the evaluation will be anonymous. Please read the statements below and sign where indicated, if you agree. Each statement is separated so you are able to disagree with any independent one. Thank you.

1) I agree to perform the evaluation session described and that the results are logged in text files.

Print Name: ______________________________

Signature: ______________________________

Date: ______________________________

2) I agree that the session is videotaped for further analysis by the researchers.

Print Name: ______________________________

Signature: ______________________________

Date: ______________________________

3) I agree that the multimedia (photos and video) captured during the evaluation may be used by the researchers in dissemination events and publications (thesis, articles, press).

Print Name: ______________________________

Signature: ______________________________

Date: ______________________________

Monitor Checklists

In this section we present the checklists used during the evaluation sessions.

Checklist 0 – The day Before the Evaluation

- Print the questionnaires (pre-questionnaire), consent form and evaluation notes
- Print the task descriptions
- Print the evaluation sheets and checklists
- Install and configure on the laptop the software to capture the image from the video camera
- Test the video camera and ensure that the disk space is enough
- Fully charge the camera and mobile device
- Perform one test

Checklist 1 – Day of the Evaluation

Introducing the Session

- Greet the participant
- Give an overview of the evaluation session
- Explain the goal of the evaluation (to test the system)
- Have the participant complete the background questionnaire
- Have the participant do simple tasks to get to know their limitations
- Explain the logging and have the participant complete the consent form

Before the Evaluation

- Create the recording setup
- Ask the user to stay in a comfortable position
- Turn on the mobile device
- Start the mobile application
- Be sure that there are no distractions
- Start video and audio recording

Before each Task

- Ask the user if he/she is comfortable
- Ask the user to adjust the comfort position
- Explain the task and ask the user to perform the required actions

During the Task

- Annotate questions, opinions, interruptions and other notes
- Try to keep the user relaxed

After the evaluation

- Stop the recording
- Debrief the participant

- Thank the participant for his availability, opinions and help
- Clean up and pack the camera and mobile device

Checklist 2 – Day after the Evaluation

- Review the evaluation logs and recordings
- Insert the gathered data in raw tables

Checklist 3 – After all evaluation sessions

- Write the report
- Contact the participants and thank them again for their help

Motor-Impaired Participants Profile

Participant	Level of Cervical Lesion	Age	Gender
1	C4 – C5	28	M
2	C5	28	M
3	C5 – C6	29	F
4	C5	28	F
5	C5	61	M
6	C4 – C5	30	M
7	C4	34	M
8	C6 – C7	40	M
9	C4 – C5 – C6	61	M
10	C5 – C6	42	M
11	C5 – C6	58	M
12	C4 - C5	58	M
13	C5 - C6 - C7	44	M
14	C5	64	M
15	C4 - C5	27	M

Functional Assessment Results

	Tip	Palmar	Two-finger	Lateral	Cylindrica	Tee	Spherical	Press	Preferred Hand
1	0	0	1	1	0	0	0	3	L
2	0	0	3	3	2	1	1	2	L
3	0	0	3	3	3	3	3	3	R
4	0	3	3	3	3	3	3	3	R
5	3	3	3	3	3	3	3	3	L
6	0	0	1	1	0	0	0	1	R
7	3	0	3	1	2	0	3	3	R
8	0	3	3	3	3	3	3	3	R
9	3	3	3	3	3	3	3	3	R
10	0	1	3	3	3	3	3	3	R
11	0	3	3	3	3	3	3	3	R
12	0	0	0	1	0	0	0	2	L
13	0	3	3	3	2	2	1	3	L
14	0	1	2	2	2	2	3	3	R
15	0	0	0	1	0	0	0	3	L

Motor-Impaired Participants Profile

Participant	Level of Cervical Lesion	Age	Gender
1	C4 – C5	28	M
2	C5	28	M
3	C5 – C6	29	F
4	C5	28	F
5	C5	61	M
6	C4 – C5	30	M
7	C4	34	M
8	C6 – C7	40	M
9	C4 – C5 – C6	61	M
10	C5 – C6	42	M
11	C5 – C6	58	M
12	C4 - C5	58	M
13	C5 - C6 - C7	44	M
14	C5	64	M
15	C4 - C5	27	M

Functional Assessment Results

	Tip	Palmar	Two-finger	Lateral	Cylindrica	Tee	Spherical	Press	Preferred Hand
1	0	0	1	1	0	0	0	3	L
2	0	0	3	3	2	1	1	2	L
3	0	0	3	3	3	3	3	3	R
4	0	3	3	3	3	3	3	3	R
5	3	3	3	3	3	3	3	3	L
6	0	0	1	1	0	0	0	1	R
7	3	0	3	1	2	0	3	3	R
8	0	3	3	3	3	3	3	3	R
9	3	3	3	3	3	3	3	3	R
10	0	1	3	3	3	3	3	3	R
11	0	3	3	3	3	3	3	3	R
12	0	0	0	1	0	0	0	2	L
13	0	3	3	3	2	2	1	3	L
14	0	1	2	2	2	2	3	3	R
15	0	0	0	1	0	0	0	3	L

A3
Visual Demands Study Materials

This appendix contains:

- Background questionnaire
- Evaluation monitor checklist
- Evaluation monitor script
- Debriefing questionnaire

Background Questionnaire

Interview #____

1. Gender
   a) Male
   b) Female

2. Age?
   a) < 18
   b) 18 - 34
   c) 35 - 54
   d) > 54

3. Habilitations?
   a) 9th grade
   b) 12th grade
   c) BSc
   d) MSc
   e) PhD

4. Do you perform texto-entry tasks in your mobile phone?
   a) Yes
   b) No

5. What method do you use?
   a) Virtual QWERTY keyboard
   b) Physical QWERTY keyboard
   c) Virtual MultiTap
   d) Physical MultiTap
   e) Other

6. What applications do you use where it is necessary to input text?
   a) Contact managing
   b) SMS
   c) Other: ____

7. How many times do you input texto in your mobile phone?
   a) Several times a day
   b) Daily
   c) Weekly
   d) Monthly
   e) Very rarely

8. Do you usually typed whilst walking?
   a) Yes
   b) No

Interview #____

Consent Form

Thank you for participating in our research studies. These studies aim to understand how people type whilst mobile. You will be asked to copy several sentences in different mobility conditions. We will log the interactions performed with the system as well as their timings. In addition, we will be videotaping your session to allow further analysis. All the results collected from the evaluation will be anonymous. Please read the statements below and sign where indicated, if you agree. Each statement is separated so you are able to disagree with any independent one. Thank you.

1) I agree to perform the evaluation session described and that the results are logged in text files.

Print Name: ____________________

Signature: ____________________

Date: ____________________

2) I agree that the session is videotaped to further analysis by the researchers.

Print Name: ____________________

Signature: ____________________

Date: ____________________

3) I agree that the multimedia (photos and video) captured during the evaluation are used by the researchers in dissemination events and publications (thesis, articles, press).

Print Name: ____________________

Signature: ____________________

Date: ____________________

Evaluation Monitor Checklist

Material:

- Obstacle course
- Device and Text-entry tasks
- Recording device
- Questionnaires

Before each session:

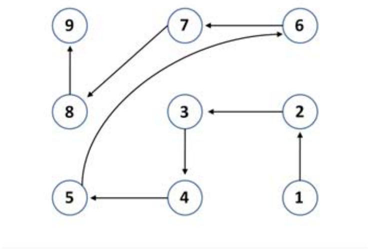


Figure 1 - Percurso

- Build obstacle course (Figure 1)

- Check mobile device and recording device

During each session:

Seated:

- Ask participant to use all text-entry methods

Walking without obstacles:

- First, measure preferred walking speed
- Ask participant to use all text-entry methods

Walking in obstacle course:

- First, measure preferred walking speed
- Ask participant to use all text-entry methods

Evaluation Monitor Script

Boa tarde a todos os presentes e obrigado pela vossa participação neste estudo.

O seguinte estudo destina-se a perceber os níveis de atenção necessários à introdução de texto em dispositivos móveis em diferentes situações e usando diferentes métodos de introdução de texto.

Como situações temos o andar num percurso sem obstáculos, num percurso com obstáculos fixos e um percurso de navegação.

Relativamente aos métodos de introdução de texto irá ser utilizado o qwerty virtual com/sem retorno sonoro e um método alternativo (NavTouch).

Îrá consistir em introduzir texto em telemóveis em três situações: Percurso sem obstáculos, percurso com obstáculos fixos e um percurso com navegação. A ordem pela qual serão executadas estas tarefas será aleatória e calculada por um programa.

Após ser seleccionado um percurso, um método e uma frase, irá ser entregue um telemóvel ao utilizador, este será posicionado no início do percurso respectivo e será dita a frase para decorar. Caso o utilizador se esqueça da frase poderá sempre perguntar novamente ao observador, mesmo durante a execução da tarefa.

Quando o utilizador acabar de escrever a frase deverá indicar ao observador que acabou de a inserir e esperar que este dê indicações acerca do próximo método frase e percurso.

Antes de cada utilizador começar a executar as tarefas do estudo terá 5 minutos para se ambientar ao telemóvel de modo a garantir que sabe manuseá-lo.

(Momento de ambientação)

Para garantir a correcção deste estudo é importante que não restem dúvidas acerca do manuseamento dos aparelhos e do estudo em si. Desta forma alguma dúvida que subsista deverá ser esclarecida agora.

(Momento para tirar dúvidas)

Vamos então iniciar o estudo.

Debriefing Questionnaire

Qwerty	**Qwerty with Audio feedback**
Rate the *cognitive load* of this method	Rate the *cognitive load* of this method
Very Low — Very High	Very Low — Very High
Rate the *physical demand* of this method	Rate the *physical demand* of this method
Very Low — Very High	Very Low — Very High
How was your *pace* during the task?	How was your *pace* during the task?
Very Slow — Very Fast	Very Slow — Very Fast
How accurate were you with this method?	How accurate were you with this method?
Low — High	Low — High
Rate the overall effort to complete the task	Rate the overall effort to complete the task
Low effort — High Effort	Low Effort — High Effort
Rate the level of insecurity felt using this method	Rate the level of insecurity felt using this method
Very Low — Very High	Very Low — Very High

NavTouch

Rate the *cognitive load* of this method

Very Low | Very High

Rate the *physical demand* of this method

Very Low | Very High

How was your *pace* during the task?

Very Slow | Very Fast

How accurate were you with this method?

Low | High

Rate the overall effort to complete the task

Low effort | High Effort

Rate the level of insecurity felt using this method

Very Low | Very High

What were the major challenges while typing in the obstacle course?

Which one of the three methods was the easiest to use, and why?

Which one of the three methods was the hardest to use, and why?

A4
SIID Study Materials

This appendix contains:

- Background questionnaire
- Debriefing questionnaire
- Evaluation monitor script

Background Questionnaire

The main goal of this work is to assess the effect of mobility on typing performance. This questionnaire aims to gather some demographic data about our participants. Data will not be revealed to a third party.

1. Age: _____

2. Gender: ___ Male ____ Female

3. Dominant hand: ___ Right ____ Left

4. How often do you input text with your mobile device?

___ Never ___Rarely ___ Weekly ___ Daily

5. Have you ever used a mobile touchscreen device?

___ Never used ___ Rarely use ___ Weekly ___ Daily

6. For how long do you use touchscreen mobile devices? _____

7. How often do you type in your touchscreen device?

___ Never ___Rarely ___ Weekly ___ Daily

8. What keyboard do you use?

___ Do not know ___QWERTY ___ MultiTap ___ Other: ____________

9. Do you use word prediction or correction? ___ Yes ___ No

10. Rate you level of expertise in typing tasks:

Very bad 1 - - - - - - - - 2 - - - - - - - - 3 - - - - - - - - 4 - - - - - - - - 5 Very good

Debriefing Questionnaire

1. What is your preferred hand posture? Why?

2. Which hand posture is the hardest to type? Why?

3. What were the main challenges of typing whilst walking?

4. Any other comments about the experiment?

Evaluation Monitor Script

Antes de mais obrigado por participar neste estudo. O objectivo deste trabalho é verificar qual o impacto do andar na escrita de texto em dispositivos móveis. Nos próximos 40 minutos irá escrever algumas frases neste telemóvel em 3 condições de mobilidade: sentado e a andar a duas velocidades diferentes. Irá também escrever de várias formas, com uma e duas mãos.

Agora vou-lhe dar alguns minutos para preencher este pequeno questionário. ***Administrar questionário.***

Alguma questão até agora?

O teclado que irá usar é um teclado tradicional em ecrãs tácteis. Aparecer-lhe-á uma frase no topo do ecrã e o objectivo é copiar esta frase o mais precisa e rapidamente possível. Atenção que o que estamos aqui a avaliar é o método de introdução de texto e como este se adequa aos utilizadores e nunca a pessoa em questão. Nesta tarefa não irá poder corrigir o que escreve, ou seja, caso se engane passe para a letra seguinte sem se preocupar em apagar.

Ao iniciar o teste irá aparecer a frase e quando achar que terminou pressione a tecla seguinte. Terá de transcrever 5 frases diferentes. Ser-lhe-á também pedido para segurar o dispositivo de uma determinada forma (uma ou duas mãos), mas irá sempre interagir com os polegares. Antes de iniciar o teste terá um período de treino de duas frases.

Alguma questão?

Antes de iniciar gostaria ainda de saber se podemos recolher vídeos e imagens do teste para análise futura.

Podemos então começar, iniciando a aplicação. Irá ser-lhe transmitida a condição de mobilidade em que irá efectuar o teste e terá um período de ambientação e após ser-lhe-á dito como segurar o dispositivo para a escrita.

Antes de iniciar as condições de mobilidade

Nesta situação o objectivo é seguir a pessoa que está à sua frente mantendo sempre o mesmo ritmo de deslocamento. Para além de ter de manter esta velocidade de andar, terá de copiar as frases que lhe vai aparecendo o mais precisa e rapidamente possível. Atenção, não se esqueça de manter a velocidade de deslocamento e seguir o monitor de teste.

Questões?

Após cada condição de teste

Recolher comentários: Principais dificuldades (comparação com outras situações e postura de mãos)

No final do teste

Aplicar questionário final

A5
HIID Study Materials

This appendix contains:

- Background questionnaire
- UPDRS questionnaire
- Archimedes spiral exercise
- Debriefing questionnaire
- Evaluation monitor script

Background Questionnaire

The main goal of this Project is to assess people's main difficulties when typing in touchscreen devices. This questionnaire aims to gather some demographic data about our participants.

1. Age: ___

2. Gender: ___

3. Dominant hand: ___

4. Do you have any visual impairment? ___ Yes ___ No

Observations:

____

____

5. Do you have any hearing impairment? ___ Yes ___ No

Observations:

____

____

6. Do you have any tremor impairment? ___ Yes ___ No

Observations:

____

____

7. Have you ever used a personal computer? ___ Yes ___ No

8. For how long do you use personal computers? ___

9. How frequently do you use physical keyboards?

___ Never used ___ Rarely ___ Weekly ___ Daily

10. Rate your expertise level:

Very bad 1 - - - - - - - - 2 - - - - - - - - 3 - - - - - - - - 4 - - - - - - - - 5 Very good

11. Do you own a mobile phone? ___ Yes ___ No

12. For how long do you use mobile phones? ___

13. What do you do with your mobile phone?

____ Make calls ___ Receive calls ___ Manage contacts

___ Send SMS's ___ Read SMS's ___ Internet browsing

___ Other: ____

14. How frequently do you input text in your mobile phone?

___ Never ___Rarely ___ Weekly ___ Daily

15. Rate your expertise level typing in your phone:

Very bad 1 - - - - - - - - 2 - - - - - - - - 3 - - - - - - - - 4 - - - - - - - - 5 Very good

16. How often do you use touchscreen mobile devices?

___ Never used ___ Rarely ___ Weekly ___ Daily

17. For how long do you use touchscreen mobile devices? ___

18. How frequently do you input text in your touchscreen mobile device?

___ Never ___Rarely ___ Weekly ___ Daily

19. What keyboard do you use?

___ Do not know ___QWERTY ___ MultiTap ___ Other: ____________

20. Do you use word prediction or correction? ___ Yes ___ No

21. Rate you level of expertise in typing tasks:

Very bad 1 - - - - - - - - 2 - - - - - - - - 3 - - - - - - - - 4 - - - - - - - - 5 Very good

Tremor Assessment: UPDRS Questionnaire

1. Place both hand on the table. *(rest tremor)*

Right hand:
0 – Absent.
1 – Slight and infrequently present
2 – Mild in amplitude and persistent. Or moderate in amplitude, but only intermittently present.
3 – Moderate in amplitude and present most of the time.
4 – Marked in amplitude and present most of the time.

Left hand:
0 – Absent.
1 – Slight and infrequently present
2 – Mild in amplitude and persistent. Or moderate in amplitude, but only intermittently present.
3 – Moderate in amplitude and present most of the time.
4 – Marked in amplitude and present most of the time.

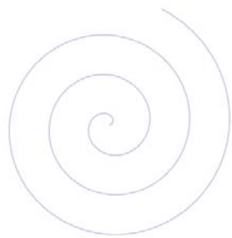
2. Touch your nose with your index finger. *(action – intention - tremor)*

Right hand:
0 – Absent.
1 – Slight; present with action.
2 – Moderate in amplitude, present with action.
3 – Moderate in amplitude with posture holding as well as action
4 – Marked in amplitude; interferes with feeding.

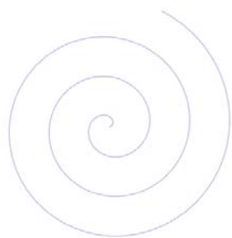
Left hand:
0 – Absent.
1 – Slight; present with action.
2 – Moderate in amplitude, present with action.
3 – Moderate in amplitude with posture holding as well as action
4 – Marked in amplitude; interferes with feeding.

Draw a spiral, with each hand, without leaning the hand or arm on the table

Right Hand:



Left Hand:



Debriefing Questionnaire

1. Rate the ease of use when typing with the mobile phone:

Very easy 1 - - - - - - - - 2 - - - - - - - - 3 - - - - - - - - 4 - - - - - - - - 5 Very hard

2. What were the main challenges?

____
____
____

3. Rate the ease of use when typing with the tablet:

Very easy 1 - - - - - - - - 2 - - - - - - - - 3 - - - - - - - - 4 - - - - - - - - 5 Very hard

4. What were the main challenges?

____
____
____

5. In which one of the devices do you prefer to type?

___ Mobile ___ Tablet ___ Do not know

6. Other comments:

____
____
____
____

Evaluation Monitor Script

Fase de Familiarização

Antes de mais obrigado por aceitar participar neste trabalho. O meu nome é Hugo, eu sou aluno no Instituto Superior Técnico, em Lisboa, e estou neste momento a fazer o meu Doutoramento. O objectivo do meu trabalho é tornar os telemóveis mais fáceis de usar, principalmente estes mais recentes que não têm quaisquer botões. Para atingir este meu objectivo, preciso de primeiro perceber quais as dificuldades que existem e quais os problemas com os telemóveis actuais.

Até agora, alguma questão?

Vou-lhe agora fazer umas perguntas simples. ***Administrar questionário.***

O que vamos fazer é brincar um pouco com estes dispositivos e vamos realizar uma tarefa muito específica: a escrita de texto.

Iniciar aplicação "practice" e mostrar teclado/caixa de texto.

Explicar e exemplificar (tap, on release)

- Jogo pressionar teclas e construir frases (5 minutos por dispositivo); experimentar as diferentes posturas
- Jogo transcrever palavras (5 minutos por dispositivo);

Fase de Teste

Boa tarde, hoje vamos começar com uns testes muito simples. Vou-lhe agora fazer algumas perguntas:

- Recolha de tremor subjectiva
- Teste espiral
- Acelerómetro

Agora vamos voltar a escrever texto. O que lhe vou pedir para fazer é copiar algumas frases o mais rapidamente e precisamente possível. Vai usar estes dois aparelhos, tanto na posição vertical como horizontal. No caso do telemóvel terá de o segurar na mão.

Alguma questão?

Antes de iniciar gostaria ainda de saber se podemos recolher vídeos e imagens do teste para análise futura.

Explicar procedimento na fase de treino (5 minutos com cada condição) – practice application. Agora ainda estamos em treino. Terá de copiar a frase que está no topo o mais rapidamente e precisamente que conseguir. Caso se engane em alguma letra, passa para a próxima. Como não existe o botão de apagar, não tenta corrigir, passa para a próxima letra. Quando chegar ao fim da frase passa para a próxima, pressionando o botão de avançar.

Alguma questão?

<ver ordem do teste>

Podemos então começar, iniciando a aplicação. As duas primeiras frases são sempre de treino e não contam.

Alguma questão?

Após cada dispositivo

Recolher comentários: Principais dificuldades (comparação entre tamanhos e com outro dispositivo)

No final do teste - Administrar questionário de debriefing.