UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE INFORMÁTICA

BRAILLESHAPES: Efficient Text Input On Smartwatches For Blind People

Pedro Jorge Amado Marques

Mestrado em Engenharia Informática

Dissertação orientada por:
Professor Doutor Tiago João Vieira Guerreiro e co-orientada Professor Doutor João Pedro Vieira Guerreiro

2022

Acknowledgments

To every single person that is or was in my life thus far, thank you. Thank you for being a part of this long-lasting journey and in some way contributing to me getting to where I am today. With that being said, there are a few people to whom I owe a special thanks.

First and foremost, I would like to thank my supervisors Professor Tiago Guerreiro and Professor João Guerreiro. Thank you for your time, your guidance, your overall support, and for believing in me and trusting me with this project. But most importantly, thank you for allowing me to have such a rewarding and life changing experience.

To the Faculdade de Ciências da Universidade de Lisboa and its department of informatics, thank you for the conditions provided which allowed me to accomplish this work. I would also like to thank LASIGE and all the people that are a part of it, for the welcoming and thriving environment, and the opportunity to know some amazing people.

To the Fundação Raquel e Martin Sain, it was an enormous pleasure to work with you and to be given the opportunity to increase my perception of new realities. A special thanks to every person attending the foundation that either participated in the study or I had the opportunity to meet and interact with. Thank you for the unforgettable experience.

Atenção, atenção, eu quero ainda agradecer to all my fellow colleagues that accompanied me throughout my academic course and with whom I spent sleepless nights working, specially to the 15/16 with which I spent some of the most amazing years of my life. Years I will never forget.

To Mariana Henriques, thank you for everything, no matter what.

I want to thank Diogo Martins, João Serpa and André Pereira, for always being at my corner, or at the bar with me. Because *"Whatever you do in this life, it's not legendary unless your friends are there to see it"*, and I wouldn't want anyone else beside me.

To the Power Rangers of Alcobaça, what a journey. For all this time, I know you know, but I'm forever thankful.

To my parents, this one is for you. Thank you for making me the person I am today. Thank you for giving me this and every other opportunity you ever gave me. Thank you for allowing me to grow the way I did, and become the person I am today. No words will ever be enough.

To myself, *we made it*. Thank you.

Seek Discomfort

Resumo

Os dispositivos móveis com ecrã tátil, como smartphones ou smartwatches, são um aspeto predominante nas nossas vidas. Eles foram evoluindo ao longo do tempo bem como as suas funcionalidades, e devido ao constante crescimento e avanços na tecnologia, usar estes dispositivos para realizar uma grande diversidade de tarefas tornou-se uma prática comum no nosso dia a dia.

No entanto, a sua dependência de interações à base de toques, a exigência de boa habilidade espacial e de boa memorização inerentes a tais dispositivos, e a falta de indicações táteis suficientes, faz com que este tipo de dispositivos seja visualmente exigente, acabando assim por proporcionar uma experiência de interação extenuante para pessoas com insuficiência visual. Isto é algo que se torna ainda mais evidente em cenários que ocorrem em contextos à base de movimento ou onde o uso de uma só mão é enfaticamente necessário.

Mesmo que os leitores de ecrã atuais e outras tecnologias existentes já comuniquem alguns avanços em relação a questões de acessibilidade direcionada para pessoas com deficiência visual, ainda existem diversas interações que requerem soluções mais sofisticadas de modo a tornarem-se mais acessíveis, como por exemplo inserção de texto ou navegação em menus. Uma das opções existentes bastante usada para inserir texto é através do reconhecimento de voz. Contudo, esta modalidade nem sempre é a mais apropriada podendo inclusive levar a resultados diferentes dos pretendidos, ou acaba por levantar questões sociais e de privacidade.

Sendo a comunicação de extrema importância e intrínseca ao ser humano, uma tarefa, em particular, para a qual é imperativo fornecer soluções que abordem as preocupações relacionadas com acessibilidade circundantes é a inserção de texto.

Acreditamos que dispositivos tais como smartwatches podem oferecer inúmeras vantagens na abordagem dos temas mencionados. No entanto, ao fazermos uma revisão sobre estudos realizados na área, foi possível perceber que estes carecem de soluções focadas em acessibilidade para a realização de diversas tarefas, sendo que a maioria das opções existentes para dispositivos móveis com ecrãs táteis é desenhada para smartphones. A falta de soluções apropriadas para esta modalidade de interação é um indício da falta de padronização existente para este tipo de dispositivo, e que acaba por contribuir para a falta de inclusão de pessoas cegas aos mesmos.

Sendo o Braille um padrão de leitura bem estabelecido para pessoas cegas e tendo mostrado resultados positivos quando usado em trabalhos anteriores relacionados com abordagens acessíveis de inserção de texto, acreditamos que usá-lo como base para uma solução de entrada de texto acessível pode ajudar a solidificar um padrão para este tipo de modalidade de interação. Acreditamos ainda que pode permitir os possíveis utilizadores tirarem proveito do seu conhecimento prévio de Braille, reduzindo dessa forma qualquer carga cognitiva extra. Através da revisão do estado da arte realizada no início da tese, foi nos possível identificar quais as abordagens que permitiram obter os melhores resultados, podendo assim destacar um artigo em particular [9]. A abordagem explorada neste artigo, denominada BrailleTouch, consiste numa abordagem à base de múltiplos toques em simultâneo num teclado virtual tátil mapeado à semelhança de uma célula Braille. Atualmente, soluções semelhantes podem ser encontradas embutidas diretamente em smartphones fazendo parte das opções de acessibilidade. No entanto, embora soluções como esta que requerem interações com múltiplos alvos em simultâneo baseadas em Braille tenham alcançado bons resultados, devido ao espaço reduzido do ecrã tátil do smartwatch, uma abordagem de toque não é a mais viável. Desta forma, decidimos que uma opção mais exequível passa por recorrer a uma solução à base de gestos.

Assim, foi possível delinear o objetivo desta tese como sendo a exploração do conceito de formas baseadas em Braille, e a validação da viabilidade do seu uso como base para um método acessível de entrada de texto para smartwatch baseado em gestos, direcionado para pessoas com deficiência visual. Apesar do foco deste método de escrita ser em smartwatches, esta abordagem foi pensada com o intuito de ser facilmente traduzível para outros formatos.

De forma a alcançar o nosso objetivo, começámos por definir o conceito de formas baseadas em Braille, o qual denominámos de Braille Shapes, bem como conceitos adjacentes identificados como relevantes. A definição alcançada descreve uma Braille Shape como uma forma obtida realizando um único traço, passando por todos os pontos destacados de uma determinada célula Braille sem passar mais do que uma vez por nenhum desses pontos. Após conseguida esta definição, foi realizado um estudo com utilizadores cegos e normovisuais para recolha de Braille Shapes e subsequente análise. Este primeiro estudo foi concretizado com o intuito de obter uma melhor compreensão de como estas eram assimiladas e concebidas pelos participantes, bem como avaliar a sua aceitação por parte dos mesmos.

Sendo que o desenvolvimento de um sistema à base da conceção de formas livres requer o reconhecimento das mesmas, as Braille Shapes recolhidas no primeiro estudo foram ainda utilizadas para avaliação de diferentes mecanismos de reconhecimento. Mecanismos estes selecionados com base na revisão de trabalhos anteriores realizados na área. Explorámos duas abordagens em específico, uma baseada na correspondência de modelos e outra no reconhecimento de imagens. Para ambas, diferentes técnicas e reco-

nhecedores foram ainda testados. Com base nesta análise, pudemos concluir que analogamente ao trabalho revisto previamente, os mecanismos de reconhecimento testados não produziram resultados aceitáveis para posterior validação em cenários do mundo real.

Obtendo feedback positivo por parte dos participantes relativamente à ideia a ser explorada e vendo uma rápida assimilação de conceitos, procedemos para uma validação do uso de Braille Shapes num contexto de inserção de texto. Para tal, foi realizado um segundo estudo com utilizadores para recolha de gestos, desta vez num contexto mais real. Neste caso, tendo os mecanismos de reconhecimento testados produzido resultados aquém do aceitável para tal validação, realizámos o que considerámos ser uma validação preliminar em que nenhum tipo de reconhecimento foi utilizado aquando da recolha, somente na análise posterior. Neste estudo foi pedido aos participantes que inserissem determinadas frases utilizando para tal Braille Shapes, sendo lhes dadas ainda algumas funcionalidades de edição de texto de forma a que inserissem espaços, apagassem caracteres, ou confirmassem a inserção de frases quando terminassem.

Terminando este estudo e analisando os resultados obtidos, foi possível observar um forte interesse e uma grande adesão por parte dos participantes à abordagem explorada, com os mesmos a indicar que a nossa solução abordava os seus aspetos de maior preocupação. Apesar das métricas avaliadas relativas à eficiência e eficácia de um sistema usando esta abordagem terem ficado aquém das espectativas, este mostrou proporcionar uma experiência ágil e de fácil habituação, com os participantes a melhorarem os resultados alcançados quanto mais tempo ficavam expostos ao método a ser estudado.

Estas conclusões permitem-nos acreditar no potencial de um método de escrita aplicando esta abordagem, e levam-nos a perceber que este não só é alcançável como é também desejável, podendo trazer inúmeras vantagens para as pessoas cegas. Mesmo que a abordagem explorada não seja utilizada para implementação de um método de escrita como inicialmente previsto, devido à sua flexibilidade de implementação é possível idealizar-mos diversas modalidades para a qual uma abordagem deste tipo possa ser aplicada, como por exemplo para atalhos de navegação. Em todo o caso, é do nosso entender que este trabalho seja um contributo para uma área de investigação de grande interesse que precisava de ser abordada.

Palavras-chave: Não-visual, entrada de texto, Braille, ecrã tátil, reconhecimento de gestos

Abstract

Mobile touchscreen devices like smartphones or smartwatches are a predominant part of our lives. They have evolved, and so have their applications. Due to the constant growth and advancements in technology, using such devices as a means to accomplish a vast amount of tasks has become common practice.

Nonetheless, relying on touch-based interactions, requiring good spatial ability and memorization inherent to mobile devices, and lacking sufficient tactile cues, makes these devices visually demanding, thus providing a strenuous interaction modality for visually impaired people. In scenarios occurring in movement-based contexts or where one-handed use is required, it is even more apparent.

We believe devices like smartwatches can provide numerous advantages when addressing such topics. However, they lack accessible solutions for several tasks, with most of the existing ones for mobile touchscreen devices targeting smartphones. With communication being of the utmost importance and intrinsic to humankind, one task, in particular, for which it is imperative to provide solutions addressing its surrounding accessibility concerns is text entry.

Since Braille is a reading standard for blind people and provided positive results in prior work regarding accessible text entry approaches, we believe using it as the basis for an accessible text entry solution can help solidify a standardization for this type of interaction modality. It can also allow users to leverage previous knowledge, reducing possible extra cognitive load. Yet, even though Braille-based chording solutions achieved good results, due to the reduced space of the smartwatch's touchscreen, a tapping approach is not the most feasible. Hence, we found the best option to be a gesture-based solution.

Therefore, with this thesis, we explored and validated the concept and feasibility of Braille-based shapes as the foundation for an accessible gesture-based smartwatch text entry method for visually impaired people.

Keywords: Non-visual, text entry, Braille, touchscreen, gesture recognition

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Over the years, advancements in technology propelled the usage of mobile touchscreen devices to perform mundane tasks. These devices, such as smartphones and smartwatches, are already a huge and ever-growing part of our lives. They allow us to perform numerous actions in different contexts, relying mostly on touch-based interactions. Nonetheless, these devices still need adaptations to be viable in some scenarios and for specific users.

1.1 Motivation

Mobile devices applications have extended beyond basic communications, but although they offer a wide variety of functionalities, interacting with them is usually visually demanding. This, alongside the lack of tactile cues, the need for good spatial ability and good memorization, can be hindering factors for visually impaired people when trying to get the most out of these devices.

Even though screen-readers and other assistive technologies already communicate advancements regarding accessibility for visually impaired people, there are still several interactions that require more sophisticated solutions to become more accessible, such as text entry or menu navigation. Since communication is such an essential aspect of our lives, and text entry is one of the most common and visually demanding tasks, there is an imperative need to provide solutions for this issue.

Blind people have the option to input text via speech recognition, but this modality may not always be appropriate or raise privacy and social concerns [16]. There are already multiple other approaches focused on improving text entry's accessibility, whether built-in on existing devices like Apple's VoiceOver[1], or others like the ones described in section 2, however, most of these usually target a specific type of devices – smartphones – and are not suitable for scenarios requiring single-hand usage or movement-based contexts.

Even if the use of smartwatches facilitates interactions in those cases, the vast diversity of existing text input methods for such devices, not only is normally not suited for visually

[1] https://www.apple.com/accessibility/vision/ (last visited 02/11/2021)

impaired people but is also a representation of the lack of standardization for this type of device [19].

With Braille being a reading standard for blind people, implementing a text entry solution based on the Braille alphabet offers a solution to the standardization issue while allowing blind users to leverage previous Braille knowledge. The usage of Braille in such techniques [9, 33, 31, 16, 36] has also been shown to provide efficient results when used in Braille chording approaches, but even so, with the reduced space of the smartwatch's touchscreen, an approach like that would not be the most feasible.

1.2 Goals

With this in mind, our end goal for this thesis is to explore the concept of Braille-based shapes and their feasibility as the basis for a gesture-based smartwatch text entry method for visually impaired people.

We aim to explore how people perceive and perform *Braille Shapes*, specifically visually impaired people, understanding their consistency and ease in doing so. We also aim to subsequently validate its usage as the foundation for a text entry method by conceiving and evaluating such a system. This solution would allow leveraging from all of Braille's advantages and could be used single-handed and "on the go" in a small screen device.

Even though the focus is on smaller-sized devices, we believe this approach will be easily translatable to other form factors.

1.3 Contributions

To accomplish this, we were required to undergo several stages which, after carrying out this work, are expected to provide contributions.

1. Our first contribution is a detailed literature review in section 2 regarding text entry methods for mobile devices, focusing on touchscreen devices and approaches relying on Braille.

2. Our second contribution is the conceptualization of Braille-based shapes and how people perceive and perform them. In sections 3.4 and 4.2, we provide an overview of shapes performed by sighted and visually impaired people and a comparative analysis between the shapes of the two groups. Our analysis also highlights some nuances to account for in future developments of systems based on this concept.

3. Section 5 provides our third contribution, a comparative analysis of various recognition approaches evaluated on the shapes collected from the participants of this study. Here we describe the process of training and testing several recognition

mechanisms to help find the most-suited solution required for future implementation of a gesture-based system using *Braille Shapes*.

4. All the shapes collected for this study compose our fourth contribution, a data set of *Braille Shapes* performed by sighted and visually impaired people, which is unique and intended to assist in future similar studies.

5. Lastly, we contribute with a prototype of a Braille-based text entry approach and its subsequent preliminary validation to understand its efficiency, efficacy, and usability characteristics.

1.4 Context

This project is developed in collaboration with LASIGE, a research unit at the Faculdade de Ciências da Unicersidade de Lisboa, in the field of Computer Science and Engineering. LASIGE is organised around six Research Lines of Excellence, namely Digital Accessibility and Systems for the Ageing Population, Cyber-Physical Systems, Data and Systems Intelligence, Health and Biomedical Informatics, Reliable Software Systems, and Resilient Distributed and Networked Systems.

Since one of the main areas of expertise of LASIGE is Digital Accessibility, the realization of this project will benefit it, possibly adding to the foundation already existing.

Due to this project's nature and specific requirements, it was also carried out in collaboration with Raquel e Martin Sain Foundation, a Private Institution of Social Solidarity. This institution has the main purpose of carrying out a work of typhlological education and occupation. More specifically, its objective is to conduct the professional training of the blind, focusing on ensuring future paid work possibilities.

The author of this work started his master's degree in Informatics Engineering, with an emphasis in the field of Human-Computer Interactions as well as Mobile and Ubiquitous Computing, which allowed him to acquire some of the necessary capacities to take on the project in question.

1.5 Structure of the document

This document is organised as follows:

- Chapter 2 - **Related Work:** Provides an overview and discussion of prior work related to the topic in question as well as the state of the art for text entry methods on mobile devices.

- Chapter 3 – **BrailleShapes:** This chapter provides a background overview of the Braille code and Braille literacy, as well as the impact they have on visually impaired people's lives. It comprises the definition of a *BrailleShape*, a description of

the text entry system envisioned, and the framing of certain concepts relevant for a better understanding of this project. It also presents the auxiliary systems devised for the purpose of this study.

- Chapter 4 - **Exploring the Feasibility of BrailleShapes:** In this chapter, the data collection process is detailed alongside subsequent analysis, and are identified the conclusions drawn from its results regarding understanding and performance of BrailleShapes by the participants.
- Chapter 5 - **Automatically Recognizing BrailleShapes:** Provides an evaluation of various recognition approaches tested on the gestures collected.
- Chapter 6 - **Exploring BrailleShapes for a Text Entry Approach:** This chapter details the preliminary validation of a system using the approach in question. It describes the procedure as well as the results and conclusions obtained from a performance evaluation. It also presents the participants opinions on how they perceived and accepted this type of approach.
- Chapter 7 - **Conclusion:** Summarizes and analyzes the contributions of this work, explaining any challenges found and providing directions for future endeavours.

Chapter 2

Related Work

In this chapter, is discussed prior work regarding 3 topics: non-visual text input on mobile devices where we highlight Braille-based approaches, both visual and non-visual text input on smartwatches, and shape and gesture recognition.

2.1 Non-visual text entry on mobile devices

Since the appearance of mobile devices, accessibility in that area has been an evolving concern. Prior to the dominance of mobile touchscreen devices, solutions like NavTap [12] had been developed to improve accessibility regarding text entry on mobile devices. This approach allowed blind users to input text on a mobile device provided with a keypad, by allowing them to navigate through the letters by tapping one of four keys (2, 4, 6 and 8) before selecting the desired letter. The alphabet would be rearranged in a vowel-indexed manner which would eliminate the need for memorization. Later on, with the emergence of mobile touchscreen devices, such an approach was then translated to a touch-based method, where instead of tapping keys, the users would perform directional gestures [11]. Although other systems [41] also used this type of multi-directional approach, Bonner et al. went a step further and provided what they thought to be three key qualities an eyes-free text entry system should have: a robust entry technique, a familiar layout, and painless exploration [3].

With the arrival of Apple's iPhone, interest in touchscreen devices increased and with the release of the iPhone 3GS, VoiceOver[1][2] was added to iOS. This accessibility feature is a screen reader built into the iPhone, that guides the users by providing TTS output for every on-screen item selected by sliding a finger over it. VoiceOver's text entry component works in the same way over a soft QWERTY-based keyboard and requires split-tapping or double-tapping to enter a selected key. This type of QWERTY-based approach presents the user with numerous visual targets, that can induce accuracy errors, especially for users

[1] https://www.apple.com/accessibility/vision/ (last visited 02/11/2021)
[2] https://en.wikipedia.org/wiki/VoiceOver (last visited 02/11/2021)

not familiar with computer keyboards [3].

To avoid the need for multiple operation processes and precise location interactions, Vidal et al. presented a system based on handwriting recognition, named oPhone, that allows the users to compose a dialling number without visual feedback, by simply drawing a number from 0 to 9 one at a time. This system is based on coupling gesture recognition and vocal synthesis [40]. Although oPhone only allowed to input digits, Tinwala et al. provided another handwriting-based approach that combined the single-stroke based alphabet from Graffiti [22] shown in figure 2.1 with audio and vibrotactile feedback, thus allowing the users to input text, a character at a time, with just one finger. The audio and vibrotactile feedback provided were at character-level, and the system required a space (double tap) to enter a word [35].

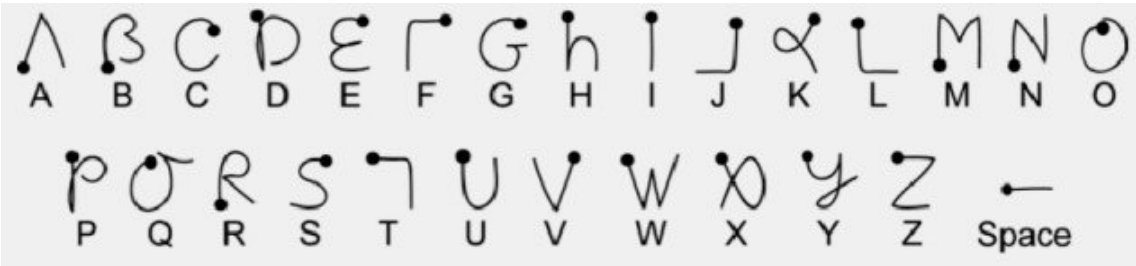


Figure 2.1: The Graffiti alphabet. Image taken from [35].

Physical gestures are usually performed alongside verbal expression and help convey communicative information. Studies have shown that blind people perform gestures while speaking with other blind people, and since smartphones are equipped with sophisticated motion sensors, Dim et al. focused their attention on the usage of motion gestures for blind people's interaction with such devices. They show [6] that motion gestures are not only usable as mobile interactions for blind people but are also well received for certain types of interactions being capable of producing acceptable results.

2.1.1 Braille-based approaches

Since the Braille writing system, developed specifically for blind people communication, is common knowledge for many blind users and allows for reasonably efficient inputting, researchers have also explored Braille-based approaches for mobile touchscreen devices text entry [31, 28, 9, 33, 25, 24, 26, 16]. Oliveira et al. presented BrailleType [28], a single-touch text entry method, that consists of a single screen embodying a Braille cell, composed of 6 targets representing the dots in the Braille matrix. It allowed blind users to input Braille chords, by tapping targets one at a time – not requiring a specific order – to select the corresponding dots of the desired braille chord. To mark a dot in the Braille cell, the user must press the desired target and wait for audible confirmation. Touching a dot already marked would remove it and provide audio feedback. When all the

necessary dots for a Braille character are selected, a double-tap in any part of the screen is required to accept it. A space would be made by entering an empty Braille cell, and a swipe left gesture would clear the Braille cell if any dots were marked, or delete the last entered character if the matrix was empty. BrailleTouch [9] consists of a similar approach although it includes an extra soft button for the spacebar and allows for multi-touch input just like a traditional Braille typewriter. To help the users with finger placement, the system also comprised of a case for the device used as shown in figure 2.2.

(a) BrailleTouch's input surface facing away from the user

(b) BrailleTouch's back facing the user

Figure 2.2: BrailleTouch on a smartphone. Images taken from [9].

In their study [33], Siqueira et al. evaluated the performance of BrailleTouch for three groups of participants A, B and C, with different levels of experience, ranging from expert users to participants with poor performance respectively. Participants from groups A and B achieved an average of 23.2 WPM (words per minute) and 21 WPM respectively, while participants from group C only achieved speeds of 9.4 WPM. Even though the majority of the groups showed good results, only group A exhibited error rates (14.5%) suitable enough for real-world use. Although multi-touch approaches are faster than single-touch ones, they are also more error-prone [25, 28]. Nicolau et al. try to mitigate this issue by providing an approach for a multi-touch Braille input correction system that uses chords as an atomic unit of information instead of characters, meaning it looks at the chord itself and uses its information to correct its character counterpart [25]. This chord-based spellchecker was shown to be consistently more accurate at word-level correction than Android's (AOSP) spellchecker, which suggested that leveraging chord information plays a major effect on correction accuracy. This study also showed that combined character-word-level correction was ineffective.

In the same way that audio isn't always the most suitable modality for text input, it may also not always be the most suitable for output. Nicolau et al. presented an approach where a Braille-based vibrotactile reading device is used as an output/reading method [24]. This method enables blind users to read textual information, by leveraging the users' Braille knowledge, since it draws inspiration from the standard writing system of the Perkins Brailler. The device consists of six vibrotactile actuators, that are used to code a Braille cell and communicate single characters, through simultaneous vibrotactile

feedback from such actuators. The actuators are placed in the index, middle and ring fingers of both hands, and each actuator represents one dot of the Braille cell, or character. Overall, results showed that recognition rates are usually high with the right stimuli and interval duration, even though some characters were more troublesome to identify. This indicates potential for the use of vibrotactile feedback on mobile devices.

Understanding the potential of vibrotactile feedback and taking advantage of braille multi-touch approaches' high input speeds [9], Nicolau et al. presented Holibraille [26], a system that enables Braille input and output on mobile devices. The input method used is based on BrailleTouch's approach, hence, the major contribution being the output mechanism. It consists of six vibration motors on the top and bottom of the device (on a custom-made case), with each actuator representing a dot of the braille cell. This allows for localized feedback regarding the outputted information. Although this system was only evaluated on character discrimination and had some limitations regarding the implementation of the physical actuators, it showed improvements in recognition speed in comparison to other mobile Braille feedback solutions. It also suggested that vibrotactile feedback should be carefully designed in order to mitigate different types of errors, either omissions or insertions, according to the fingers that are being actuated.

Even with solutions for input error correction [25], some typing errors still remain and need to be amended by the users. Trindade et al. presented an approach that allows Braille text entry and edit through the use of both physical and gestural interactions on a mobile device [36]. To do so, they developed a system consisting of a physical Braille keyboard built-in on a custom case as well as an interface that provides editing controls - caret movement, text selection and clipboard operations - via touchscreen gesture interaction. The keyboard draws inspiration from the traditional Braille writing system that contains seven keys, although it has extra buttons to enter different input operations and modes such as edit mode, backspace, and space. The addition of these buttons also allows for the device to be used both in portrait and landscape mode. Although this system didn't show significant improvements regarding input speed in comparison to another touch-based system, it showed to be more accurate and to allow better error correction.

Other studies [40, 35] have already provided gesture-based solutions for blind text input, however, these approaches have either limitations or require the learning of a new alphabet. Li et al. presented BrailleSketch [16], an approach to a gesture-based text input method for mobile devices that leverages previous Braille knowledge. This method allows for users to input a letter by sketching a gesture, or more specifically, by drawing a path that passes by the corresponding dots of the Braille code as seen in figure 2.3. The gesture can be performed from anywhere on the screen and the Braille code can be drawn in many ways. This solution utilizing an auto-correction method produced better results regarding input speed than others mentioned in the study except for BrailleTouch in certain conditions. It achieved average input speeds of 14.53 WPM without signs of

plateauing, and an average error rate of 11%. It also proved that word-level feedback instead of character level can be advantageous, and all this combined, makes the system potentially better than most tapping and typing methods, for visually impaired users with Braille knowledge.

All these different existing solutions for mobile text entry accessibility not only introduce more demanding systems from a skillset point of view but also disregard the differences amongst blind people [27]. Oliveira et al. propose to identify and quantify individual characteristics that make a difference in a blind user's experience when using a mobile touchscreen device, more specifically in scenarios with mobile touch-based text entry. In their study [27], they conclude that age and age of blindness onset have a great impact on overall abilities and that tactile sensibility, spatial ability and verbal IQ affect the ability of blind users to perform touch-based interactions. They also mention that different system designs suit different people better and that this should be taken into consideration in order to provide more inclusive solutions.

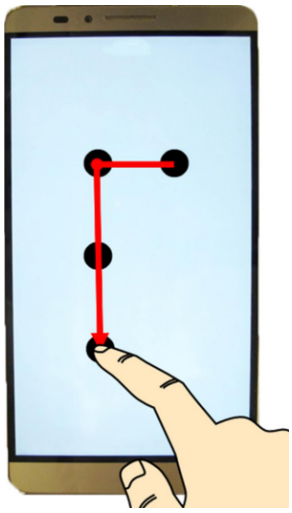


Figure 2.3: BrailleSketch on a smartphone. Images taken from [16].

2.2 Text entry on smartwatches

Smartwatches have been emerging and are becoming a bigger presence in consumers' lives. They allow for quicker and more practical interactions and can ease scenarios where only one hand is available. Since blind people are usually guided with the aid of a cane or a guide dog, or can simply be carrying something with one hand making it unavailable, smartwatches are easily a powerful tool for people with visual disabilities. However, this means accessibility-related issues present in other mobile devices have emerged in this type of device as well. One task, in particular, that is not fully covered yet is text entry [19].

Yi et al. proposed a non-touch bezel-based rotational approach for text entry on smartwatches [42]. To do so, they presented a system comprised of multiple cursors on a circular keyboard, dynamically positioned to optimize rotational distance. The system allowed the users to enter text by selecting letters with the nearby cursors positioned, by the rotation of the smartwatch bezel. Continuous gesture recognition was also used as an approach for smartwatch text entry. Nascimento et al. presented a gesture-based system, that allowed the input of letters with at most two interactions, without requiring the full drawing of the letter [23]. The system relies on an incremental gesture recognition algorithm developed in another study, as well as on a Naïve Bayes classifier for gesture classification.

Luna et al. show [19] that there is a large range and diversity of text entry methods for smartwatches, where it is possible to see substantial potential as well as some limitations. However, this diversity also indicates the lack of standardization in this area. Not only that but the solutions identified don't necessarily target text entry accessibility concerns.

2.2.1 Non-visual text entry

One concern in particular - the effect of different target sizes on non-visual text entry – was addressed by Rodrigues et al. [29]. They showed that the participants' input speed, landing accuracy as well as movement/exploration efficiency decreased overall as target size also decreased, however, results [29] suggest that even though larger targets presented better text entry results than smaller ones, there is an upper limit to larger sized targets, which was understood to be 10mm.

In order to address the lack of smartwatch text entry accessible methods, Luna et al. provided an evaluation for five different Braille input methods for smartwatches, as well as some relevant information on the topic [18]. To do so, they developed five different prototypes with different input approaches each - Connect, Touch, Swipe, Serial and Perkins - as well as with three extra features for complementing interactions: vibration patterns per dot, multi-frequency dual-tone feedback and screen rotation. From his study, it was concluded that a connect the dots approach, a swipe gestures approach, and a touch approach were all considered acceptable for future evaluation, with the first being the preferred method. Dual-tone feedback and screen rotation were also considered to be useful for the system's usage.

Figure 2.4: Prototype screens of some of the methods developed in use. Image taken from [18].

With approaches narrowed down to three (i.e., Connect, Swipe, and Touch) following their pilot study, Luna et al. proceeded their investigation by evaluating said approaches [20]. Their study shows the Connect approach providing the best results out of the three

averaging 10.89 WPM, followed by the Touch approach (mean=7.51 WPM), and then the Swipe approach (mean=5.78 WPM).

Since smartwatches support a collection of motion gestures, an interaction modality that can be appealing for visually impaired people, Feiz et al. assessed [8] how accessible wrist gestures are for people with visual impairments' interactions with wearable devices. The study showed that wrist gestures still have some barriers to overcome and provided principles for more accessible wrist gestures. It also showed that although users performed android provided gestures more accurately, they preferred and found custom gestures to be easier.

2.3 Gesture Recognition

One barrier that still stands in the way when designing and developing accessible touchscreen interfaces, is the lack of understanding regarding the way blind people use such technologies. A study [14] by Kane et al. suggests that blind people have different gesture preferences from sighted people, preferring gestures coming from corners or edges, and are more likely to perform multi-touch gestures. It also shows that blind users perform larger gestures, more slowly and with less accuracy, less form closure, and less line steadiness. With this in mind, it was possible to define some guidelines for more accessible touchscreen interfaces [14].

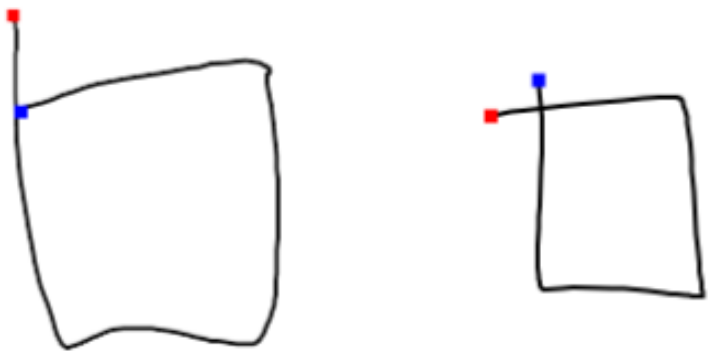
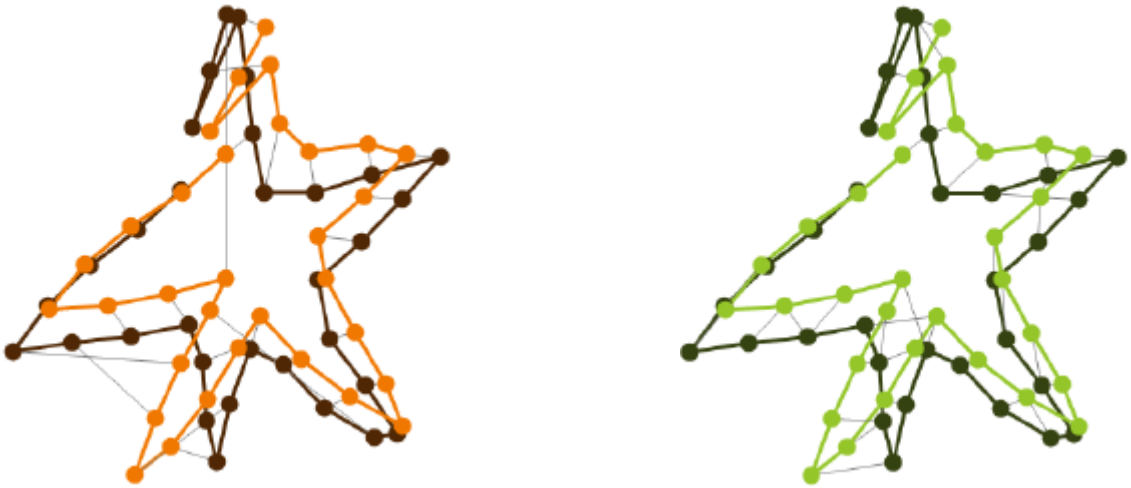


Figure 2.5: Lines drawn by a blind and a sighted participant respectively. Image taken from [14].

Understanding gesture preference and performance is a step in the right direction, but there is still the need to find more and better ways to recognize such gestures. There have been considered [4] two approaches for creating a multi-touch gesture recognizer: formally defining multi-touch gestures, and then trying to identify them based on their formal model; or by example where learning techniques try to identify a certain gesture from a set of other possible gestures. Both have their own advantages and disadvantages, and the most promising results are understood to come from mixed solutions.

Kristensson et al. presented a technique that allows visualization as "feedback" and continuous recognition of pen strokes and touchscreen gestures [15]. This technique utilizes an incremental recognition algorithm, that estimates the probabilities of a user's currently partial or complete stroke within a set of template classes. From their study, it was understood that recognizers should use Turning Angle with an end-point bias instead of the Euclidean Distance for the recognition algorithm implementation, to achieve better accuracy performance.

Amongst some of the recognizers previously mentioned [4], the "$-family" is a popular group of recognizers based on the Nearest-Neighbour approach. Vatavu et al. proposed an updated algorithmic design [37] of the one that at the time provided the best results, $P [38]. In this design update, the proposed algorithm $P+ consisted of a more flexible point matching strategy that did not rely on stroke's chronological order. It ended up providing a statistically significant difference in recognition performance compared to any other of the recognizers evaluated, producing the point matching seen in figure 2.6. Although the $P is considered to be the most flexible of the "$-family", it's not suitable even with optimizations for low-resource devices. With this in mind, Vatavu et al. presented $Q, "a super-quick, articulation-invariant point-cloud stroke-gesture recognizer for mobile, wearable, and embedded devices with low computing resources" [39]. $Q ended up unexpectedly outperforming $P in every parameter but accomplished even greater improvement in classification time. It also achieved a reduction of 97% CPU time using less than 3% of the computations of $P.



(a) Point matchings by $P (b) Point matchings by $P+

Figure 2.6: Different point matchings by $P and $P+, for a "star" gesture produced by a participant with low vision. Images taken from [37].

This type of recognizer showed great recognition results, yet, most present inherent limitations regarding the types of gestures they can discriminate. For example, some of the "$-family" recognizers require explicitly defined templates for each gesture articula-

tion, while others ignore such aspects. Both approaches have their advantages and disadvantages however, knowing more about how users perform their gestures would enable the design of a recognizer that could capitalize on such characteristics. This knowledge could also be applied to the design of gesture sets to minimize existing conflicts. With this in mind, Anthony et al. defined a methodology based on a set of articulation features to measure the consistency with which people produce gestures. In their study, they also measured the consistency of 4 gesture data sets and concluded that people are more internally consistent than between themselves. These findings had also been made in prior work [34, 5] that defined people as highly individual and internally more consistent regarding handwriting recognition and multi-touch gestures. With this, they established some guidelines for designing gesture sets and implemented a toolkit - *GECKo* - to help measure gesture consistency.

Aslam et al. presented a gesture recognition algorithm optimization using the Crow Search Algorithm, capable of predicting Braille coded gesture patterns. It consisted of an Artificial Neural Network with structure optimization using the Crow Search Algorithm, that could predict the gesture coordinates in four directions (left, right, top, bottom). The system showed a maximum prediction accuracy range for top, bottom, left and right directions to be 99.9%, 99.5%, 95.5% and 94.23% respectively. It also showed that the proposed NN optimization outperformed other existing techniques in both accuracy, sensitivity (98.5%), and specificity (99.3%) [2].

Hidden Markovian Models (HMM) had already been applied to recognition problems such as speech recognition and gesture and handwriting recognition, however, this model class presents some limitations. It does not include explicit information regarding time and duration, and only Markovian systems can be modelled exactly. Dittmar et al. tried to mitigate these issues and tried to prove that a system based on Hidden non-Markovian Models could recognize gestures that only differed in execution speed. To do so, they developed a touch gesture recognition system that uses Conversive Hidden non-Markovian Models (CHnMM). In his approach, a single model only represents a single gesture. In a first study, the devised CHnMM system had an overall precision of 95.5% and a recall of 88.3%, while in a second experiment it even reached “perfect precision”. Those results combined with a processing time between 31ms and 47ms made the system feasible for real-time scenarios [7].

2.3.1 Convolutional Neural Networks

Deep learning is used in numerous areas thanks to its diverse range of applications, and it has been improving thanks to the increase of Artificial Neural Networks (ANN). One big propeller of advances in deep learning is Convolutional Neural Networks (CNN), systems that combine ANN with current deep learning strategies. CNNs are used in deep learning for many areas [30, 17, 13] such as analysis of visual imagery, pattern recognition, im-

age classification and more, and with tools like TensorFlow Lite[3] – an open-source deep learning framework for on-device inference – they have been made more efficient for low resource devices.

Siddique et al. presented an evaluation of a Convolutional Neural Network's accuracy variation, for handwritten digits' classification, developed using TensorFlow [30]. To do so, they designed a CNN with one input layer, one output layer, and five hidden layers. These five hidden layers consisted of two convolutional layers, two pooling layers, and a fully connected layer. The study results of the training and validation accuracy obtained when altering the combination of the hidden layers, showed that the best arrangement of hidden layers consists of each convolutional layer being followed by a pooling layer, with a dropout regularization method followed by a flatten layer and another dropout, before the fully connected layer. Such arrangement provided a maximum accuracy of 99.21% for 15 epochs, and when compared to other approaches, it also provided better results.

Huang et al. utilized a different grouping of hidden layers for a CNN-based recognition model for Braille music notation. He presented [13] a CNN consisting of 3 convolutional layers, 2 pooling layers and 2 full connection layers, as well as a training and testing algorithm for the recognition model. Although it provided better results than other approaches it was compared to, it achieved worse accuracy when compared to them [30]. Li et al. also used a CNN and TensorFlow for the implementation of a handwriting recognition model. They presented a personal Neural Network Model for Chinese handwritten characters' recognition [17]. One particular aspect of this approach is that it allows the user to train and adapt their own model according to their own handwriting characteristics and writing style. This CNN-based system showed a recognition accuracy of 98.137%, with 44 errors, in a total of 2362 test sets with not only Chinese characters, but also English letters, punctuation marks and Arabic numerals.

2.4 Discussion

Table 2.1 summarizes the studies previously mentioned regarding non-visual text entry methods for smartphones as well as the text entry methods for smartwatches.

With all the aforementioned information and by analysing table 2.1, we can understand that there is already a diverse variety of non-visual text entry approaches for smartphones. Even though most of the methods analysed were designed for a specific type of device, there is still some variety regarding the approaches used, whether gesture-, navigation-, or touch-based. Amongst those, the ones that rely on the Braille writing system take advantage of previous Braille knowledge from the user, hence easing some of the cognitive load. Out of those, it is clear to see that the ones allowing multi-touch provide faster text entry results, however, those are also more susceptible to accuracy errors.

[3] https://www.tensorflow.org/lite (last visited 30/11/2021)

Such issue can be mitigated by the use of correction systems [25], some form of physical contraption to aid with spatial awareness [9], and by providing good feedback mechanisms. In cases where word-level audio feedback can be used, it has been proved to be more beneficial than character-level feedback increasing the users' typing speed [16]. Vibrotactile feedback showed great potential when applying the right stimuli, although findings suggested it should be carefully implemented when used.

These approaches can easily be translated to larger screen devices like tablets since the change from a smaller screen to a larger one doesn't necessarily constrain the previously implemented systems' interfaces. However, the same is not applicable for smaller sized devices. As seen, the amount of text entry methods specifically designed for smartwatches is still scarce, and the existing methods are not exactly accessibility oriented. In order to fill this gap, and since smartwatches smaller screen size makes it difficult to implement complex touch-based approaches, gesture-based methods would be an appropriate choice.

This decision can also be supported by the now better understanding of blind people's gestures [14], and the fact that recognizers have been evolving and more solutions for low resource devices [39] are being developed. Most of the recognizers mentioned work by analysing gesture data but with technologies like TensorFlow Lite also evolving, using Convolutional Neural Networks for image recognition is being made more efficient for low resource devices like smartwatches. This variety in recognizers provides multiple options for a gesture-based system's implementation.

Having in mind some of the users' individual characteristics mentioned to affect their touch-based interactions' performance [27], all this information could be of good use to establish some guidelines for the design and development of non-visual text entry methods for smartphones. Nevertheless, the majority of the studies mentioned were conducted in controlled environments, which leaves open questions regarding the usability of such approaches in real-life scenarios.

So, by analysing this, there is a visible area to be explored - non-visual text entry methods for smartwatches – as well as guidelines to tackle this effort in the most effective way.

Table 2.1: Overview of text entry methods for mobile devices and respective studies

Method	Devices Tested	IT	TT	WPM	ER (%)	Feedback	Env.
NavTap [12]	SP	K	S	1.25	10	A C-level	C, R-L
NavTouch [11]	SP	T	S	1.79	4	A C-level	C
No-look Notes [3]	SP	N	S / M	1.32	11	A C/W-level	C
Yfantidis *et al.* [41]	CRT	N	S	7.00	-	A/V C-level	C
oPhone [40]	SP	G	S	-	-	A/V C-level	C
Tinwala *et al.* [35]	SP	G	S	7.60	0.4	A/V C-level	C
COMPASS [42]	SW	B	Non	9.30	< 0.5	T	C
Nascimento *et al.* [23]	SW	G / T	S	-	< 6.86	-	C
Braille-based							
BrailleType [28]	SP	T	S	1.49	8.91	A D-level	C
BrailleTouch [9]	SP	T	M	17.86	28.60	A C-level	C
Hybrid-Brailler [36]	SP	T	M	6.10	10.1	A C/W-level	C
BrailleSketch [16]	SP	G	S	14.53	11	A D/W-level	C
Luna *et al.* [20] (Min)	SW	T	S / M	5.78	9.53	A/V D-level	C
Luna *et al.* [20] (Max)	SW	T	S / M	10.89	18.75	A/V D-level	C

Columns observations: IT - Input Type, TT - Touch Type, Env. - Environment
Devices Tested observations: SP - Smartphone, SW - Smartwatch, CRT - CRT Monitor
TT observations: S - Single Touch, M - Multi Touch
IT observations: G - Gesture-based, N - Navigation-based, B - Bezel-based, T - Touch-based, K - Key-based
Feedback observations: A - Audio, V - Vibrotactile, T - Tactile, D - Dot, C - Character, W - Word
Env. observations: C - Controlled, R-L - Real-Life

Chapter 3

BrailleShapes

The main goal of this work is to assess the validity and usability of Braille-based shapes as the basis for a gesture-based text entry approach for smartwatches. This comes as an attempt to mitigate some of the problems identified in section 1.1, highlighting the need for an agile and accessible text entry method. To do so, we defined the concept of a *Braille Shape*, and devised and validated a system based on such concept.

We opted to use shapes composed of single strokes so that our approach does not require precise multi-touch interactions. Doing so makes it a more appropriate option for smaller-sized devices like smartwatches, amongst other benefits later mentioned, which provides a more agile and ”on the go” approach.

With Braille providing positive results as the foundation for several accessible text entry methods, the shapes are also based on the Braille alphabet. It is a standard for visually impaired people, meaning its prior knowledge that they can leverage, helping reduce any cognitive load required. It is also a way to establish a standard for this type of interaction.

Before describing and discussing the process and results of doing so, it is helpful to establish some key concepts. For starters, let us expose a definition of Braille.

3.1 Braille

Braille[1] is a tactile writing and reading system that is used by visually impaired people. It was developed in 1824 by Louis Braille[2] who became blind at age three. This tactile code allows the reading and writing of many different languages as well as musical and mathematical notations. Additionally to text, Braille also allows to create graphs with different types of lines and illustrations.

Braille encodes up to 64 characters, by a combination of dots in a 3x2 matrix called Braille cell. A single cell can represent alphabet letters, numbers, punctuation marks,

[1] https://www.britannica.com/topic/Braille-writing-system#ref281452 (last visited 10/11/2021)
[2] https://www.biography.com/scholar/louis-braille (last visited 29/08/2022)

mathematical symbols, musical notes, or even a whole word. Its mapping differs from language to language and it may even vary within the same language. For example in the case of English Braille there are 3 levels: *contracted braille, uncontracted braille and grade 3 braille*.

It is common knowledge for many blind users and paramount to their daily lives.

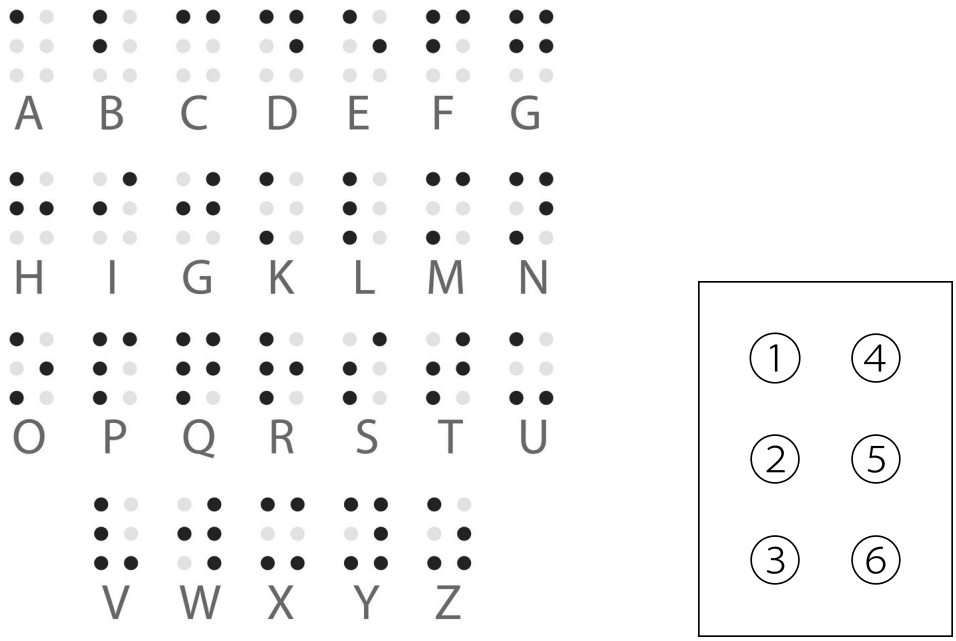


Figure 3.1: English Braille alphabet and cell labelling. Images taken from the internet[5][6]

3.2 Braille Literacy

It has been shown[3] that Braille reading proficiency is a vital tool that, if learned early on, could potentiate visually impaired people to do as well as, if not better than, sighted people in numerous areas. It can increase their exposure to more and better opportunities, thus enabling a better quality of life. For example, regarding professional life, of the estimated[4] 85 thousand blind adults in the United States in 2018, 90 per cent deemed Braille literate were employed.

Moreover, knowing Braille allows blind people to indulge in certain activities like reading physical format books while providing them access to fundamental aspects of any written language as punctuation and spelling, which are less accessible via audio. These activities can go beyond leisure, also aiding in various tasks. Reading a book using Braille instead of listening to it can be helpful when there is the need to consult something mid-

[3]https://en.wikipedia.org/wiki/Braille_literacy (last visited 10/11/2021)
[4]https://www.afb.org/blindness-and-low-vision (last visited 10/11/2021)

speech. For instance, a school teacher can consult a book while explaining it to students.

3.3 Spatial Ability

Spatial ability is another key concept and is referred to as one's capability for reasoning, understanding and remembering the spatial relations amongst objects or space[5]. It's a combination of multiple sub-skills interrelated amongst each other, vital to everyday life.

It's of great importance for numerous tasks and a key factor for success in various fields of study, however, with the always emerging new technologies and consequent increase of spatially demanding tasks, it has become even more important.

3.4 Braille Shapes

Having depicted some fundamental concepts, we now expose how they relate to the idea of a gesture-based text entry approach by explaining what it means to have a gesture-based system that takes advantage of Braille. For this, we also need to understand how to conceive gestures translated from characters encoded by dots.

We consider a Braille-based gesture-based text entry system as a system in which users can input a character by drawing its corresponding Braille cell using what we call a *Braille Shape*. We describe a *Braille Shape* as the shape obtained from performing a single stroke, passing over all the raised dots of a given Braille cell without going more than once over any of those dots.

By exploring this type of approach, we address three main aspects already mentioned and contextualized in chapter 2:

- The need for good *spatial ability* - described in 3.3 - required to perform multi-touch interactions or ones that require precise locations
- The cognitive load of learning a new alphabet
- The need for a new standard for accessible text entry in small devices

We address the need for good *spatial ability* and the complexity of performing multi-touch and precise location interactions by restraining the shape to be performed with a single stroke. This way, a system using this approach does not require its users to lift their fingers while performing the gesture, which could cause them to lose track of the touch space of the device. That is even more important for small-screen devices like a smartwatch and beneficial for usage in nomadic contexts or when single-handed use is

[5]https://www.tbase.com/the-braille-alphabet/
[6]https://commons.wikimedia.org/wiki/File Braille_cell.svg
[5]https://www.yumpu.com/s/s4uQeSqaWoli2GvW (last visited 08/08/2022)

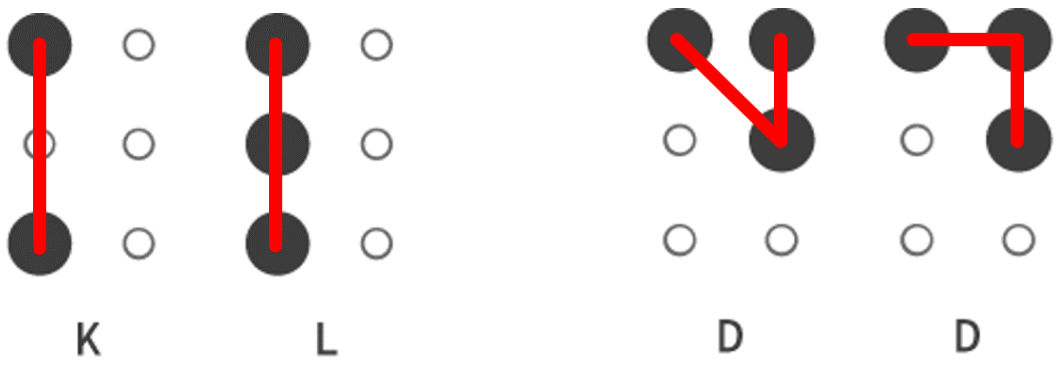
required. It is also advantageous since a system with this approach does not need any type of confirmation after each shape drawn. This decision is also supported by previous studies [1, 34, 5] that identify single-stroke gestures as more consistent and subsequently an adequate choice for recognition purposes.

Basing our approach on the Braille system helps reduce the learning and memorization required to interact with any system that uses it. This way, we address the cognitive load of learning a new alphabet since users with any prior Braille knowledge can leverage it to devise the shapes performed.

Since Braille has proven to be advantageous by providing good results in other domain-related approaches, being of common knowledge for blind people, and being versatile in its applications in various domains, it can be referred to as a good standard for accessible text entry in small devices hence addressing that topic as well.

3.4.1 Ambiguity

While using this approach, even though it is possible to represent each character in the alphabet using a unique shape, we acknowledge the possibility of ambiguous representations for several letters. For example, inputting the letters “L” and “K” could be done the same way, producing the same outcome, a vertical line, as seen in figure 3.2a. This ambiguity increases the difficulty in the recognition tasks making this approach more challenging to implement for real-life applications. It must be noted that throughout this study we only considered the 26 letters of the alphabet, meaning no ambiguity resulting from numbers and punctuation was accounted for.



(a) Same *Braille Shapes* representation of two different characters

(b) Different *Braille Shapes* representation for the same character

Figure 3.2: Examples of ambiguous character representations using *Braille Shapes*

One way our approach attempts to mitigate this issue while also decreasing its complexity is by requiring that the shape composed must not go over the same cell dots more than once. This way, the shapes tend to be more ”open”, and the number of shape pos-

sibilities for each character diminishes. However, there is still a significant amount of possible variations leading to ambiguity in systems using this approach.

For that reason, we analysed different ways a system with this approach can further attempt to mitigate those scenarios:

1. Using a Braille-aware spellchecker to help with character-level and/or word-level correction
2. Providing the user with a choice between multiple options per shape drawn
3. Using a user-dependent recognizer
4. Using a recognizer that takes into account the size and position of the gestures performed

This analysis is presented in the following chapters.

3.5 BrailleShapes System

Having established the concept of a *Braille Shape* and bearing in mind the information regarding prior work, we have conceived a system using our approach. This section encompasses a rundown of the features envisioned for that system as well as the reasoning behind such design decisions and how the system helps address topics highlighted in chapter 2, further extending the beneficial impact of using *Braille Shapes*.

We describe our system as an ambiguous keyboard where users insert text by inputting *Braille Shapes* and where words are suggested based on a Braille-aware spellchecker. Even though the spellchecker reduces the need to amend each wrongly recognized character, the system still provides the user with multiple character options per shape drawn. The strokes composing the shapes can be started anywhere on the screen without specific orientation, as long as they are entirely within screen bounds. Our approach relies on a lightweight recognizer capable of real-time recognition on low-resource devices. It resorts to the Android TextToSpeech engine to provide audio feedback, and it also provides haptic feedback when needed.

The system is focused on smaller-sized touchscreen devices like smartwatches to help mitigate the lack of accessible text entry methods for those devices. Such a decision is supported by *Braille Shapes'* effort to reduce the need for good *spatial ability* and multi-touch and precise location interactions, which could prove to be an even bigger challenge in such a form factor. The system also reinforces this effort by allowing gestures to be performed anywhere on the touch surface without specific orientation.

This implementation does not require any interface components other than the touch surface. Hence, it does not present significant difficulties in being converted to larger form factors, also helping support the decision of focusing on smaller-sized devices. Any

information of the text being composed is also absent, thus addressing one particular concern depicted in section 2, the privacy and social concerns raised by other approaches like ones requiring voice input.

By using a Braille-aware spellchecker and providing options at a character level, the system tackles the existence of ambiguity in character representation, resorting to the ways mentioned in points 1 and 2 of section 3.4.1. However, providing character options is thought of as an optional togglable feature since the aim is to focus on word-level feedback, due to conclusions drawn from prior work [16] that mentions word-level feedback to provide better results than at a character level.

To help with that matter, audio feedback is also defaulted to word-level, with the option to have character-level feedback only when using the character options feature. By default, the system reproduces distinct sounds for any wrongly performed action (i.e., unrecognized gesture, a gesture performed out of bounds) and speaks the recognized and corrected word after confirmation, with the option to reproduce every character recognized. Since the system focuses on word-level feedback and correction, it also defaults the delete option to be applied to the entire word unless the character options mode is enabled. However, in a situation where a user knowingly wrongfully performs a gesture, the system also allows character-level deletion with a different interaction mechanism.

Based on already existing approaches and the information gathered from the data collection study presented in section 4.2, we decided on a possible set of actions to interact with the system. Some of them were already mentioned but are further detailed and contextualized in table 3.1.

3.6 Auxiliary Systems

Throughout this work, two user studies took place. These studies are in more detail in sections 4 and 6. Nonetheless, each required a set of custom applications for a specific purpose, each specified in this section.

3.6.1 Collecting BrailleShapes

Our goal with the first study was to collect *Braille Shapes* performed on a smartwatch from sighted and visually impaired people. For that, we developed an inputting application and a control application.

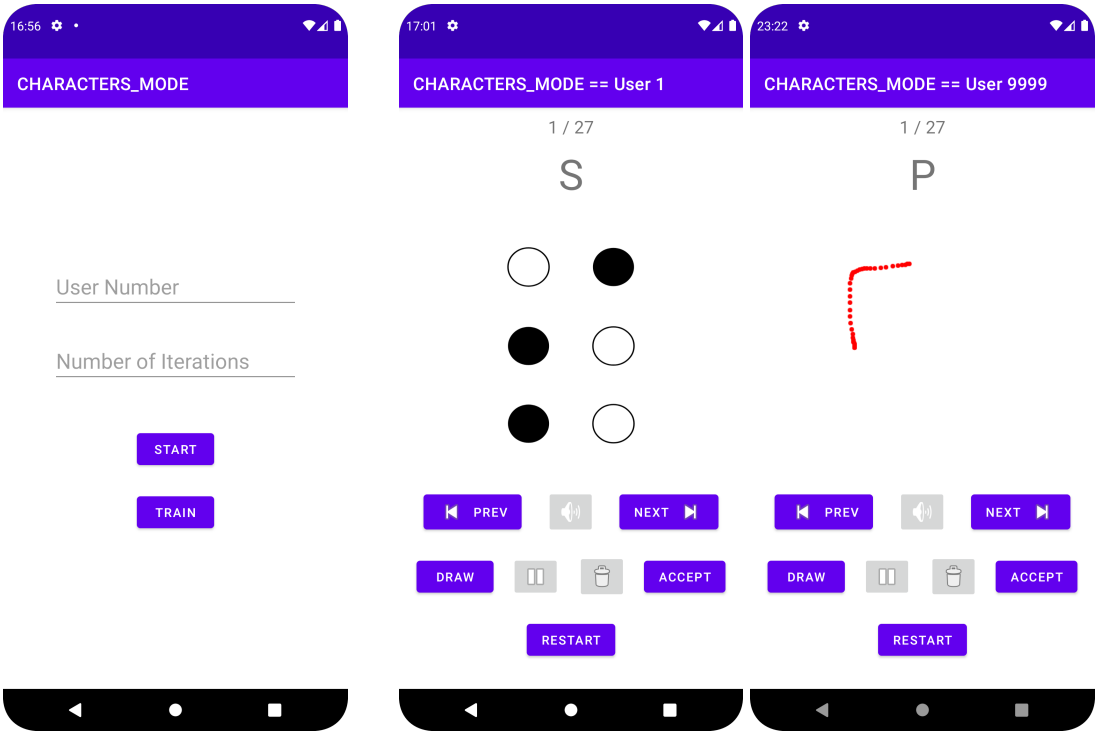
With the desired mean of participant interaction being a smartwatch device, we developed the inputting application for that specific form factor. When conceiving it, we did so that participants only input requested shapes and receive different types of feedback. We developed the control application to manage the state of the inputting application and visualize data and information during the procedure. It was conceived for a larger

Table 3.1: System's interaction possibilities

Action	**Description**	**Outcome**	**Feedback Type**	**Feedback**
Input Character	Drawing *Braille Shapes* on the touch surface - only one shape at a time	Recognized character added to current word	Haptic	Short vibration
Double-tap (1)	Performing a double-tap a first time - after inputting a character	Word confirmation. **"Space"** added after current word	Haptic & Audio	Pattern vibration, word spoken
Double-tap (2)	Performing a double-tap a second time - after performing a double-tap	Sentence confirmation. **"."** added after current word	Haptic & Audio	Pattern vibration, sentence spoken
Input Shape (*"Circle"*)	Drawing a "Circle" on the touch surface	Last entered word deleted	Haptic & Audio	Short vibration, the now current last word is spoken
Long-press	Performing a long-press	"Scroll" through word suggestions - select word on release	Haptic & Audio	Pattern vibration, suggested words spoken
Two fingers downwards slide	Performing a downwards slide w/ 2 fingers	Last entered word provided via text to speech	Audio	Word spoken
(2 x) Two fingers downwards slide	Performing 2 downwards slides w/ 2 fingers sequentially	Last entered sentence provided via text to speech	Audio	Sentence spoken
Error	Wrongly performed action - i.e., unrecognized gesture, a gesture performed out of bounds	Audio feedback provided	Audio	"No" and a buzzer like sound provided via text to speech respectively

and more powerful device (e.g., smartphone) to provide more flexibility and performance when interacting with it.

The inputting application consists of a single blank screen with two possible states managed by the control application - *blocked and unblocked* - intended to prevent unwanted interactions with the system. When in its *unblocked* state, it listens for touch events. If any is detected, it saves information about the event until it stops. Oppositely, no interaction data is recorded when in a *blocked* state. In both states, it allows different types of feedback- i.e., audio and vibrotactile - when requested by the control application. According to the action performed or being performed, the vibrations produced can be of different frequencies, duration, and intensities. The audio reproduced can be a simple beep-like sound or an actual character or word, processed using the ToneGenerator class or Android's TextToSpeech Engine, depending on which it is. By request of the control application, the state of the inputting one can be in training or collecting mode. The difference between the two is that the training mode provides continuous audio feedback during a touch event, meaning that while a participant touches the screen, a beep sound gets reproduced. We implemented this to help participants get used to the approach and the device.



(a) Initial screen. (b) Control screen.

Figure 3.3: Data collection study control (smartphone / *host*) application's interface in different stages of the process.

Contrarily, the control application is more complex due to its interactions' requirements and the space available in its designated form factor. It allows the selection of

which shape to request from the participant and subsequently acceptance of the gesture produced, and controlling the state and functioning of the inputting application throughout. Besides managing the procedure, it also processes and stores data.

Its interface consists of two screens, as seen in figure 3.3, an initial screen to set up the procedure for the participant and stage in question and a second one to control and perform the actual procedure. Setting things up is done by inputting participant data (i.e., participant ID) and procedure settings (i.e., the number of iterations and the state of the procedure). For managing the procedure there is a visualization area (the top area of the screen) and one for interaction purposes (the bottom area of the screen). The visualization area shows the character requested and the corresponding braille cell, and the gestures being performed in real-time. It is a helpful feature to monitor the procedure and ensure it is performed correctly by the participants. In the interaction area are the buttons that control the procedure. These buttons send commands to the inputting application to perform actions such as changing its state or providing a certain type of feedback while also updating the control application's state.

A two-way communication exists between the applications through the Wearable Data Layer. The inputting application sends the data collected from the touch events in the form of XML files and receives commands controlling its state and the actions it performs. Those commands are sent by the control application, which also receives the gesture data produced by the participants on the inputting application.

3.6.2 BrailleShapes for Text Entry

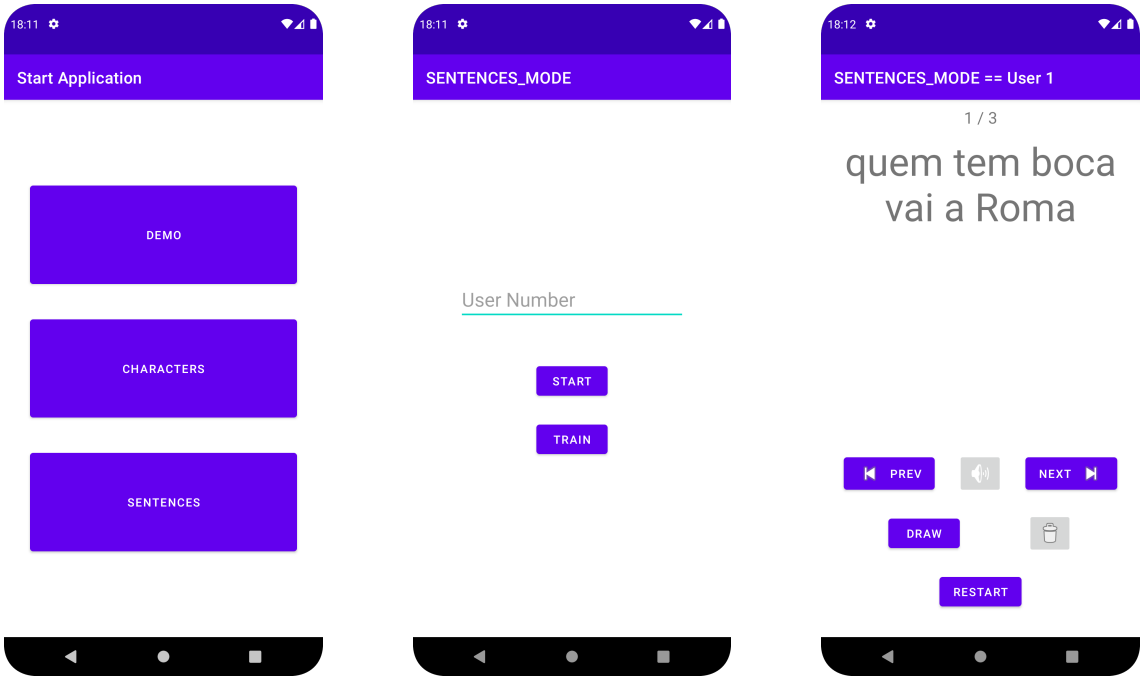
The second study was to provide a preliminary validation of the approach as a text entry method, which was done by performing a similar type of data collection as in the previous one. For this one, however, besides collecting *Braille Shapes* for each character of the alphabet individually, we also collected them as part of sentences. For it, updated versions of the previously built control and inputting applications were conceived.

This new iteration of the control application allows changing between *characters* and *sentence* collection modes on the initial screen. Selecting the *character*'s option leads to the same control screen as the previous version of the system. The real change is when opting for the *sentences* mode, which leads to the screen shown in figure 3.4c. Its interface consists of two areas, as before, with the visualization area adapted to handle sentences. The control area allows the selection of which phrase is requested to the participant while monitoring its progress throughout the sentence, i.e., visualize which character of the sentence the participant is currently on. The application's design and functioning regarding data storage and processing are the same as in the previous version.

Having this second study focusing on validating the approach from a text entry perspective, inputting features were added to the smartwatch application allowing a more complete usage. Apart from collecting data from the touch events, it also allows inputting

spaces between words, deleting inserted gestures, and confirming inputted sentences. These actions are completed by performing a double-tap for inserting *space* characters and sentence confirmation (which is dealt with by the system), and a long-press for gesture deletion, with each providing different types of audio and haptic feedback.

There is still a two-way communication between the applications using the Wearable Data Layer, and all the existing features of the previous version are also inherited.



(a) Initial screen. (b) Secondary Screen. (c) Control screen.

Figure 3.4: Text entry study control (smartphone / *host*) application's interface.

Chapter 4

Exploring the Feasibility of BrailleShapes

As previously mentioned, to accomplish our goal of validating Braille-based shapes as the foundation for a text entry method, we first need to collect and analyse *Braille Shapes*. This chapter is divided into two sections, one describing the steps taken and another discussing the results obtained.

The first section – 4.1 – describes the data collection process detailing its different stages and requirements, the study's population and the data collected.

The second section – 4.2 – presents the results of analysing the data collected in terms of its properties and the way it was conceived.

4.1 Data Collection

We started by conducting a user study to collect information and gestures – *Braille Shapes* – from both sighted and visually impaired people. This allowed us to compose our data set and to acquire important information for our system's design and development. It also allowed us to better understand our population, how they perceived our approach, and how they conceived and performed the gestures requested.

Even though the system is meant for people with visual impairments, we saw it as beneficial to also collect gestures from sighted people since using their gestures to recognize gestures from blind people showed promising results in previous studies [14].

4.1.1 Participants

We recruited 20 participants with visual impairments (9 self-identified males, 11 self-identified females) and 18 without visual impairments (12 self-identified males, 6 self-identified females) for this study, as described in table 4.1. Participants were recruited from the institution Fundação Raquel e Martin Sain and via word of mouth.

Sighted participants had an average age between 24 and 35 (SD=0.51) and were predominantly right-handed, with only one participant being ambidextrous. They all reported to be at ease with touchscreen devices on a 5-level Likert scale (mean = 4.9) which can be a result of all participants using their smartphones daily (self reportedly). Regarding smartwatches, 7 participants mentioned having had any form of contact with such devices, with only 4 of them mentioning wearing one daily.

Visually impaired participants had an average age between 36 and 45 (SD=1.17) and were also mostly right-handed, with only 3 participants being ambidextrous and 3 being left-handed. Of those participants, 12 were totally blind and the rest reported to have up to 10% vision with 9 of them having this impairment since birth. Out of those who acquired vision loss with age, only one reported to having had visual contact with a touchscreen device (smartphone). Visually impaired participants reported an average of 4 on a 5-level Likert scale regarding being at ease with touchscreen devices, even though all but one reported using their smartphones daily. This is hypothesized to be due to some of the participants' minimal use of those devices since several of them reported only performing actions such as making and receiving phone calls and sending and receiving messages. Regarding smartwatches, 6 participants mentioned wearing one every day, while the remainders reported never having contact with such devices. Those who mentioned wearing one regularly reported to interacting minimally with the device only performing basic tasks such as hearing the time and notifications. In comparison to other studies, our visually impaired population is very specific and can be considered as an older and less tech-savvy population.

4.1.2 Requirements

Our only requirement for visually impaired participants was that they possessed Braille knowledge. This information is detailed in table 4.1 where participants are referred to as numbers from 1 to 20 to preserve anonymity. This requirement did not apply to sighted participants since we could show them the Braille alphabet in an agile manner, and it would probably restrain their recruitment.

4.1.3 Apparatus

The devices used for the data collection task were a Google Pixel 4a running Android 12 and a 44 mm Samsung Galaxy Watch4 with Google WearOS. Both devices were running the custom apps mentioned in section 3.6.1 which were devised for the purpose of the study, however, participants only interacted with the smartwatch.

The smartphone was running a "host app" that allowed the investigator in charge to communicate with the smartwatch application, controlling its execution flow and thus the study's procedure. This application allowed the investigator to go through the various

Table 4.1: Visually impaired participants' collected information

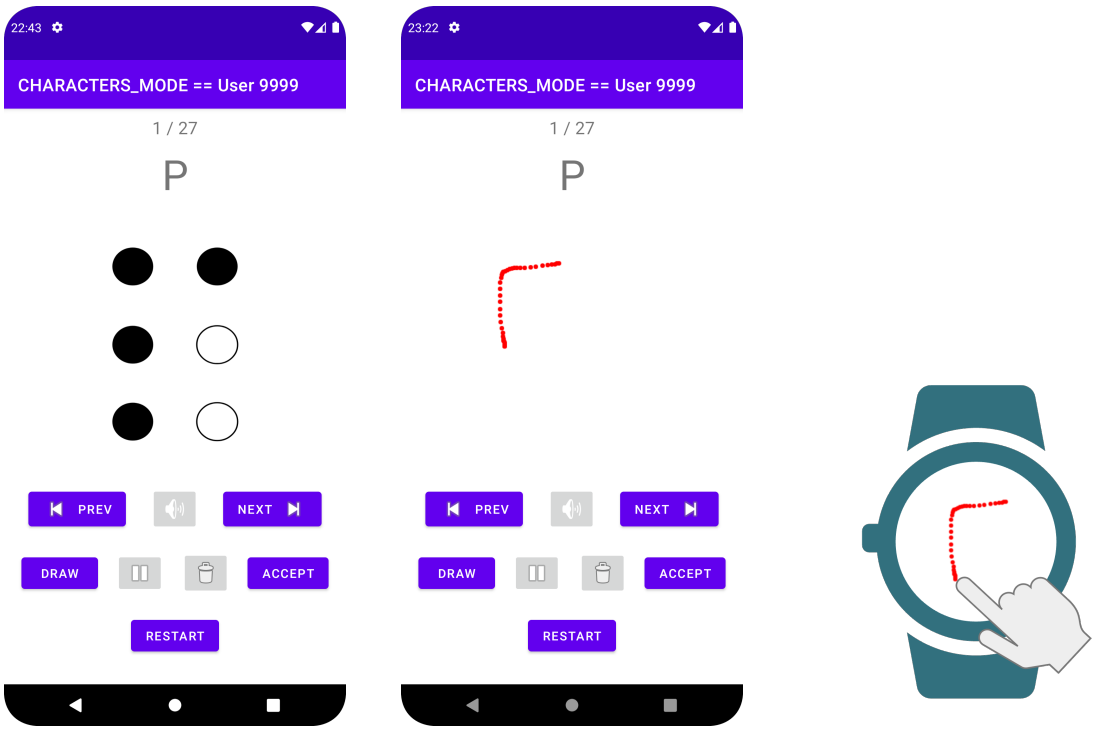
#	Age Group	Age of Blindness	Literacy	Braille Proficiency[1]	Used typing method(s)[2]
1	24 - 35	10	Higher Education	4 / 5	QWERTY (direct touch)
2	46 - 55	19	Primary Education	3 / 3	Dictation
3	56 - 65	47	Higher Education	3 / 2	Dictation
4	24 - 35	11	Higher Education	5 / 4	QWERTY
5	24 - 35	0	Secondary Education	4 / 5	QWERTY (standard mode)
6	46 - 55	0	Higher Education	5 / 5	QWERTY
7	36 - 45	31	Secondary Education	4 / 0	Dictation
8	24 - 35	0	Primary Education	4 / 3	QWERTY (standard mode) & Dictation
9	24 - 35	0	Secondary Education	0 / 3	QWERTY (magnified)
10	36 - 45	10	Primary Education	5 / 5	Dictation
11	24 - 35	0	Secondary Education	5 / 4	QWERTY (touch typing mode)
12	46 - 55	23	Higher Education	4 / 4	QWERTY (standard mode)
13	56 - 65	3	Primary Education	4 / 4	-
14	56 - 65	44	Secondary Education	4 / 5	Dictation
15	24 - 35	18	Secondary Education	5 / 4	QWERTY (standard mode)
16	36 - 45	0	Higher Education	5 / 5	QWERTY (touch typing mode) & Dictation
17	24 - 35	0	Primary Education	5 / 5	QWERTY (touch typing mode) & Dictation
18	36 - 45	0	Primary Education	5 / 4	QWERTY (standard mode) & Dictation
19	46 - 55	28	Secondary Education	5 / 4	QWERTY (standard mode)
20	36 - 45	0	Secondary Education	4 / 4	QWERTY (standard mode) & Dictation

[1] Reading / Writing - Self-declared in a Likert scale of 1 to 5
[2] All participants that used QWERTY but participant 9 used TalkBack or VoiceOver

characters one by one, whether backwards or forwards, to change the state of the smartwatch's application and to make the smartwatch audibly reproduce a specific character. It also dealt with all the necessary processing and storage of any gesture data received from the smartwatch, allowing its visualization both in real-time and after all data was collected.

The smartwatch application was designed solely for inputting and initial testing of the different feedback types we intended to use. For that reason, its interface consisted only of an empty screen whose state was controlled by the investigator. It possessed both audio and vibrotactile feedback, as well as safety mechanisms to prevent undesired interactions.

Both applications' layout configurations and some stages of their workflow can be seen in figure 4.1.



(a) Smartphone *host app* waiting for gesture data.

(b) Smartphone *host app* while receiving gesture data.

(c) Smartwatch app while performing gesture.

Figure 4.1: Data collection study applications' workflow stages.

4.1.4 Procedure

This first study started with an introduction of the investigators, and the study itself, mentioning its purpose and the procedure. While introducing the study, participants had the opportunity to interact with the smartwatch device with the screen *blocked* by the custom application to accustom themselves to it. Then, an initial characterization session took place consisting of an oral questionnaire regarding demographic data and technology pro-

ficiency. Visually impaired participants were also enquired about their Braille proficiency in reading and writing.

Afterwards, a brief training session of around 5 minutes was performed to allow the participants to better understand both the system and the procedure. In this session, participants were asked to input a small number of letters from the alphabet by drawing their corresponding *Braille Shapes*. To do so, for each letter we asked the participants to "envision the shape resulting from drawing a single stroke that would go over all the raised dots in its corresponding Braille cell". It also had to be done without going over any dot more than once. Those were the only two restrictions imposed on the participants throughout the study. We also advised them not to perform the gestures starting or ending at the edge of the smartwatch. This was intended to minimize situations where the gesture would not be recorded due to the device not detecting any interaction. The letters selected for this particular session - ABDGKMORUXZ - were the ones thought to go over most scenarios (e.g., a single dot, ambiguous representation, missing dots in the middle row, etc.). Other than the letters mentioned, participants were asked to input a circle as in print writing and not in Braille. It was not explained the intention behind such a request, however, this shape was collected for recognition and interaction purposes.

The session started with the smartwatch application screen in a *blocked* state to prevent any unwanted touches. The application state and the entirety of the procedure (e.g., the actions it performed) were controlled by the investigator using the smartphone "host app", as previously mentioned. During the training session, the smartwatch application presented the participants with a randomly selected letter at a time by reading it aloud using Android's TextToSpeech engine. Participants were instructed to say the letter heard back to confirm they had perceived it correctly. If so, the investigator *unblocked* the smartwatch application screen, and participants performed the gesture of the letter requested. While touching the screen, a beep-like sound played so that participants knew if they were performing the interaction correctly. Haptic feedback was also provided on different occasions to aid the participants in situating the procedure and the action they were performing. If no situation arose requiring the repetition of the gesture, it was accepted and the data was sent to the smartphone application. The smartwatch screen would go back to a *blocked* state. This process was repeated for all the aforementioned letters, or until reaching the time limit.

For the data collection session, the same procedure took place only this time we requested each participant to input all 26 letters of the alphabet plus a circle, a total of 3 times for visually impaired participants and 2 times for sighted participants minus the "circle" shape. In this session, no sound was emitted when touching the smartwatch screen since it is a mechanism meant for participants to adapt to the device, but the haptic feedback remained the same. Even though no time limit was established, we tried to cap the session to a maximum of 45 minutes to limit this study to an hour per participant.

Finally, after collecting the gestures' data, participants were prompted to opine regarding interaction mechanisms for a possible finalized version of a system using this approach. First, they were asked to imagine a full functioning system that would recognize the gestures performed, allowing them to input characters and write words and sentences. Then, having that in mind, they were asked 2 questions:

- *"Of the following options, which would make the most sense for you to use to insert a space when you finish writing a word?"*
- *"Of the following options, which would make the most sense for you to use to erase the last character you entered?"*

For each question, they were presented with the same options to chose from: *double-tap, single two-finger touch, long press*, and *draw a circle*. They were also given the option to suggest a new way to perform any of those actions.

This procedure was idealized to a maximum of an hour per participant, and toke about 45 minutes to an hour to complete depending on the participants' performance.

4.1.5 Data

The data collected from the participants of the gestures performed was stored in XML files.

For each gesture class per user a file was created, and it harvests all the instances of that class for that user. It contains information regarding the entire gesture, i.e., the corresponding character and the number of existing points and strokes, and it is structured according to the gesture's components, i.e., the strokes composing the gesture. Each stroke is characterized by an index, and it comprises all the respective points of that specific stroke. Each point is made of X and Y coordinates and a timestamp.

Information regarding the collection itself is also present, with the starting and finishing times of the collection – since the user is asked to perform the gesture until its completion – also being stored. This way of structuring and storing the data allows for a quicker visualization of the raw data, and an easier utilization for future need. It also allows a gesture to be reconstructed if a visual representation is needed.

The questionnaires regarding demographic data and technological and Braille proficiency were created and performed using Google Forms.

4.2 Data Analysis

4.2.1 Gestures Collected

As mentioned, visually impaired participants were asked to input 3 *Braille Shapes* for each of the 26 letters of the English alphabet plus 3 "circle" shapes as in print writing.

Only 2 gestures per letter were requested for sighted participants, and no ”circle” shape was requested. This makes a total of 27 * 3 * 20 = 1620 gestures performed by visually impaired participants and 26 * 2 * 18 = 936 by sighted participants. This decision was required to expedite the procedure due to time constraints created by the COVID-19 pandemic, and since our focus was on visually impaired participants' gestures, it served as a safety measure to ensure we gathered the most data from them to mitigate any situation where gestures end up not meeting our requirements.

One visually impaired participant in particular performed all of the gestures as if the Braille cells were inverted vertically, something the participant mentioned to be due to using a slate and stylus to write. This method allows typing Braille by using a stylus to punch holes in a slate, which is to be done in reverse order so the slate can be flipped and read. When performing any analysis on the gestures, however, those were inverted to match the orientation of the remainder.

Gestures from 4 visually impaired participants are not accounted for in this analysis. None of these participants completed the task as requested, with some not even performing the number of gestures required within the stipulated time frame. Hence, the number of gestures from visually impaired participants used throughout this analysis is 27 * 3 * 16 = 1296, making a combined total of 2232 gestures collected.

Each participant whose data was discarded from this analysis presented similar reasons for this, with the most noticeable being, in frequency order:

1. *The uneasiness when interacting with the smartwatch.* This led to the participants not performing the gestures in their entirety within screen bounds, with some of them even starting the gestures on their arm or hand instead of the watch. They also performed several touches on the screen each time they went to perform a gesture, to situate/identify the touch surface.

2. *The lack of hand steadiness* made it hard to perform single-stroke gestures. Participants unwillingly raised their hands mid-gesture.

3. *The inability to envision different shapes for each letter.* Participants performed the strokes similarly to the way they read a Braille cell. They performed the strokes going through the numbers corresponding to the dots of the cell in sequential order from 1 to 6, with the only difference being not going through some of them depending on the character, making their shapes indistinguishable from each other.

4.2.2 Gestures' Properties

In addition to collecting the gestures from the participants, we examined various properties of such gestures to better understand the way they were performed and if there is any significant difference between gestures from visually impaired and sighted participants and even between the gestures performed by each participant.

Size

For each gesture, using its bounding box area, we measured its overall size.

Sighted participants produced gestures with an average size of 85371 (SD=92890.2) pixels^2, whereas visually impaired participants produced gestures with an average size of 91290 (SD=11664.3) pixels^2. Unlike [14], no significant differences in size were found between the gestures of both participants' groups ($t(28.6)=-1.6$, $p>0.1$), however visually impaired participants still performed larger gestures overall.

Size Variation

Other than the gesture size, we also examined the size variation of multiple instances of the same gesture. We did this for each participant using the standard deviation of the gesture size for each gesture class.

Sighted participants averaged a deviation of 4700 (SD=1309.0) pixels^2, while visually impaired participants averaged 11956 (SD=3074) pixels^2. There is a significant difference in the standard deviation of sizes between gestures from sighted and visually impaired participants ($W=0$, $p<0.05$), with the latter producing greater deviations when performing the same gesture multiple times.

Aspect Ratio

Regarding the aspect ratio of each gesture's bounding box (width/height), the average for sighted participants is 1.04 (SD=0.07) and for visually impaired participants is 1.06 (SD=0.16). This suggests that visually impaired participants tend to create slightly wider gestures however, no significant difference was found between the two groups ($t(19.9)=-0.4$, $p>0.1$). This findings go in accordance with the findings from [14].

Length

We also measured the length of the gestures, meaning we measured the distance between their points in sequential order, from the first chronological point to the last one.

Sighted participants produced gestures averaging 332.2 (SD=56.7) pixels long, while visually impaired participants averaged 338.0 (SD=81.8) pixels in length. When comparing the length of the gestures produced, no significant difference is found ($t(26.3)=-0.2$, $p>0.1$).

Thinking Duration

One aspect in particular we wanted to understand was the amount of time participants toke to think of a *Braille Shapes*. We intended to see whether participants' Braille knowledge had any impact on the time they took to conceive the shapes, and if there were other aspects impacting it as well.

For that, we analysed the time they toke since a gesture was requested until they started to perform the gesture. On average, sighted participants toke 3340 (SD=964.2) milliseconds to perform a gesture from the moment it was requested, while visually impaired participants toke on average 5924 (SD=3292.6) milliseconds. A Wilcoxon rank-sum test found a significant difference in the thinking duration required by both groups of participants (W=38, $p<0.05$), which indicates that visually impaired participants toke significantly more time to idealize the shapes performed.

When trying to understand what affected participants' thinking duration, we can see that its significantly correlated to the number of dots per Braille cell with a correlation coefficient of 0.7 and p-value below 0.05. If we consider cells with more dots to be more complex, this can mean that more complex cells impact the time visually impaired participants take to conceive the gestures however, the cell's complexity is more than just its number of points. The position of the dots in comparison to each other, is another aspect that can further impact the complexity of devising a shape. For example, the fact that the dots of the Braille cell are all in the same column (e.g., character "L"), or that the middle dot is in a different column then the rest or is even absent (e.g., character "W"), can decrease and increase the thinking duration respectively as can be seen in figure 4.2.

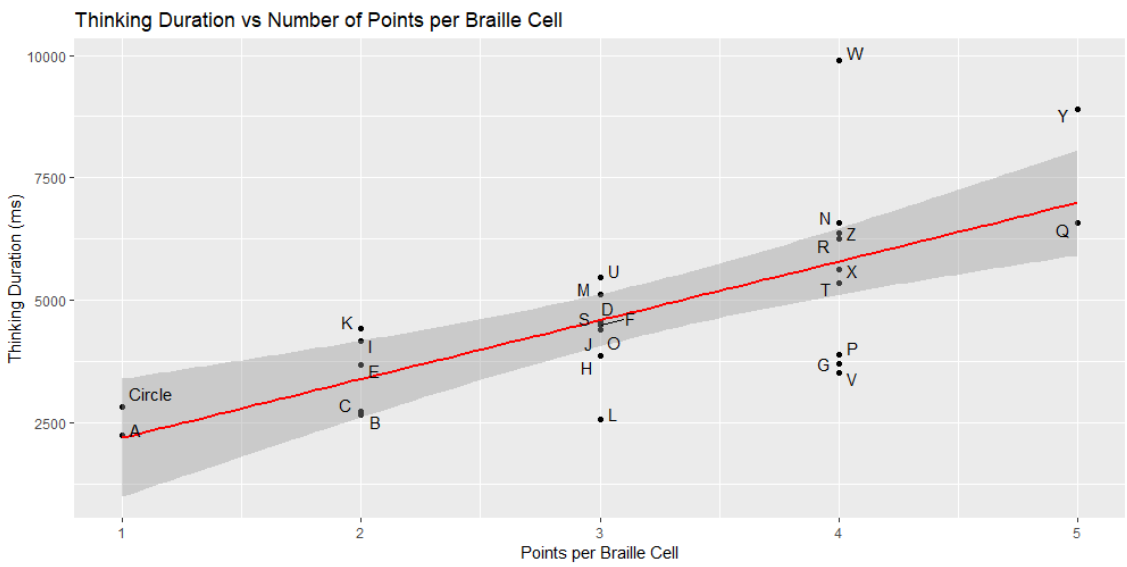


Figure 4.2: Thinking Duration vs Number of Dots in a Braille Cell

Another aspect to take into consideration is the participants' Braille knowledge. By basing our approach on the Braille alphabet, we intended to allow participants to leverage their Braille knowledge, thus reducing the cognitive load of learning a new alphabet and possibly reduce the time taken to process the characters and translate them into shapes. However, no significant correlation was found between visually impaired participants' Braille knowledge and thinking duration. One thing/aspect that is worth mentioning is that participants self-reported Braille proficiency was not necessarily accurately representative

of their actual knowledge, since several participants required assistance in remembering several Braille cells.

Duration

On average, sighted participants toke 833.5 (SD=169.4) milliseconds to perform their gestures from the moment they touched the screen until the moment they raised their fingers, while visually impaired participants toke 1705.9 (SD=558.9) milliseconds to perform theirs. A Wilcoxon rank-sum test showed a significant difference in the duration of which sighted and visually impaired participants performed their gestures ($W=8$, $p<0.05$), indicating that visually impaired participants take longer to do so.

As we can see in figure 4.3, the same reasons that impact thinking duration (i.e., what we mentioned as complexity of a Braille cell) can impact the time each gesture takes to perform. There is a significant correlation between the number of dots per Braille cell and the time both groups toke to perform the corresponding shape, with a correlation coefficient of 0.80 and p-value below 0.05. This correlation is even stronger for visually impaired participants alone with a correlation coefficient of 0.83 and p-value also below 0.05.

Of all the characters, ”a” is the one that toke the least time to gesture with an average duration of 139.6 milliseconds also being the one with the least amount of dots, while the character ”X” is the one that toke the most time to compose.

Again, the cell’s complexity is more than just its number of points which can be highlighted by the outliers (i.e., the characters ”L” and ”X”) in figure 4.3.

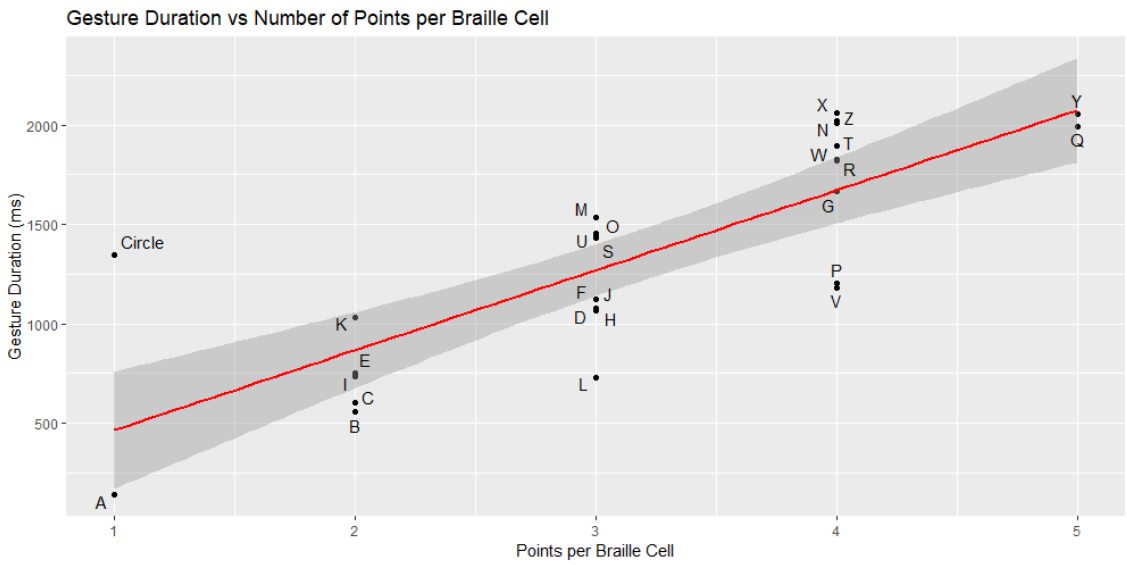


Figure 4.3: Gesture Duration vs Number of Dots in a Braille Cell

Nonetheless, even though the duration a gesture takes to be produced can be influenced by the cells’ complexity, it can also be impacted by the shape’s complexity which

in part can be reflected by its thinking duration. There is a significant correlation between the duration a gesture takes to produce and the time its shape takes to conceive with a correlation coefficient of 0.8 and a p-value below 0.05.

4.2.3 Gesture Consistency

Besides looking at the properties that characterize the gestures performed by the two groups to compare them, measuring the gestures' consistency within users and between users can improve the understanding of how participants perform their gestures and subsequently help decide the recognition mechanisms to apply.

So, using GECKo [1], we measured the consistency of the gestures between users and within-users for all participants. We did this both considering and not considering articulation direction as differentiating aspect since some recognizers take it into account.

Clusters

For both cases, we started by clustering the gesture data, which allowed us to know the average number of variations per gesture class. GECKo [1] groups the gestures with its clustering technique however, it also allows us to manually manage the clusters of gestures according to our preferences and needs. This way, we managed to group the gestures taking and not taking their articulation direction into account as illustrated in figure 4.4.

Accounting for the articulation direction, based on our gesture clustering, visually impaired participants produced an average of 4.9 (SD=2.6) variations per gesture class, while sighted participants produced an average of 3.4 (SD=2.1). A Wilcoxon rank-sum test confirmed that this difference is significant (W=491.5, $p<0.025$), with visually impaired participants producing more variations per gesture class. For visually impaired participants, the characters "B", "C", "E", "H", "I", "L", and "V" had the least amount of representations with 2 variations per gesture class, and the "circle" shape had the most with 13 followed by the character "R", with 9 different representations. For sighted participants, the characters "A", "D", "E", and "V" had the least amount of variations with only 1 representation per gesture class, and the character "W" had the most variations with 8 different representations.

Discarding the articulation direction as a clustering feature, visually impaired participants produced an average of 2.9 (SD=1.6) variations per gesture class while sighted participants produced an average of 2.3 (SD=1.7). In this case, there is no significant difference between the 2 groups. Visually impaired participants produced a single representation for the characters "B", "C", "D", "E", "H", "I", "L", "V" and the "circle" shape, while the character "O" had 6 different representations. For sighted participants, half of the characters had a single representation, including the "circle" shape, and the character "X" had the most variations with 6 different representations.

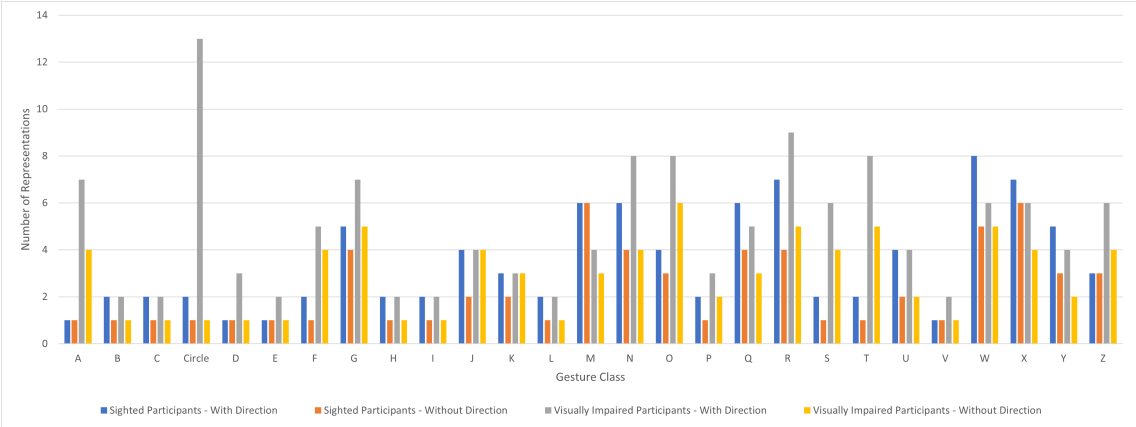


Figure 4.4: Number of representations per gesture class for each group (with and without accounting for articulation direction)

Even though the character ”A” has only a single dot in its corresponding Braille cell, it is not amongst the characters with the fewest variations in its representation for visually impaired participants as it is for sighted participants. This is due to the interactions visually impaired participants sometimes had with the touch surface, in which they would slightly move their finger when attempting to tap it to perform the gesture. This led to more and different shapes for the character in question with and without accounting for articulation direction. Nonetheless, when looking at the gestures with the least amount of representations, we can see a slight similarity between groups. In both cases and for both groups, no significant correlation was found between the number of representations per gesture class and what we considered as some of the complexity features of a gesture. However, there is a significant positive correlation between the two groups when discarding articulation direction with a correlation coefficient of 0.5 and p-value below 0.05, which can complement the similarity found between the characters with the least representations.

With these results, we can understand the impact the gesture’s articulation direction has on the number of representations produced when comparing both groups. As we can see, there is only a significant difference in the number of representations per gesture class between groups when taking articulation direction into account, with visually impaired participants performing more variations for each gesture class. When further analysing the gestures from each group, it is noticeable that sighted participants started most of theirs from the top, while visually impaired participants had a more mixed approach which can explain that difference.

This impact is also felt within each group. There is a significant difference in the number of representations each group produced with and without accounting for articulation direction. Sighted participants produced significantly more variations for each gesture class when accounting for articulation direction ($W=494$, $p<0.05$), and so did visually

impaired participants (W=531, $p<0.05$), with the difference being more significant for the latter.

Agreement Rate

After clustering the gestures as found appropriate, we measured their agreement rate using GECKo [1], which allowed us to understand how similar the gestures produced are. The higher the rate, the higher the similarity between the gestures produced. We did this again between and within groups, accounting and not accounting for the articulation direction, which can be seen in 4.5. It is expected that a lower number of representations for each gesture class would translate to a higher agreement rate.

When considering articulation direction, we found a low degree of agreement for both groups individually, with sighted participants showing an agreement rate of 0.64 (SD=0.28) and visually impaired participants displaying a rate of 0.49 (SD=0.27). A higher degree of consistency, however, is found when not considering articulation direction, as it would be expected based on the above-mentioned results that show a lower number of representations when not taking articulation direction into account. Sighted participants showed an agreement rate of 0.79 (SD=0.29), and visually impaired participants had a rate of 0.72 (SD=0.28). This difference caused by articulation direction is significant for both groups, being more noticeable for visually impaired participants (W=190, $p<0.01$) than for sighted participants (W=227, $p<0.05$). However, there is no significant difference when comparing each group against the other, meaning visually impaired participants are as consistent amongst each other as sighted participants. If we take the gestures from both groups all together, the results are similar with an agreement rate of 0.54 (SD=0.26) considering articulation direction, and 0.72 (SD=0.27) not considering it.

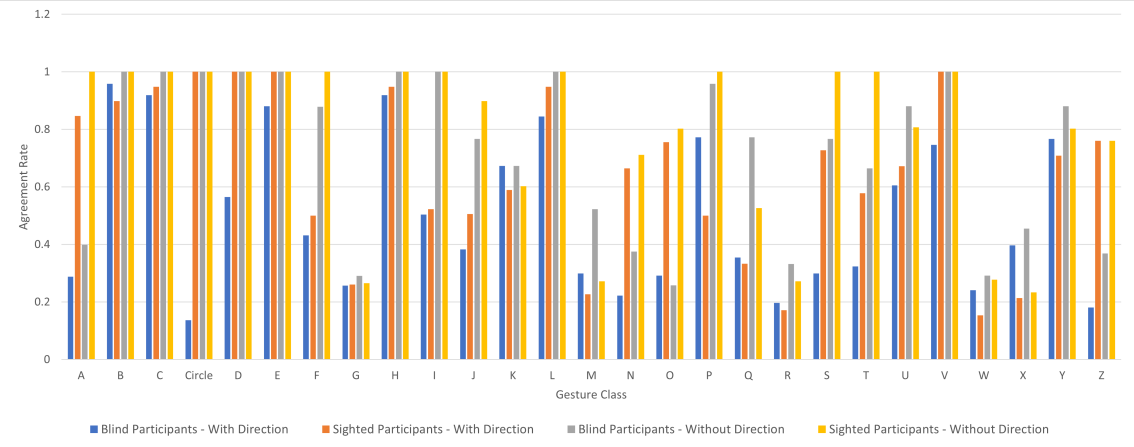


Figure 4.5: Agreement Rate per gesture class for each group of participants (with and without accounting for articulation direction)

Analysing each participant individually, we find a degree of agreement of the gestures

produced to be overall higher than the agreement rate measured between the groups. This is in agreement with the findings from prior studies [1, 34, 5] that mention users to be highly individual and internally consistent. For sighted participants, accounting and not accounting for articulation direction, the agreement rates averaged 0.83 (SD=0.12) and 0.92 (SD=0.07) respectively. Visually impaired participants averaged an agreement rate of 0.78 (SD=0.11) considering articulation direction and 0.89 (SD=0.06) not considering articulation direction. However, this difference between the groups' agreement rate and individual consistency is only significant for visually impaired participants and when accounting for articulation direction.

4.2.4 Interaction Modalities

Other than the gestures conceived, we also collected participants' opinions regarding interaction mechanisms for a possible finalized version of a system using this approach. As previously stated, participants were asked to select from a list - *double-tap, single two-finger touch, long press*, and *draw a circle* - or to suggest a new interaction modality to perform a word confirmation and a word deletion action.

To perform a confirmation action, i.e., to add a space or a full stop after a word, the majority of participants indicated a *double-tap* as their preferred way. Several participants supported this choice by saying it is what they are used to using when interacting with touch-screen devices. Other than the options listed, participants also suggested a *two-finger double-tap*.

As a way to perform a deletion action, i.e., to delete an element from the composed text, participants chose a *single two-finger touch*. They also suggested a *triple-tap*, *hand covering the screen*, and *drawing a triangle* beside the options listed.

4.2.5 Device Feedback

Another aspect we also intended to understand was whether the feedback used was adequate and could be kept for further stages of the study.

Regarding haptic feedback, we questioned participants on whether the vibrations were perceptible in every case, whether the vibration patterns used in each situation were appropriate and if the intensity was properly calibrated. The response was positive, with all participants accepting the haptic feedback employed.

Since at this stage we only used audio feedback to reproduce the characters requested, our main concerns were the volume and perceptibility of the output. Even though the audio was at max level, there were not many complaints since the study occurred in a controlled and quiet environment. However, the characters' output using Google's Text-ToSpeech Engine was sometimes imperceptible. Participants mentioned having significant difficulties distinguishing characters with similar sounds like "N" and "M" or "P"

and ”B”. The efforts made to mitigate this issue were somewhat effective nonetheless, the only reliable way to overcome this adversity was to orally repeat the characters to the participants after they were announced.

4.3 Discussion

This first study aims to help understand people’s perception and conceptualization of *Braille Shapes* by collecting and analysing their gestures and additional information, which serves as the foundation for the remainder of this work.

To properly understand and contextualize our findings, we first need to know our study’s population. It consists of sighted and visually impaired people, with the latter being significantly older, averaging ages between 36 and 45, leaning more towards the older side of that interval. Even though visually impaired participants reported a 4 out of 5 regarding being at ease with touchscreen devices, this does not transpire when performing the study’s procedure. It is especially noticeable when compared against sighted participants and is hypothesized to be due to the minimal use participants reported to give their touchscreen devices, mostly making phone calls, messaging and checking the time.

Compared with other studies [42, 41, 35, 15], our population can be considered older and less tech-savvy, making their interactions with the touch devices a bit more cumbersome, leading in some cases to the inability to complete the tasks requested. The participants unable to complete the procedure presented similar reasons for it to happen: *uneasiness when interacting with the smartwatch*, *lack of hand steadiness*, and *inability to envision different shapes for each letter*.

While performing the study, it also came to attention that their Braille proficiency is below the self-reported. Throughout the study, several participants forgot the Braille cell coding of several characters, which had to be reminded by the investigator present.

When understanding how participants conceive and perform the gestures, similarly to prior work, we can see that visually impaired participants did it somewhat differently from sighted participants. In a general way, visually impaired participants produced gestures overall bigger, slower, and varying more in size. They also required more time to think of *Braille Shapes* for each character, which correlates to what we consider part of the complexity of a Braille cell, i.e., the number of dots and their position relative to each other. The higher the complexity, the more time it took participants to devise a shape. The duration participants took to produce the gestures correlates as well with the complexity of the Braille cell.

To compare the actual shapes devised by the participants, we did so considering and not considering articulation direction. For both cases, visually impaired participants produced more variations per gesture class (i.e., per character), however, there is a significantly higher impact in the number of representations when accounting for articulation

direction. Sighted participants showed to be more consistent than visually impaired participants when devising shapes, however, when combining the two groups, the consistency measured is similar to each group separately. Individually, sighted participants showed to be more consistent. Nonetheless, participants from both groups showed higher individual consistency than when in a group. In all cases, articulation direction significantly negatively impacted the measurements.

Even with the highlighted differences between groups, there are slight similarities in which characters have the least number of variations in both groups. It is also noticeable that visually impaired participants start their gestures mostly from the top, and in one particular occurrence, a participant devised the shapes considering the Braille cell to be vertically inverted.

Considering all the above-mentioned aspects can help idealize some guidelines for future development:

Pre-defining some of the gestures with the least number of variations. As mentioned, the characters with the least amount of different representations were similar among visually impaired people. Defining how they are performed can help reduce possible ambiguity, with a low probability of not going over all the possible cases. This way, if opting for a template-match-based approach, ”perfect templates” can be used for those specific characters, and disambiguation mechanisms can also be applied.

Using gestures' starting points as a recognition feature. Since visually impaired participants started most of their gestures primarily from the top of the touch surface, knowing which gestures can help differentiate from the gestures started elsewhere.

Opting for user-dependent recognition approaches. With participants being individually more consistent than in a group, opting for tailored solutions like user-dependent recognizers can provide the best results.

Providing a mirroring option to allow people to perform gestures invertedly. With visually impaired people knowing how to write using a slate and stylus, they may prefer to think of the gestures the same way they write them, vertically inverted. Providing an option to leverage that, would reduce any extra cognitive load associated with the use of the system.

In the end, we believe that the number of sessions and iterations held was not enough to allow participants to adjust and become more comfortable with the system, especially since they were interacting with a type of device they were not used to. With each session, participants showed signs of improvement and increased ease with the method.

Chapter 5

Automatically Recognizing BrailleShapes

Conceiving a system relying on gestures as an interaction modality requires an appropriate recognition mechanism. Aiming to explore a broader range of options and not focus solely on the data analysis findings, we evaluated several recognition approaches mentioned in section 2 as having the best results and being the most adequate for our end goal. This way, we can ensure we explore how the differences described above in the gestures produced by visually impaired and sighted participants affect recognition accuracy and influence our decision regarding the type of recognition to employ.

5.1 Template-Matching

We started by testing our data on some of the recognizers from the ”$-family”. This family of recognizers is meant to be easy to implement for rapid prototyping of gesture-based user interfaces and has shown great results in stroke-gesture recognition. These recognizers use template-based matching algorithms for gesture recognition, meaning each candidate gesture is compared against a predefined set of templates provided until the one with the most similarities is found. These templates serve as a blueprint for the gestures in question however, in this case, participants did not always produce them in the same manner. Some participants did not conceive the gestures for each class the same way in every iteration, which can prove to be a hindering factor for recognition accuracy in this type of recognizer. The same can be said about the previously highlighted issue regarding ambiguous representation between and within participants.

We tested the $P [38], $P+ [37], and $Q [39]. The $P was the first in the family to consider gestures as clouds of points allowing it to overcome some limitations of prior members that reasoned about the gestures as a chronological order of points drawn. This makes it a direction-invariant recognizer and provides more freedom when producing the gestures. Then, focusing on improving accessibility in touch screens for visually impaired people, $P+ was developed as an improvement from $P. It provided slightly better accu-

racy with reduced execution times. Superseding it, however, is $Q, a recognizer optimized for low-power and mobile devices. It is built upon $P, providing utterly faster recognition speeds while slightly improving its accuracy. It is the $-family's most performant recognizer so far and is also intended for devices with low capabilities making it the optimal choice for our use case. Nonetheless, we tested the other two recognizers mentioned to have a point of comparison.

All these recognizers are invariant to sampling, scaling, translation, and articulation, meaning the strokes can be performed anywhere on the touch surface and with any size, in whichever direction desired, and disregarding the sampling rate of the device. They are only susceptible to rotation which is beneficial for this case in particular since slight changes in the angle of the shape produced can correspond to an entirely different character.

5.1.1 Evaluation Methodology

We measured all recognizers' user-dependent and user-independent accuracy and classification times as follows.

For user-dependent recognition rates, we selected T samples as templates for each gesture class from a participant P, and one additional sample from P, different from the first T, as the gesture to be classified. Due to the diversity of representations per gesture class, T is always the total amount of samples from the participant in question minus the one used as a candidate gesture. We made sure to use all the gestures produced by the participant as a candidate to go over over as many cases as possible. This process repeats for each participant (sighted and visually impaired), and the recognition results are averaged and expressed as a percentage in table 5.1 and classification times in table 5.3.

As mentioned in the literature, the optimal number of templates for the dollar family recognizers is around 8 samples per gesture class. Since we only collected a maximum of 3 per user, we also measured user-dependent recognition rates by applying some form of augmentation to the data to have 8 samples per gesture class. With these recognizers being susceptible to rotation, we augmented the data by applying slight rotations to the gestures that would not change the shape completely.

To measure user-independent recognition rates, we used T samples from a group G of participants as templates to match against, and all of the samples from an additional participant P different from any in G as candidate gestures. This process repeats for every participant, and whether P is sighted or visually impaired, G is always composed of all the remaining participants of a single group (i.e., sighted and visually impaired), and T is always equal to the total amount of gestures of each group. Even though the optimal number of templates is around 8 samples per gesture class and using all of the samples as templates reduced performance, it allowed us to measure the max recognition rate

possible for each participant. Results are averaged and expressed as a percentage in table 5.1 and classification times in table 5.3.

For both user-dependent and user-independent recognition rates, we also measured the accuracy of the top-k results produced by the recognizers to understand the potential of providing options at a character level, and it could also prove beneficial for spellchecking. However, these recognizers are only meant to produce a single outcome which meant the k fluctuated and did not guarantee that the correct suggestion was present. Results were averaged and presented as percentage in table 5.2.

5.1.2 Benchmark Devices

To properly measure and understand the recognizers' performance and usability in real-world use cases, we performed these evaluations on a real device instead of an emulator. We used the same device as in the data collection study, a 44 mm Samsung Galaxy Watch4 with Google WearOS. When this study took place, this was the latest and most advanced Android-based smartwatch with a dual-core 1.18 GHz Cortex-A55 processor and 1.5 GB of RAM. The implementation of the $-family recognizers was done on Android Studio and fully in Kotlin.

5.1.3 Results

Tables 5.1 and 5.2 show user-dependent and user-independent recognition rates for every recognizer tested and for every case (e.g., using data from different participant groups as training data). When analysing the accuracy of those recognizers, no significant difference is shown between them in any evaluation case, with all of them presenting similar recognition rates for each. The same can not be said regarding classification times described in table 5.3, in which $Q presented significantly quicker results averaging less than a second for every evaluation case. Hence, the following results described are from testing the $Q recognizer only.

User-dependent recognition rates were significantly lower for visually impaired participants than for sighted participants ($W=237$, $p<0.01$), averaging 55.1% and 71.2% respectively. The same can be seen when evaluating the accuracy of the top-k results, where visually impaired participants averaged 73.3% while sighted participants 81.6%, also presenting a significant difference ($W=209$, $p<0.05$).

When measuring user-dependent accuracy with augmented data (8 templates per gesture class), however, no significant difference was found between the two groups, not even when analysing the top-k results. Visually impaired participants averaged a recognition rate of 84.6% and sighted participants a rate of 85.3%, and for top-k results 91.9% and 92.2% respectively.

User-independent recognition rates were also significantly lower for visually impaired

participants than for sighted participants (W=282, $p<0.05$), respectively averaging 43.9% and 66.5%. The same goes for recognition rates of the top-k results with 59.9% accuracy for visually impaired participants and 79.5% for sighted participants. When comparing the two main conditions, i.e., using sighted participants' data or visually impaired participants' data as training data, our findings go according to prior work highlighting the first condition to show better results than the second. Each group presented recognition rates significantly higher for the first condition (W=0.96, $p<0.05$). Sighted participants averaged accuracies of 72% with sighted data as training data vs 61% with visually impaired participants' as training data, while visually impaired participants presented an average of 46.8% vs 40.9% respectively. Again, the same can be seen in the top-k results.

Overall, we can see that user-dependent recognition rates were mostly higher than user-independent recognition rates in every condition, specially with augmented data which showed better results than the rest.

Table 5.1: User-dependent and user-independent recognition rates for both participants groups of $P, $P+, and $Q.

	***vs* Self**	***vs* Self Augmented**	***vs* Sighted**	***vs* Visually Impaired**
$P				
Shapes by Sighted	72.0%	84.9%	70.7%	60.0%
Shapes by Visually Impaired	54.2%	83.6%	45.6%	40.3%
$P+				
Shapes by Sighted	74.8%	87.9%	72.6%	63%
Shapes by Visually Impaired	57.7%	90.1%	48.2%	43.1%
$Q				
Shapes by Sighted	71.2%	85.3%	72.1%	60.5%
Shapes by Visually Impaired	55.1%	84.6%	46.8%	40.9%

Observations: Each gesture was tested against (*Self*) the participant's other gestures, (*Sighted*) gestures from sighted participants, and (*Visually Impaired*) and gestures from visually impaired participants. Table cells report mean recognition accuracy as percentage.

5.1.4 Spellchecking

To further increase the accuracy of the output produced and to help mitigate the ambiguity-related issues, as mentioned in the previous chapter, we evaluated a spellchecker on the gestures recognized. To do so, we grouped gestures from isolated characters to form sentences and then passed them to the spellchecker. Similarly to prior work [20], the phrases

Table 5.2: User-dependent and user-independent top-k recognition rates for both participants groups of $P, $P+, and $Q.

	***vs* Self**	***vs* Self Augmented**	***vs* Sighted**	***vs* Visually Impaired**
$P				
Shapes by Sighted	82%	91.8%	82.7%	74.2%
Shapes by Visually Impaired	72%	91.5%	60.1%	56.3%
$P+				
Shapes by Sighted	84%	93.9%	84.8%	78%
Shapes by Visually Impaired	74%	95.2%	64.1%	59.9%
$Q				
Shapes by Sighted	81.6%	92.1%	82.5%	75.8%
Shapes by Visually Impaired	73.3%	91.9%	62.2%	57.6%

Observations: Each gesture was tested against (*Self*) the participant's other gestures, (*Sighted*) gestures from sighted participants, and (*Visually Impaired*) and gestures from visually impaired participants. Table cells report mean top-k recognition accuracy as percentage.

Table 5.3: User-dependent and user-independent classification times (ms) per gesture for both participants groups of $P, $P+, and $Q.

	***vs* Self**	***vs* Sighted**	***vs* Visually Impaired**
$P			
Shapes by Sighted	84	1770	1954
Shapes by Visually Impaired	126	1810	2100
$P+			
Shapes by Sighted	43	721	719
Shapes by Visually Impaired	65	780	810
$Q			
Shapes by Sighted	8	112	109
Shapes by Visually Impaired	14	120	118

Columns observations: Each gesture was tested against (*Self*) the participant's other gestures, (*Sighted*) gestures from sighted participants, and (*Visually Impaired*) and gestures from visually impaired participants. Table cells report approximate mean classification times in milliseconds.

selection process was aware of MacKenzie's phrase set and its challenges for this case. This study's participants only understood Portuguese and no other similar phrase sets for

this language were found, despite the efforts. Therefore, the three phrases conceived in prior work were used and can be seen in figure 5.1. These phrases were devised containing the seven most frequent letters of the Portuguese vocabulary and to resemble sentences that people would use in the context of short reply messaging. Since our data set contains no punctuation or accents, the same goes for the phrases used.

Due to time constraints, conceiving a Braille-aware spellchecker was deemed unfeasible and intended for future work, so in this case, we resorted to the Android Spell checker framework. However, it showed to be inadequate for our needs. The spellchecker did not provide sufficient improvements in the outcome frequently producing unwanted suggestions and sometimes not even providing suggestions at all.

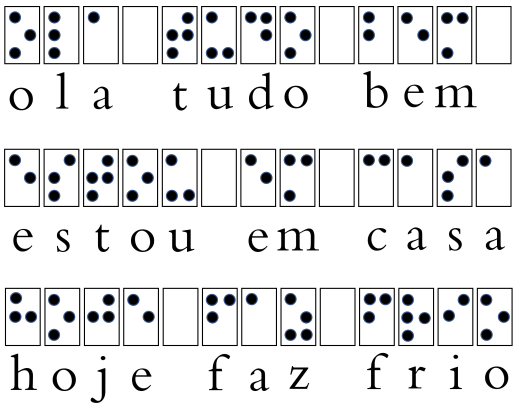


Figure 5.1: Braille cells for spellchecking sentences in Portuguese. Sentences equivalent to “Hi, how are you?”, “I’m home” and “It’s cold today”

5.2 Image Recognition

Using image recognition for handwriting classification tasks is another popular strategy at present, providing outstanding results. As mentioned, deep learning is used in numerous areas - including the field of computer vision - thanks to its diverse range of applications, and its biggest propeller has been Convolutional Neural Networks (CNN). Tools like TensorFlow Lite make it even more feasible to use these techniques in mobile contexts and devices with low resources, making it a potential solution for our system’s recognition approach.

A CNN is a class of Artificial Neural Networks commonly used in visual imagery analysis. Its models are composed of a combination of 5 different layers that usually present the architectural structure seen in figure 5.2. There is an input layer containing the image data to work on, a group of hidden layers, and an output layer. The group of hidden layers starts with a convolutional layer, also known as the feature extraction layer since it extracts features from the inputted image. It is accompanied by an activation func-

tion which is usually *ReLU*. This layer convolves the input passing its result to the next layer. It is followed by a pooling layer that reduces the spatial volume of the input image after convolution, with the most common type being *max pooling*. A fully connected layer follows to connect neurons in one layer to neurons in another layer and to map the representations between the input and the output. With both the convolutional layer's and the pooling layer's outputs being 3D volumes, they need to be flattened - by a flattening layer - to be inputted in the fully connected layer which requires a 1D vector of numbers. Finally, when performing multi-class classifications, a *softmax* layer is applied to convert a vector of N real numbers into a probability distribution of N possible outcomes. These outcomes then get matched against labels in the output layer.

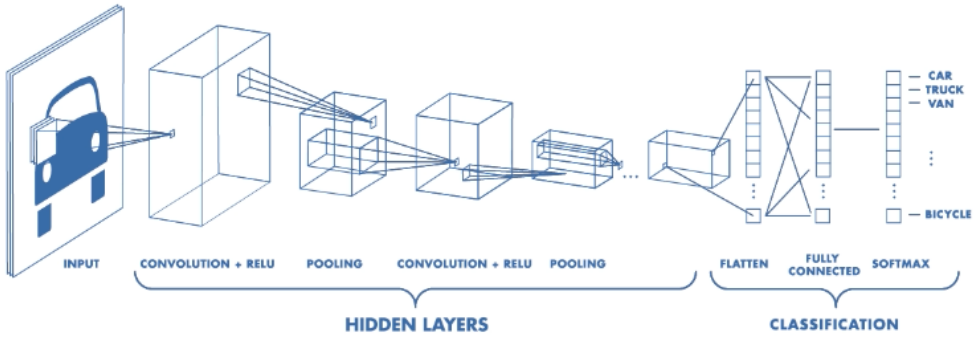


Figure 5.2: Basic CNN architecture.

5.2.1 Evaluation Methodology

In this work, we tested different CNN configurations and different approaches to achieve the best performance possible. Not discarding the findings acquired in the first study, we decided to cover the broadest range of possibilities within our time frame.

We started by training and evaluating several CNN configurations mentioned in prior work, with some adaptations to fit our purpose. We began with gestures from sighted participants to have as a baseline. After that, we tested the same configurations with gestures solely from visually impaired participants. Since sighted participants proved to have more consistent and reliable gestures, with similarities with the ones from visually impaired participants, we also tested the same CNN configurations on a combination of gestures from both groups.

For each configuration, the procedure is as follows. The gestures get converted into images, with a size equal to the one mentioned in the specific article, and the data gets split into 80% training data and 20% testing data. The testing data is divided 10% into testing data and 10% into validation data. The model trains on the training data for 50 epochs

before being evaluated on the testing and validation data. This procedure repeats for 100 trials, with the best results being presented in table 5.4. To help prevent overfitting, all data is augmented with a random zoom (with a width and height factor of -0.1) and a random rotation (with a rotation factor of 0.01 in both directions), and the model trains with an early stopping function. Layer regularizations were also applied to further attempt to mitigate overfitting.

Table 5.4: Performance of CNNs with configurations from prior work.

#	Number of Hidden Layers	Data	Maximum Training Accuracy (%)		Maximum Validation Accuracy (%)	
			Accuracy	Top-3 Accuracy	Accuracy	Top-3 Accuracy
1	4	S	93%	99%	74%	89%
1	4	VI	83%	96%	41%	59%
1	4	B	83%	97%	61%	82%
2	4	S	97%	100%	75%	89%
2	4	VI	98%	99%	40%	58%
2	4	B	98%	99%	59%	77%
3	4	S	98%	99%	77%	90%
3	4	VI	97%	97%	49%	61%
3	4	B	98%	98%	65%	87%

Column Observations: # - CNN Configuration Number
Data: S - Sighted, VI - Visually Impaired, B - Both

The first configuration has 4 hidden layers consisting of 2 groups with a convolutional layer and a pooling layer. The convolutional layers consist of 32 and 64 filters respectively, with a kernel of size 3x3 pixels and the *ReLU* activation function. Both pooling layers are Max Pooling layers and have a pool size of 2x2 pixels. The hidden layers are followed by a dropout of 25%, a flattening layer, and a fully connected layer with 128 neurons. Finally, there is another dropout of 50% before the final fully connected layer (i.e., the output layer). In the second configuration, the number and composition of the hidden layers are the same as in the previous one. The hidden layers are followed by a flattening layer and 2 fully connected layers without any dropout. For the third configuration, the structure of the hidden layers is also the same, however, it is only followed by a fully connected layer with 12544 neurons and an output layer.

In every configuration, the final fully connected layer, i.e., the output layer, has 27 neurons, 1 for each gesture class. Since it is a multi-class classification task, it uses a

softmax activation function.

With results far below the ones achieved in prior work and not better than the previously tried approaches, the same above-mentioned procedure was applied to a custom CNN configuration devised by tuning hyper-parameters, i.e., parameters that define the model's architecture. This technique allows trying different combinations of parameters to help make design choices in search of the optimal model architecture. It can be applied using a Tuner which changes the parameters with options provided, and in our case, was performed using the KerasTuner framework.

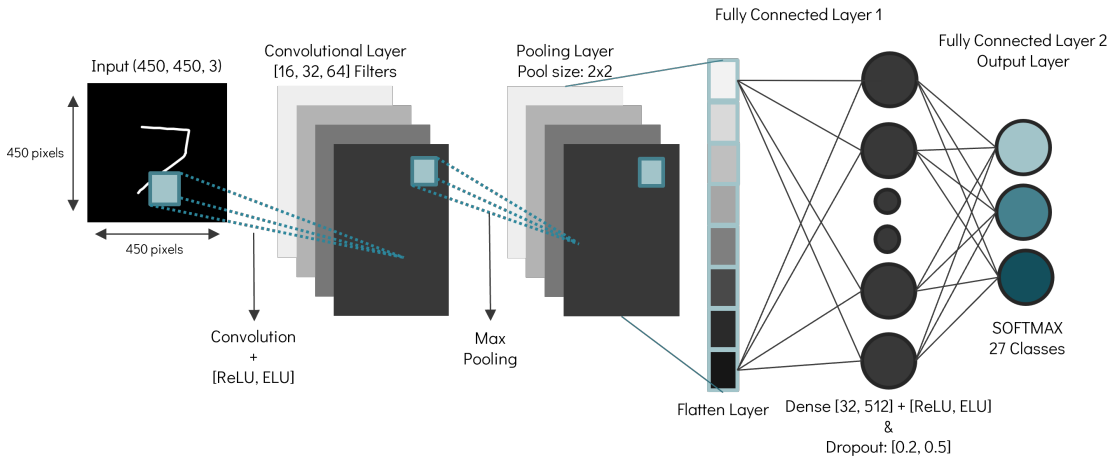


Figure 5.3: Custom CNN configuration.

The general architecture of the model is presented in figure 5.3, and similarly to the ones mentioned prior, consists of an input layer, groups of hidden layers, and an output layer. All the parameters tweaked by the tuner are also present in the figure, alongside the values they can take. Even though our input data are images sized 450x450 pixels, they are resized to 60x60 to reduce processing time, however, this did not affect recognition accuracy. To make the image data more suitable for a neural network, it is also normalized by a processing layer. The number of hidden layer groups is one of the parameters changed by the tuner ranging from 1 to 3, however, the groups always consist of a convolutional layer followed by a batch normalization layer (to make training faster and more stable) and a pooling layer. The activation function used with the convolutional layers can be *ReLU* or *ELU*, the kernel size is always 3x3 pixels, and the number of filters used varies between 16 and 64. The pooling layers always have a pooling size of 2x2 and are of the type *Max Pooling*. The hidden layers are followed by a flattening layer and by the first fully connected layer. This connected layer has units ranging from 32 to 512 and an activation function that can also be *ReLU* or *ELU*. Layer regularizations were applied to help prevent overfitting. Before the final fully connected layer or output layer, there can be a dropout. For the cases where the tuner opts to use a dropout, its values range from 0.2 to 0.5, with steps of 0.05. Finally, the output layer has 27 neurons, 1 for each gesture

class, and uses a *softmax* activation function.

With this approach, we achieved the accuracies described in table 5.5 as well as different model configurations depending on the data they were trained and tested with. The model configuration resulting from training and testing with data from sighted participants (1) is composed of 5 groups of hidden layers, each layer having 64 filters, a fully connected layer with 992 units, and a dropout of 0.5. All layers use the *ReLU* activation function. The configuration resulting from using data from visually impaired participants (2) is composed of 5 groups of hidden layers, each layer having 64 filters, a fully connected layer with 128 units, and a dropout of 0.2. All layers use the *ReLU* activation function. And finally, the configuration conceived from using data from both groups is composed of 5 groups of hidden layers, each layer having 64 filters. It has a fully connected layer with 224 units and a dropout of 0.5. Each layer used the *ReLU* activation function as well.

Table 5.5: Configuration and performance of CNN devised with a hyper-parameters tuner.

#	Data	Maximum Training Accuracy (%)		Maximum Validation Accuracy (%)	
		Accuracy	Top-3 Accuracy	Accuracy	Top-3 Accuracy
(1)	S	93%	100%	72%	96%
(2)	VI	74%	93%	49%	75%
(3)	B	83%	98%	64%	88%

Column Observations: # - CNN Configuration Number
Data: S - Sighted, VI - Visually Impaired, B - Both

Another approach evaluated was the use of transfer learning in different models. Transfer learning is a technique in which models that are pre-trained on larger data sets developed for a specific task get reused as a starting point for a model for another task. This technique allows improved performance and rapid progress when modelling the second task. First, we evaluated general image recognition purpose pre-trained models provided by the TensorFlow community. Table 5.6 describes those models and their performance. They were selected for being some of the more effective and efficient models used by the TensorFlow community. For this case, the models had to be adapted, redefining their input and output to meet our data's criteria before being trained with said data.

However, these models were designed for more general purposes. So, we tested a transfer learning approach on a model trained on a data set more similar to ours. This model was pre-trained on the Chinese MNIST, a version of a popular data set (of handwritten digits) in computer vision problems called MNIST[1]. The model is composed of 2

[1]http://yann.lecun.com/exdb/mnist/

Table 5.6: Performance of applying transfer learning to TensorFlow pre-trained models.

Model	**Data**	**Maximum Training Accuracy (%)**		**Maximum Validation Accuracy (%)**	
		Accuracy	**Top-3 Accuracy**	**Accuracy**	**Top-3 Accuracy**
EfficientNet V2 - S	S	76%	96%	61%	87%
EfficientNet V2 - S	VI	73%	94%	40%	76%
EfficientNet V2 - S	B	84%	97%	59%	88%
EfficientNet V2 - XL (21k)	S	92%	100%	76%	97%
EfficientNet V2 - XL (21k)	VI	54%	80%	32%	60%
EfficientNet V2 - XL (21k)	B	77%	85%	57%	88%
EfficientNet - Lite 4	S	78%	95%	52%	88%
EfficientNet - Lite 4	VI	54%	81%	40%	69%
EfficientNet - Lite 4	B	63%	88%	48%	79%

Data: S - Sighted, VI - Visually Impaired, B - Both

sets of hidden layers, which, differently from any model previously tested, are composed of 2 sequential convolutional layers and a pooling layer. Each convolutional layer has 16, 16, 32, and 64 filters respectively, all using *ReLU* as the activation function. A dropout of 0.4 follows each set, and the hidden layers are followed by a flattening before the output layer. The results achieved with this model are in table 5.7.

Table 5.7: Performance of applying transfer learning to model pre-trained on Chinese MNIST.

Model	**Data**	**Maximum Training Accuracy (%)**		**Maximum Validation Accuracy (%)**	
		Accuracy	**Top-3 Accuracy**	**Accuracy**	**Top-3 Accuracy**
Chinese MNIST Pre-Trained Model	S	88%	99%	67%	87%
	VI	67%	77%	35%	58%
	B	77%	94%	51%	76%

Data: S - Sighted, VI - Visually Impaired, B - Both

5.2.2 Results

Table 5.4 illustrates the performance of the different CNN configurations from prior work with the different data sets' configurations. All models provided similar results, with an overall training accuracy of over 90% for both single (avg=93.9%) and top-3 (avg=98.2%) results. However, in all cases there are signs of model overfitting, with validation results being far inferior, averaging 60% for single results and 76.9% for top-3 results. Validation accuracy also showed the most variation, with model 1 presenting the lowest value of 40% using visually impaired participant data and model 3 the highest value of 77% with data from sighted participants, both cases for single results accuracy. This variation is also visible in the top-3 results.

Table 5.5 shows the results of devising a CNN by tuning its hyper-parameters. This approach achieved training accuracy for single results averaging 83.3%, and top-3 results averaging 97%. Validation results were also significantly inferior to training averaging 61.7% and 86.3% for single and top-3 results respectively. The maximum validation accuracy achieved (72%) was using sighted participants' data and the lowest (49%) was using visually impaired participants' data.

Regarding transfer learning approaches, results were similar, whether using more general-purpose models or one pre-trained on a more similar data set to ours. Table 5.6 shows the results obtained in the first case with training accuracies, averaging 72.3% for single results and 90.6% for top-3 results, and average validation accuracies of 51.6% and 81.3% for single and top-3 results. Both the highest (76%) and the lowest (32%) validation accuracy values obtained were with the EfficientNetV2 - XL model using data from sighted and visually impaired participants respectively.

The results stemming from the model pre-trained on the Chinese MNIST averaged a training accuracy of 77.3% for single results and 90% for top-3 results and validation accuracies of 51% and 73.7% for single and top-3 results. Again, we can see model overfitting with validation results being significantly lower.

5.3 Discussion

In the process of analysing the recognizability of the shapes collected and aiming to find the recognition mechanism that best suits our system and our needs, we explored several recognition approaches that can be divided into two main categories: *template-matching-based* and *image recognition based*. In doing so, we also approached some of the findings from the first study (chapter 4).

The template-matching approaches evaluated were from the $-family - $P [38], $P+ [37], and $Q [39] - due to their positive results in prior work and design philosophy. Overall, they provided similar recognition accuracy for single and top-k results, with higher values for user-dependent recognition for each participant's group, achieving a maximum

accuracy of 74.8% and 57.7% with visually impaired participants' data. This is in line with the findings from chapter 4, which mentions people as being more internally consistent and hints at the advantages of using a user-dependent recognition approach. We managed to improve on these results by augmenting the data used, however, this method is not reliable for further testing with participants. Since it only applies rotation to the gestures, it does not replicate different possibilities of articulations for each shape. With only an average of three samples per gesture class per participant, the same participants would likely perform a different gesture variation from those collected. Regarding classification speed, $Q was by far the most prolific, being the only one to provide acceptable results for further real-life testing. It was expected since it is developed for low-resource devices and had shown the best results among the three in prior work.

Another result supporting our findings is the recognition accuracy obtained using gestures from sighted participants as templates. Even though it provided worse results than using the participants' gestures, it still provided better results than with gestures from visually impaired participants.

As a measure to counter ambiguity-related accuracy decrease, we also tested a general-purpose spellchecker which did not provide acceptable results, leaving open the question of whether to use a Braille-based spellchecker to improve the method's recognition accuracy.

We evaluate image recognition approaches by testing different techniques and model configurations, whether based on the literature or by altering each model parameter individually. To start, we tested various configurations from prior work. With the results below expected, we attempted a custom model configuration devised by tuning each hyperparameter individually using a tuner. Since this approach did not provide results at the desired level, we turned to the transfer learning technique using both general-purpose pre-trained models and a model pre-trained on a data set more similar to ours.

In any case, there was some overfitting of the models evaluated, regardless of our attempts to mitigate such issues. It led our validation accuracy to be much lower than our training accuracy achieving a maximum value of 77% for single results and 90% for top-3 results using one of the model configurations from prior work.

Due to the difficulties presented by participants in using the device and the time constraints that disabled them from further adapting to it and our approach, based on the results gathered, we deemed any of the approaches as being unfeasible for further user testing.

Chapter 6

Exploring BrailleShapes for a Text Entry Approach

To properly validate the use of *Braille Shapes* as the basis for any text entry system requires implementing and subsequently evaluating such a system. With prior results regarding recognition accuracy being below what is acceptable to perform said validation, we only performed a preliminary one without the live recognition aspect. It serves as a proof of concept and helps us understand the validity and how well people adapt to this kind of approach. This chapter describes the validation performed and is divided in two main sections.

The first section – 6.1 – describes the data collection process detailing its different stages and requirements, the study's population and the data collected.

The second section – 6.2 – presents the results of a performance analysis of the preliminary version of the system. It also shows a recognition evaluation using such data.

6.1 Data Collection & System Exploration

Similarly to the first user study, we conducted another one to collect gesture data and information. With the system aiming only to be used by visually impaired people, this time, however, they were our only target population. This study focused not only on collecting more data from the participants but doing so in a more realistic usage scenario, allowing us to grasp how participants perceived and interacted with the system.

6.1.1 Participants

Our population comprises some participants from the previous study, which helps minimize the practice and habituation required, maintains some degree of consistency, and allows using their previous respective gestures as elements for recognition analysis.

We recruited 7 visually impaired participants (5 self-identified females and 2 self-identified males) who we thought to be more at ease with the device, our approach, and

the concept of a gesture-based system using *Braille Shapes*. They were again recruited through the institution Fundação Raquel e Martin Sain and via word of mouth.

These participants averaged between 46 and 55 years old (SD=1.27) and were all right-handed. Of those participants, three had partial vision, while the rest were fully blind. All of them use a smartphone daily and three also use a smartwatch regularly. Self-reportedly, they were relatively at ease with touchscreen devices, averaging 4.1 on a 5-level Likert scale. However, they reported only performing a few simple actions like contacting others (mostly via phone calls or audio messages), looking at the time and getting notifications. This applies to both devices mentioned. It can still be considered an aged and less tech-savvy population compared to other studies.

Table 6.1: Visually impaired participants' collected information

#	Age Group	Age of Blindness	Literacy	Braille Proficiency[1]	Used typing method(s)[2]
1	24 - 35	0	Secondary Education	5 / 4	QWERTY (touch typing mode)
2	46 - 55	23	Higher Education	4 / 4	QWERTY (standard mode)
3	56 - 65	44	Secondary Education	4 / 5	Dictation
4	24 - 35	0	Primary Education	5 / 5	QWERTY (touch typing mode) & Dictation
5	36 - 45	0	Primary Education	5 / 4	QWERTY (standard mode) & Dictation
6	46 - 55	28	Secondary Education	5 / 4	QWERTY (standard mode)
7	56 - 65	0	Higher Education	5 / 5	QWERTY (standard mode) & Dictation

[1] Reading / Writing - Self-declared in a Likert scale of 1 to 5
[2] All participants that used QWERTY but participant 9 used TalkBack or VoiceOver

6.1.2 Requirements

For this study, we required participants to possess Braille knowledge and preferably regular contact and ease with touchscreen devices. These were deciding factors when selecting participants from the ones participating in the previous study. Table 6.1 details this information, with participants identified as numbers from 1 to 7 to preserve anonymity.

6.1.3 Apparatus

To perform this study, we used a Google Pixel 4a running Android 12 and a 44 mm Samsung Galaxy Watch4 with Google WearOS, both devices running the custom applications mentioned in section 3.6.2 devised for the study, with participants only interacting with the smartwatch device and application.

The "host app" on the smartphone allowed the investigator in charge to control the study's procedure by communicating with the smartwatch application and managing its execution flow. It allowed the investigator to opt between two tasks - collecting characters or sentences - and carry out each procedure accordingly. Regardless of which procedure it was, the investigator could go through the various characters or sentences one by one, both backwards or forwards, change the smartwatch's application state and make it reproduce a specific character or sentence via TextToSpeech. The application dealt with all processing and storage of any gesture data received from the smartwatch, allowing its visualization in real-time and after all data collected. Its layout configuration is displayed in figure 3.4

On the smartwatch side, for the characters' collection task, the application used was similar to the one used in the previous study. For the sentences collection task, however, other than collecting gesture data, the inputting application had editing features that allowed participants to delete characters and confirm words and sentences. It had a simplistic interface with a single empty screen whose state was controlled by the investigator. It possessed audio and vibrotactile feedback and safety mechanisms to prevent undesired interactions.

6.1.4 Procedure

The procedure started with an introduction of the investigators and a presentation of the study, explaining its purpose and providing an overview of its procedure. Since our participants all took part in the previous study, they were already familiarized with the concepts approached and the devices used. Nonetheless, they still had a brief moment to adjust to the smartwatch device. In a scenario in which participants did not participate in the first study, a more detailed explanation is provided, and characterization questionnaires are performed.

After the introductory part, we briefly allowed participants to train for around 5 minutes. Similarly to the first user study, the training session consisted of participants inputting a small set of characters from the alphabet by drawing their corresponding *Braille Shapes*. We asked them to "envision the shape resulting from drawing a single stroke that would go over all the raised dots in its corresponding Braille cell", and they had to complete it without passing over any of the dots more than once. Those were the only restrictions imposed on the participants. Starting or ending the gestures at the edge of the smartwatch was also mentioned as not the best practice to mitigate situations where the

device did not detect any interaction, not recording the gestures performed. For this particular session, we used the letter - ABDGKMORUXZ - thought to go over most scenarios (e.g., a single dot, ambiguous representation, missing dots in the middle row, etc.). Other than the letters mentioned, we asked participants to input a circle as in print writing and not in Braille. We did not explain what the intention behind such a request was, however, we collected this shape for recognition and interaction purposes and further enrichment of our data set.

At the start of the session, the smartwatch was in a *blocked* state, which was controlled by the investigator on the ”host app”. The system audibly presented participants with random letters, one at a time, which participants had to confirm to understand. After that, the investigator would *unblock* the smartwatch application, and participants would perform the gesture corresponding to the character requested. The smartwatch application played a beep-like sound alongside haptic feedback whenever the screen was touched, to help participants guide themselves on the touch surface. If no situation arose requiring the repetition of the gesture, it was accepted and the data was sent to the smartphone application, and the smartwatch application would go back to a *blocked* state. This process is repeated for each of the aforementioned letters or during the 5 minutes.

The actual collection session is divided into two phases, a character and a sentence collection phase. The character collection phase is similar to the training session, with participants having to input the 26 letters of the alphabet and a symbol (i.e., circle shape) a total of 2 times each. This time, no sound is emitted at the touch of the screen. Even though no time limit was imposed, we tried to limit the session to 30 minutes maximum so the study would not go over an hour and a half per participant.

In the sentences collection phase, the procedure, even though similar, focused more on a realistic use case scenario. We asked participants to input 3 full sentences by entering each character at a time, allowing them to delete undesired gestures and insert *spaces* between words. At the end of each sentence, they also had to *enter* the sentence to confirm it. Other than reading the sentences, the system also provided audio feedback for each of the actions performed (e.g., ”Ok” for sentence confirmation).

We used a Portuguese phrase set following MacKenzie’s methodology, with a corpus of Portuguese proverbs adapted to this context and the challenges found. The phrase set only comprised sentences with around five words for a more agile procedure, allowing better memorization of such items, and resembling a more probable scenario in which users can make sense of a more likely interaction. Since the gestures collected did not include any punctuation, capital letters, or accents, neither did the phrase set.

We finished the study by inciting participants to provide feedback on the possibility of a system using this type of approach, to understand whether it would be too complex and complicated, whether they would use it and whether they would quickly adapt to it. We did this using an adaptation of the SUS questionnaire.

6.1.5 Data

All gesture data collected was stored in XML files structured differently depending on whether it was characters or sentences.

For character gestures, we created a file per gesture class containing all the instances of that class for that user. The gesture information comprises the corresponding character and the gesture's strokes. Each stroke is characterized by an index and contains points with X and Y coordinates and a timestamp.

The sentences were stored in an XML file per sentence, identified by the sentence name, and containing gesture information for each of the characters in that sentence. The gestures are made up the same way as in the individual gestures' files, including the *space*, *delete*, and *enter* gestures.

Collection data like starting and finishing times are also present in any of the files.

The questionnaires regarding demographic data and technological and Braille proficiency were created and performed using Google Forms.

6.2 Data Analysis and System Evaluation

6.2.1 Gestures Collected

This study had two separate collection phases, one for collecting gestures individually and the other for groups of gestures composing sentences.

In the first phase, each participant inputted 2 *Braille Shapes* for each of the 26 letters in the English alphabet plus 2 "circle" shapes as in print writing, making a total of 27 * 2 * 7 = 378 individual gestures.

For the second phase, we requested each participant input 3 sentences with a random amount of characters. We recorded 21 sentences, adding up to 448 gestures after transcription, not accounting for the editing features gestures (e.g., space, delete).

All data were deemed usable and considered within our requirements.

6.2.2 Evaluation Metrics

As mentioned, we performed a preliminary evaluation of the system after all data were collected. We measured and analysed several text entry metrics to understand its performance in a more realistic scenario.

To measure the text entry performance of our approach, we calculated and analysed four key metrics: *words per minute* and *gestures per second* to measure entry rates, *gestures per character* and *minimum string distance* for error rates. These metrics are widely accepted in text entry evaluation and are presented as defined by MacKenzie et al. [21].

Words Per Minute (WPM)

Words per minute is one of the most commonly used metrics in text entry evaluation and is used for measuring typing speed. Conventionally, a word is regarded as 5 characters, including spaces. This measure only considers the length of the resulting transcribed string and the time it took to produce.

$$WPM = \frac{|T| - 1}{S} \times 60 \times \frac{1}{5} \tag{6.1}$$

It is defined in Eq. 6.1, where T is the final transcribed sentence and $|T|$ its length. In this case, T only contains letters and spaces. The S term is seconds, and it is measured from the entry of the first character until the entry from the last, counting with backspaces. The value of 60 is seconds per minute, and the value of 1/5 is words per character.

Gestures Per Second (GPS)

Another metric that allows us to know the speed at which users are performing inputting actions is the measure *gestures per second*. It is considered the ”action rate”, with the gestures being an atomic action taken during the text entry process and any unproductive action as a nonrecognition. In this case, it is considered any individual stroke used to compose the *Braille Shapes* and the interactions required to perform text entry actions.

It is defined as:

$$GPS = \frac{|IS_{\varnothing}| - 1}{S}, \tag{6.2}$$

where $IS_{\varnothing}$ stands for the input stream including all actions, and $\varnothing$ represents a nonrecognition.

Gestures per Character (GPC)

Error rates are usually more complicated to measure than entry rates. They can be considered *corrected errors* (i.e., errors during entry) and *uncorrected errors* (i.e., errors after entry).

One way to quantify errors performed during text entry is using the measure *gestures per character* (GPC), which represents the average amount of gestures taken to input a character. The GPC conveys an expression of the method’s efficiency and accuracy, with a higher GPC value indicating lower accuracy and efficiency and a value closer to 1 implying a more efficient and accurate method.

As per the GPS, we account for a gesture as an atomic action taken during the process of text entry and any unproductive action as a nonrecognition. In this case, they are any individual stroke used to compose the *Braille Shapes* and the interactions required to perform text entry actions.

$$GPC = \frac{|IS_{\varnothing}|}{|T|} \tag{6.3}$$

Since we did not have a recognition mechanism during the data collection procedure, the errors and corrections were due to the participants forgetting to input certain characters, the system not recognizing a touch interaction, and the participant acknowledging a gesture as wrongly performed.

Minimum String Distance (MSD)

To measure the accuracy of the resultant transcribed sentence, we can use the *minimum string distance* statistic (MSD). It provides the minimum distance between two sentences regarding the number of correction operations required to go from the transcribed (T) string to the originally presented (P) one. It is a well-known algorithm in statistics and is commonly used when no error corrections are allowed.

It requires a final transcribed string resulting from recognizing the characters inputted, which in this case are the characters resulting from the gestures produced by the participants. To do so, we filtered the sentences inputted without a recognition mechanism (i.e., only used the gestures resulting from the corrections performed) and ran them through the $Q recognizer to obtain a final transcribed sentence.

To calculate the minimum string distance we are using a Java implementation available online [32] and are presenting it as error values in percentage by using Eq. 6.4. $|P|$ and $|T|$ are the lengths in characters of the presented and transcribed strings respectively. We opted out of using capitalization, punctuation, and accents which required normalization of any of the sentences used and collected.

$$MSD_{error\ rate} = \frac{MSD(P,T)}{MAX(|P|,|T|)} \tag{6.4}$$

6.2.3 Performance Evaluation

Table 6.2 presents the averaging results per participant of the performance evaluation, encompassing the values obtained for the entry speed and the error rates metrics except for MSD.

Inputting Speed

Participants averaged 3.96 WPM (SD=0.99) using this approach, with every participant showing an average increase of 0.51 WPM from the first sentence collected (mean=3.74, SD=1.29) to the last one (mean=4.25, SD=1.11), with no signs of plateauing. The best performance with this approach achieved 6.3 WPM, and the worst did 1.5 WPM.

Regarding the number of gestures per second performed to input the sentences, participants averaged 0.38 GPS (SD=0.07) gesturing less than an inputting action every second. They also showed improvements from the input of the first sentence (mean=0.36, SD=0.10) to the last (mean=0.39, SD=0.08), with the best performance achieving 0.52 GPS and the worst 0.24 GPS.

Compared to the Braille-based smartwatch approaches mentioned in table 2.1, ours provided the worst results in terms of WPM. These results are below expected and desired, nonetheless looking solely at inputting speed metrics is not enough to understand the why.

Table 6.2: Mean values of analysed metrics per participant.

#	Words Per Minute	Gestures Per Second	Gestures per Character
1	4.42	0.38	1.04
2	5.45	0.47	1.04
3	2.75	0.31	1.46
4	3.28	0.31	1.15
5	3.11	0.31	1.18
6	3.82	0.39	1.22
7	4.88	0.46	1.15
Total	**3.96**	**0.38**	**1.18**

Error Rates

Similarly to the input speed metrics, participants showed improvements regarding GPC. The number of gestures performed per character decreased from 1.18 in the first sentence to 1.11 in the last sentence, indicating a high-efficiency level. In this case, the metric does not directly convey the system's accuracy since no inputting actions were performed using character recognition.

To calculate our error rate using the MSD, we used the $Q recognizer to obtain a transcribed string resulting from input recognition as required. Even though a user-dependent approach performed the best on our recognition evaluation, we calculated the MSD with sentences transcribed with different configurations. We used gestures from each participant, whether from the first study, the second study, or both combined, to recognize their own inputted sentences only. We also used gestures from the first study, from each group of participants and from a combination of both groups, accompanied by the data from the

second study, to recognize all of the sentences.

The error rates measured using a user-dependent approach for recognition were very similar regardless of the gesture combinations used. Using gestures from the second study collected in the same study as the sentences, we obtained an error rate of 51.09% (SD=13.13%, Min=17.39%, Max=72.23%). As expected, it performed better than using gestures solely from the first study, which provided an error rate of 53.35% (SD=10.17%, Min=32.26%, Max=69.29%). The best results, however, were with gestures from both studies. We achieved an error rate of 50.82% (SD=12.00%, Min=17.39%, Max=71.43%).

With a user-independent approach, contrary to the expectations resulting from the recognition evaluation performed, results improved significantly. Similarly to prior findings, using gestures from sighted participants showed the best results achieving an error rate of 42.39% (SD=8.73%, Min=24.00%, Max=61.49%) versus 42.39% (SD=8.73%, Min=24.00%, Max=61.49%) from using gestures from blind participants, and 41.72% (SD=9.44%, Min=25.80%, Max=64.00%) from using a combination of gestures from both groups.

Unpacking Errors and Delays

During the sentences phase of the data collection procedure, the system did not comprise a recognition mechanism thus no errors were derived from wrongfully recognized characters. However, situations arose where participants needed to undo the interactions performed due to erroneous actions or took extra time besides what the gestures required.

At times, while repositioning their hands on the smartwatch or simply while in a resting position, participants accidentally touched the screen which unintentionally counted as an inputting action. Other times, participants attempted to perform editing interactions like a double-tap or a long-press and ended up miss-timing them leading to wrongfully accomplished interactions. Regardless, several participants proceeded to act on these situations by completely pausing the inputting for a brief moment to think about it, requesting the assistance of the investigator present. These situations led participants to perform multiple *delete* actions, hence increasing the time and number of gestures required to input a sentence.

Occasionally, participants forgot how to spell certain words or the *Braille Shapes* of certain characters, which required them to ask the investigator or take some extra time to think about it. One participant, in particular, would often touch the screen to start a gesture and stay that way for a brief moment before actually performing it. Whether due to inputting inaccuracies or extra cognitive load, participants spent an average of 1.74 seconds between performing each inputting interaction adding up to a total of 9.86 extra seconds per word, or 57.2 extra seconds per sentence. Assuming that these types of setbacks decrease with the use and habituation of the approach, it can be estimated an increase of more than double (approximately up to 11 WPM) the number of words per

minute currently produced with this approach.

Subjective Feedback

We gathered participants' feedback regarding the usability and likeability of the system and tried to understand its value to our target population. Participants provided their opinions by answering a questionnaire consisting of items to be ranked on a 5-level Likert scale.

They expressed being very interested in the system idealized (mean=4.6), mentioning they would like to use such a system frequently and would do so if they had a smartwatch device (mean=4.7). They thought the approach was easy to use (mean=4.6) and not unnecessarily complex (mean=1.2), with quick and easy adaptation after a short exposure (mean=4.9). They felt the way to interact with the system was intuitive (mean=4.9) and reported being very confident doing so (4.5). Compared to their usual inputting methods, they mentioned believing they were much faster.

Overall, participants were generally positive and excited about a system with this approach, with some participants commenting *"I would love to see this project going forward so I could finally consider acquiring a smartwatch device."*, and *"Using this idea even if just for shortcuts would be so much more agile than what I do right now."*. One participant, in particular, emphasized how this approach would be beneficial in several ways, mentioning some aspects we had established as motivations for this work. They also denoted the idea and the study to awaken an interest in smartwatch devices.

6.3 Discussion

With this second user study, we aimed to validate using *Braille Shapes* as the foundation for a gesture-based text entry system. We did so by developing a system with such an approach and testing it with participants in a somewhat realistic scenario. It helped us measure the system's performance and compare it against other existing methods. It also allowed us to understand some of the hurdles found by the participants when interacting with such a system, and how well this type of approach was accepted by our target population.

For this study we recruited visually impaired participants from the first user study, bearing in mind how at ease they were with smartwatch devices and how sharp their Braille knowledge and subsequent translation into *Braille Shapes* was. We did so in an attempt to leverage some previous accustoming to the approach and to ensure the most similarities between the shapes used for recognition purposes and the ones collected throughout this second user study. It was a smaller and older sample than the one from the first study, with participants' ages between 46 and 55 years old.

Using this approach, participants achieved an average speed of 3.96 WPM with a minimum average error rate of 41.39% and 0.38 GPS. Improvements throughout each sentence collected and no signs of plateauing suggest that participants would benefit from a longer exposure time to the approach and the device. Nonetheless, the results obtained were below the ones of most of the Braille-based methods reviewed in section 2 and presented in table 2.1. One aspect believed to contribute to the poor results is the time taken by participants to undo undesired actions. As mentioned, they would forget or take a while to remember how to proceed in such situations, sometimes asking the investigator for assistance. Participants would also forget which character they were currently inputting and sometimes lingered when producing *Braille Shapes*. Mitigating these issues by exposing participants to the system for prolonged periods could potentially increase the efficiency and effectiveness of the system drastically.

One thing worth mentioning is that even though participants achieved an average of 1.18 GPC thus indicating a high level of efficiency, this is not the most accurate representation since no recognition mechanism was applied, and thus gestures were only corrected when a participant produced an unwanted interaction or thought the shape to be poorly gestured.

Regardless of the results obtained from the metrics evaluated, the system was well received by our participants, indicating a strong interest in both this type of entry method as well as smartwatch devices. They thought the system to be intuitive and uncomplicated, providing a quick and easy adaptation, mentioning it allowed a faster and more interesting inputting experience than the ones they already possessed. Participants also felt this approach addressed several concerns raised when inputting text, most coinciding with the ones we presented as motivations for this work.

Chapter 7

Conclusion

One aspect, in particular, that is so paramount to our daily lives is communication. It is intrinsic to human existence and can be accomplished in various manners (e.g., using voice, gestures, technology, etc.). Advancements in technology allow for numerous new ways of performing such tasks, but these also translate into an increment of accessibility concerns. With touchscreen devices being visually demanding, it can be a constraining factor for visually impaired people. It is even more noticeable for specific scenarios that, for example, require single-hand usage or occur in movement-based contexts. We believe that devices such as smartwatches can prove beneficial in mitigating such issues and others mentioned throughout this work. However, these lack accessible alternatives for performing such a crucial task.

Our literature review shows that the already existing text entry approaches focused on addressing accessibility concerns do not usually target different types of devices other than smartphones. It not only limits the inclusion of visually impaired people to other types of devices like smartwatches but also results in a lack of standardization for those.

Recognizing these concerns motivated the development of this thesis, resulting in the elaboration and validation of the concept for a new smartwatch-accessible text entry approach for visually impaired people based on the Braille writing system.

We started by defining the concept of a Braille Shape which is the foundation for the inputting method devised. We defined a character's Braille Shape as the shape obtained from performing a single stroke, passing over all the raised dots of its corresponding Braille cell, without going over any of those dots more than once.

We then performed two user studies to collect and analyse data and validate our idea. Our first study focused on collecting Braille Shapes from sighted and visually impaired participants to understand how they perceived the concept and how they would perform such gestures. The data collected was also used to perform a recognition evaluation and help select the best recognition mechanism required by an approach like this. The second study also consisted of a data collection process, however, this time in a more realistic scenario and only from visually impaired participants. Participants were instructed to input

complete sentences using Braille Shapes, including spaces and sentence confirmations. They could delete any inputted characters when needed. We also collected participants' opinions regarding the usability of the system and some demographic data.

Our findings show that sighted and visually impaired participants gesture Braille Shapes differently and that visually impaired participants are less consistent when doing so. Nonetheless, consistency was higher when evaluating each group member individually, indicating that a user-dependent approach to a recognition mechanism holds the best potential for this type of application. We also concluded that accounting for articulation direction significantly impacts recognition, which suggests that an articulation-invariant recognizer would be the best alternative. Another aspect worth mentioning is the similarities of some of the Braille Shapes produced for certain characters in terms of their articulation and a low number of variations among participants, which can be a leverage point when considering the possibility of pre-establishing Braille Shapes for such characters.

When performing a preliminary evaluation using this approach, the values obtained for speed and error rates are not comparable to the best-performing existing methods. Even so, participants showed improvements at each stage of the study, indicating signs of improvement with prolonged exposure to this approach. Furthermore, the system proved to be well accepted by the participants showing signs of covering their main concerns regarding this type of interaction modality.

The results of this work allow us to believe that pursuing this idea is not only desirable but also feasible, with the possibility of bringing great improvements to visually impaired people's lives. Even if the utility of this approach ends up not being what initially intended (i.e., text entry), we believe it is flexible and adaptable enough for other use cases such as inputting shortcuts and commanding actions. In any case, we believe we helped complement an area of investigation for topics of great interest and importance that need to be addressed.

7.1 Limitations and Future Work

Throughout this work, the most significant limitation encountered came with the COVID-19 pandemic. It changed the timing of the user studies and restrained the recruitment of the participants.

Having a population with almost nonexistent prior interaction with the type of devices in question and whose Braille knowledge was not as solidified as reported could have negatively impacted the study's results. Nevertheless, we did our best and recruited an acceptable number of participants for the population in each study.

The time constraints inhibited a more prolonged study, thus impeding the collection of more data, which affected other stages of the study as well. We believe that more

data would prove helpful in evaluating and selecting a recognition mechanism, allowing a more comprehensive validation of the system.

Resorting to the guidelines and findings provided by this work, future work can help mitigate such limitations and evaluate the approach explored to its full potential and implementation.

Appendix A

First study's questionnaire applied to sighted participants

BrailleShapes Questionnaire

*Obrigatório

Dados Pessoais

Visto que eu já me apresentei, vou agora fazer-lhe umas perguntas em forma de questionário de forma a ficar a saber mais sobre si.

NOTA: As questões assinaladas com ** não são para perguntar ao participante

1. ** ID Utilizador *

2. Nome *

3. Idade *

Marcar apenas uma oval.

- < 18
- 18 - 23
- 24 - 35
- 36 - 45
- 46 - 55
- 56 - 65
- > 65

4. Sexo *

Marcar apenas uma oval.

- Feminino
- Masculino
- Outro / Prefiro não revelar

5. Habilitações *

Marcar apenas uma oval.

- Ensino Básico 2º Ciclo (5º e 6º anos)
- Ensino Básico 3º Ciclo (7º, 8º e 9º anos)
- Ensino Secundário (10º, 11º e 12º anos)
- Ensino Superior
- Outro / Prefiro não revelar

6. Mão Dominante *

Marcar apenas uma oval.

- Direita
- Esquerda
- Ambidestro

7. Usa Relógio *

Marcar apenas uma oval.

- Sim
- Não

8. Mão em que usa/faria mais sentido usar Relógio *

Marcar apenas uma oval.

- Direita
- Esquerda
- Ambas

9. Notas:

Tecnologia

Entrando agora numa secção mais tecnológica,

10. Qual o seu à vontade com dispositivos com ecrãs tácteis? *

Marcar apenas uma oval.

Muito pouco à vontade

1

2

3

4

5

Muito à vontade

11. Com que frequência utiliza os seguintes dispositivos? *

Marcar apenas uma oval por linha.

	Nunca	Menos de uma vez por semana	Um vez por semana	Quase todos os dias	Diariamente
Computador					
Smartphone					
Tablet					
Smartwatch					

12. Qual o seu telemóvel atual? ("Nenhum" ou o modelo do telemóvel) *

13. Quando insere texto no seu telemóvel, de que maneira o faz? (e.g., voz, qwerty) *

14. Alguma vez sentiu interesse ou necessidade em utilizar um smartwatch? *

Marcar tudo o que for aplicável.

- [ ] Interesse
- [ ] Necessidade
- [ ] Não

15. Notas:

Avançar para a pergunta 16

Estudo

< REALIZAR ESTUDO >

16. ** Duração do estudo *

Exemplo: 08:30

Agradecimentos	Sendo assim, damos por terminada esta sessão do estudo. Espero que tenha sido uma experiência no mínimo agradável. Mais uma vez muito obrigado pelo tempo dispendido, a sua participação foi fundamental. Da nossa parte é tudo, Obrigado!

Appendix B

First study's questionnaire applied to blind participants

BrailleShapes Questionnaire

*Obrigatório

Dados Pessoais

Visto que eu já me apresentei, vou agora fazer-lhe umas perguntas em forma de questionário de forma a ficar a saber mais sobre si.

NOTA: As questões assinaladas com ** não são para perguntar ao participante

1. ** ID Utilizador *

2. Nome *

3. Idade *

Marcar apenas uma oval.

- < 18
- 18 - 23
- 24 - 35
- 36 - 45
- 46 - 55
- 56 - 65
- > 65

4. Sexo *

Marcar apenas uma oval.

- Feminino
- Masculino
- Outro / Prefiro não revelar

5. Habilitações *

Marcar apenas uma oval.

- Ensino Básico 2º Ciclo (5º e 6º anos)
- Ensino Básico 3º Ciclo (7º, 8º e 9º anos)
- Ensino Secundário (10º, 11º e 12º anos)
- Ensino Superior
- Outro / Prefiro não revelar

6. Mão Dominante *

Marcar apenas uma oval.

- Direita
- Esquerda
- Ambidestro

7. Usa Relógio *

Marcar apenas uma oval.

- Sim
- Não

8. Mão em que usa/faria mais sentido usar Relógio *

Marcar apenas uma oval.

- Direita
- Esquerda
- Ambas (Justificar)

9. Grau de Cegueira ("Total", "Parcial" ou % correspondente) *

10. Idade de Aquisição ("Nascença" ou Valor Numérico) *

11. Possui alguma condição que lhe diminua a sensibilidade nas mãos? *

12. NOTA: Só no caso de cegueira "tardia". Teve contacto visual com smartphones, smartwatches ou outros dispositivos com superficies táteis antes de adquirir a deficiência visual ("Não" ou qual o dispositivo) *

13. Notas:

Avançar para a pergunta 14

Braille

Relativamente ao sistema de escrita tátil Braille:

14. Qual o seu nível de escrita numa máquina de escrever Braille / Perkins Brailler? *

Marcar apenas uma oval.

Inexistente

- 0
- 1
- 2
- 3
- 4
- 5

Fluente

15. Nível de Leitura *

Marcar apenas uma oval.

Inexistente

0

1

2

3

4

5

Fluente

16. Notas:

Avançar para a pergunta 17

Tecnologia

Entrando agora numa secção mais tecnológica,

17. Qual o seu à vontade com dispositivos com ecrãs tácteis? *

Marcar apenas uma oval.

Muito pouco à vontade

1

2

3

4

5

Muito à vontade

18. Com que frequência utiliza os seguintes dispositivos? *

Marcar apenas uma oval por linha.

	Nunca	Menos de uma vez por semana	Um vez por semana	Quase todos os dias	Diariamente
Computador					
Smartphone					
Tablet					
Smartwatch					

19. Qual o seu telemóvel atual? ("Nenhum" ou o modelo do telemóvel) *

20. Quando insere texto no seu telemóvel, de que maneira o faz? (e.g., voz, qwerty + VoiceOver(Padrão, Datilografia profissional, Toque Direto)) *

21. Alguma vez sentiu interesse ou necessidade em utilizar um smartwatch? *

Marcar tudo o que for aplicável.

- [ ] Interesse
- [ ] Necessidade
- [ ] Não

22. Notas:

Avançar para a pergunta 23

Estudo

< REALIZAR ESTUDO >

23. ** Duração do estudo *

Text Entry Controls

24. Das seguintes opções, qual lhe faria mais sentido usar para inserir um espaço quando termina de escrever uma palavra? *

Marcar apenas uma oval.

- Toque duplo (1 só dedo, sequência rápida de 2 toques)
- Toque com 2 dedos em simultâneo
- Desenhar um circulo
- Outra: _____

25. Sugestão:

_____

26. Das seguintes opções, qual lhe faria mais sentido usar para apagar a última letra que inseriu? *

Marcar apenas uma oval.

- Toque duplo (1 só dedo, sequência rápida de 2 toques)
- Toque com 2 dedos em simultâneo
- Toque longo (1 só dedo, tocar no ecrã e ficar a pressionar)
- Desenhar um circulo
- Outra: _____

27. Sugestão:

_____

Agradecimentos

Sendo assim, damos por terminada esta sessão do estudo.
Espero que tenha sido uma experiência no mínimo agradável.
Mais uma vez muito obrigado pelo tempo dispendido, a sua participação foi fundamental.
Da nossa parte é tudo,
Obrigado!

Appendix C

First study's script applied to sighted participants

This script is performed alongside a questionnaire/form
Questionnaire: https://forms.gle/akih1KJMnVJqdEVw7
This script is performed alongside a Braille alphabet sheet

Take notes of what feels relevant

< Iniciar ambas as aplicações >
< Iniciar cronômetro >

Bom/Boa dia/tarde/noite,

Antes de mais muito obrigado por ter aceite participar neste estudo, esperamos que corra tudo da melhor forma e que seja uma experiência o mais agradável possível para si.
O meu nome é Pedro Marques, sou estudante de mestrado em Engenharia Informática na Faculdade de Ciências da Universidade de Lisboa, e estou aqui hoje acompanhado pelo Professor Tiago Guerreiro que é o orientador responsável por este projeto.

Este estudo faz parte da minha dissertação que tem como objetivo principal desenvolver e avaliar um método de escrita para smartwatches para pessoas cegas, através do desenho de formas que representam caracteres Braille.
Nesta etapa do projeto, o nosso objetivo é recolher o máximo de dados e informação possível e por este motivo, este estudo em específico vai focar-se na recolha de gestos/formas desenhadas representativas de caracteres Braille, e vai ainda ter em conta os comentários ou sugestões que possa ter.
Quero deixar claro que ao longo deste estudo, o nosso foco vai estar sempre na recolha de informação e na avaliação da nossa abordagem, sendo que nunca o/a iremos avaliar a si. Para além disso, toda a informação recolhida será tratada única e exclusivamente no âmbito do projeto e para fins académicos. Deste modo, quero pedir-lhe que partilhe todas as observações e sugestões que possa ter, visto que são aspetos fundamentais para a melhor compreensão do problema e por consequência da solução. Quero ainda pedir-lhe que exponha qualquer dúvida que tenha a qualquer momento.

Dito isto, tem alguma questão até agora?

Então antes de começarmos vou só pedir-lhe que leia e assine este formulário de consentimento relativamente à participação neste estudo. < Pedir para assinar consentimento >

< Realizar Questionário Até Secção de Tecnologia >

Como dito anteriormente, o objetivo é desenvolver um método de escrita para smartwatches, o que quer dizer que ao longo deste estudo vamos utilizar um. Caso não tenha conhecimento, um smartwatch é um relógio com ecrã tátil que permite realizar algumas das mesmas tarefas que um smartphone < dar smartwatch e pedir para colocar >. Este smartwatch em particular tem, para além de uma face tátil, dois botões do seu lado direito. O seu ecrã tátil apresenta, apesar de não ser palpável, um rebordo que o limita antes do fim do relógio em si. Este rebordo é pequeno o que acaba por não ter grande efeito no uso do smartwatch, no entanto vou pedir-lhe que tente não começar ou acabar os gestos mesmo no limite do ecrã do relógio < tentar indicar quais os limites do smartwatch e se necessário a localização dos botões >. Pode colocar os braços na posição que lhe parecer mais confortável para usar o relógio. Antes de prosseguirmos, vou fazer-lhe mais algumas perguntas, neste caso, relativas a dispositivos tecnológicos < voltar para questionário >.

Antes de começarmos a sessão de recolha de gestos, vamos perceber um pouco melhor o que quer dizer escrever em Braille usando gestos ou desenhando formas. Braille consiste num sistema de escrita tátil, em que cada caractere é representado por um conjunto de pontos levantados numa célula, e cada célula Braille é composta por duas colunas de 3 pontos cada, como lhe será mostrado mais a frente. Se pensarmos em unir os pontos utilizados para representar algum caractere, com um traço apenas, passando uma única vez por cada ponto, é possível imaginarmos uma forma aberta. É essa a forma que pretendemos que utilize para inserir os caracteres que lhe pedirmos.

Até aqui, alguma dúvida?

Iremos começar com uma pequena sessão de treino de cerca de 5 minutos para que se habitue ao relógio e a esta nossa abordagem, e de seguida iremos prosseguir para o estudo propriamente dito. Em ambas as sessões vou-lhe pedir para desenhar padrões Braille equivalentes a determinados caracteres. Para o fazer, pode fazer um gesto com o dedo no ecrã, da forma que lhe parecer mais ágil e intuitiva possível, começando no sítio que quiser e com a orientação que quiser, desde que comece e termine dentro dos limites do ecrã.
Vai começar por ouvir uma letra dita pelo relógio e de seguida vai-me dizer que letra ouviu. Após o fazer, eu irei ativar o ecrã e poderá então desenhar a forma correspondente à letra. Quando completar a forma total, vou-lhe pedir que descanse a mão na mesa de forma a saber que terminou. Caso se tenha enganado a tocar no ecrã e queira refazer a letra basta dizer.
Este processo vai ser repetido até passarmos pelas letras pelo menos uma vez.
A qualquer momento esteja à vontade para perguntar caso tenha alguma dúvida.

Vamos então começar a nossa sessão de treino.
< Treino: A B D G K M O R U X Z (repetir se necessário ou usar outras letras) >
< Aprox. 5 min >

Vamos agora dar então início à recolha em si. Tal como na sessão de treino ser-lhe-ão lidas letras através do smartwatch, as quais lhe vou pedir que repita para depois desenhar. O resto do processo é o mesmo, no entanto desta vez iremos desenhar todas as letras do alfabeto inglês. Se a algum momento precisar de fazer uma pausa, não hesite em pedir.
Alguma questão?

< Estudo: Todas as letras - 2 Iterações >

< FEEDBACK SESSION >

Appendix D

First study's script applied to blind participants

This script is performed alongside a questionnaire/form
Questionnaire: https://forms.gle/orSaxBKkVqBiBtWD8

Take notes of what feels relevant

< Iniciar ambas as aplicações >
< Iniciar cronômetro >

Bom/Boa dia/tarde/noite,

Muito obrigado por ter aceite participar neste estudo, esperamos que corra tudo da melhor forma e que seja uma experiência o mais agradável possível para si.
O meu nome é Pedro Marques, sou estudante de mestrado em Engenharia Informática na Faculdade de Ciências da Universidade de Lisboa, e estou aqui hoje acompanhado pelo Professor Tiago Guerreiro que é o orientador responsável por este projeto.

Antes de mais deixe-me perguntar-lhe se tem preferência que eu utilize algum dos termos "pessoa cega", "pessoa invisual" ou outro que lhe faça sentido?

Este estudo faz parte da minha dissertação que tem como objetivo principal desenvolver e avaliar um método de escrita para smartwatches para pessoas cegas, através do desenho de formas que representam caracteres Braille.
Nesta etapa do projeto, o nosso objetivo é recolher o máximo de dados e informação possível e por este motivo, este estudo em específico vai focar-se na recolha de gestos/formas desenhadas representativas de caracteres Braille, e vai ainda ter em conta os comentários ou sugestões que possa ter.
Quero deixar claro que ao longo deste estudo, o nosso foco vai estar sempre na recolha de informação e na avaliação da nossa abordagem, sendo que nunca o/a iremos avaliar a si. Para além disso, toda a informação recolhida será tratada única e exclusivamente no âmbito do projeto e para fins académicos. Deste modo, quero pedir-lhe que partilhe todas as observações e sugestões que possa ter, visto que são aspetos fundamentais para a melhor compreensão do problema e por consequência da solução. Quero ainda pedir-lhe que exponha qualquer dúvida que tenha a qualquer momento.

Dito isto, tem alguma questão até agora?

Então antes de começarmos vou só pedir-lhe que leia e assine este formulário de consentimento relativamente à participação neste estudo. < Pedir para assinar consentimento >

< Realizar Questionário Até Secção de Tecnologia >

Como dito anteriormente, o objetivo é desenvolver um método de escrita para smartwatches, o que quer dizer que ao longo deste estudo vamos utilizar um. Caso não tenha conhecimento, um smartwatch é um relógio com ecrã tátil que permite realizar algumas das mesmas tarefas que um smartphone < dar smartwatch e pedir para colocar o mais justo possível de forma a que fique confortável no pulso >. Este smartwatch em particular tem, para além de uma face tátil, dois botões do seu lado direito. O seu ecrã tátil apresenta, apesar de não ser palpável, um rebordo que o limita antes do fim do relógio em si. Este rebordo é pequeno o que acaba por não ter grande efeito no uso do smartwatch, no entanto vou pedir-lhe que tente não começar ou acabar os gestos mesmo no limite do ecrã do relógio <se necessário mostrar demo app> < tentar indicar quais os limites do smartwatch e se necessário a localização dos botões >. Pode colocar os braços na posição que lhe parecer mais confortável para usar o relógio. Antes de

prosseguirmos, vou fazer-lhe mais algumas perguntas, neste caso, relativas a dispositivos tecnológicos < voltar para questionário >.

Antes de começarmos a sessão de recolha de gestos, vamos perceber um pouco melhor o que quer dizer escrever em Braille usando gestos ou desenhando formas. Braille consiste num sistema de escrita tátil, em que cada caractere é representado por um conjunto de pontos levantados numa célula, e cada célula Braille é composta por duas colunas de 3 pontos cada. Se pensarmos em unir os pontos utilizados para representar algum caractere, com um traço apenas, passando uma única vez por cada ponto, é possível imaginarmos uma forma aberta. É essa a forma que pretendemos que utilize para inserir os caracteres que lhe pedirmos.

Até aqui, alguma dúvida?

Iremos começar com uma pequena sessão de treino de cerca de 5 minutos para que se habitue ao relógio e a esta nossa abordagem, e de seguida iremos prosseguir para o estudo propriamente dito. Em ambas as sessões vou-lhe pedir para desenhar padrões Braille equivalentes a determinados caracteres. Para o fazer, pode fazer um gesto com o dedo no ecrã, da forma que lhe parecer mais ágil e intuitiva possível, começando no sítio que quiser e com a orientação que quiser, desde que comece e termine a forma dentro dos limites do ecrã.
Vai começar por ouvir uma letra dita pelo relógio e de seguida vai-me dizer que letra ouviu. Após o fazer, eu irei ativar o ecrã e poderá então desenhar a forma correspondente à letra. Quando completar a forma total, vou-lhe pedir que descanse a mão na mesa de forma a saber que terminou. Caso se tenha enganado a tocar no ecrã e queira refazer a letra basta dizer.
Este processo vai ser repetido até passarmos pelas letras pelo menos uma vez.
A qualquer momento esteja à vontade para perguntar caso tenha alguma dúvida.

Vamos então começar a nossa sessão de treino.
< Treino: A B D G K M O R U X Z (repetir se necessário ou usar outras letras) >
< Aprox. 5 min >

Vamos agora dar então início à recolha em si. Tal como na sessão de treino ser-lhe-ão lidas letras através do smartwatch, as quais lhe vou pedir que repita para depois desenhar. O resto do processo é o mesmo, no entanto desta vez iremos desenhar todas as letras do alfabeto inglês e uma forma que será um círculo. Se a algum momento precisar de fazer uma pausa, não hesite em pedir.
Alguma questão?

< Estudo: Todas as letras - 2 Iterações >

Para terminar o estudo, vou-lhe só fazer algumas perguntas de forma a obter a sua opinião em relação a alguns aspetos, e de seguida terá a oportunidade de, se quiser, deixar os seus comentários em relação a todo este processo.

< voltar para questionário >

< FEEDBACK SESSION >

Appendix E

Second study's first phase questionnaire

BrailleShapes Questionnaire

*Obrigatório

Dados Pessoais

Visto que eu já me apresentei, vou agora fazer-lhe umas perguntas em forma de questionário de forma a ficar a saber mais sobre si.

NOTA: As questões assinaladas com ** não são para perguntar ao participante

1. ** ID Utilizador *

2. Nome *

3. Idade *

Marcar apenas uma oval.

- < 18
- 18 - 23
- 24 - 35
- 36 - 45
- 46 - 55
- 56 - 65
- > 65

4. Sexo *

Marcar apenas uma oval.

- Feminino
- Masculino
- Outro / Prefiro não revelar

5. Habilitações *

Marcar apenas uma oval.

- Ensino Básico 2º Ciclo (5º e 6º anos)
- Ensino Básico 3º Ciclo (7º, 8º e 9º anos)
- Ensino Secundário (10º, 11º e 12º anos)
- Ensino Superior
- Outro / Prefiro não revelar

6. Mão Dominante *

Marcar apenas uma oval.

- Direita
- Esquerda
- Ambidestro

7. Usa Relógio *

Marcar apenas uma oval.

- Sim
- Não

8. Mão em que usa/faria mais sentido usar Relógio *

Marcar apenas uma oval.

- Direita
- Esquerda
- Ambas (Justificar)

9. Grau de Cegueira ("Total", "Parcial" ou % correspondente) *

10. Idade de Aquisição ("Nascença" ou Valor Numérico) *

11. Possui alguma condição que lhe diminua a sensibilidade nas mãos? *

12. NOTA: Só no caso de cegueira "tardia". Teve contacto visual com smartphones, smartwatches ou outros dispositivos com superficies táteis antes de adquirir a deficiência visual ("Não" ou qual o dispositivo) *

13. Notas:

Avançar para a pergunta 14

Braille

Relativamente ao sistema de escrita tátil Braille:

14. Qual o seu nível de escrita numa máquina de escrever Braille / Perkins Brailler? *

Marcar apenas uma oval.

Inexistente

0

1

2

3

4

5

Fluente

15. Nível de Leitura *

Marcar apenas uma oval.

Inexistente

0

1

2

3

4

5

Fluente

16. Notas:

Avançar para a pergunta 17

Tecnologia

Entrando agora numa secção mais tecnológica,

17. Qual o seu à vontade com dispositivos com ecrãs tácteis? *

Marcar apenas uma oval.

Muito pouco à vontade

1

2

3

4

5

Muito à vontade

18. Com que frequência utiliza os seguintes dispositivos? *

Marcar apenas uma oval por linha.

	Nunca	Menos de uma vez por semana	Um vez por semana	Quase todos os dias	Diariamente
Computador					
Smartphone					
Tablet					
Smartwatch					

19. Qual o seu telemóvel atual? ("Nenhum" ou o modelo do telemóvel) *

20. Quando insere texto no seu telemóvel, de que maneira o faz? (e.g., voz, qwerty + VoiceOver(Padrão, Datilografia profissional, Toque Direto)) *

21. Alguma vez sentiu interesse ou necessidade em utilizar um smartwatch? *

Marcar tudo o que for aplicável.

- [ ] Interesse
- [ ] Necessidade
- [ ] Não

22. Notas:

Avançar para a pergunta 23

Estudo

< REALIZAR ESTUDO >

23. ** Duração do estudo *

Text Entry Controls

24. Das seguintes opções, qual lhe faria mais sentido usar para inserir um espaço quando termina de escrever uma palavra? *

Marcar apenas uma oval.

- Toque duplo (1 só dedo, sequência rápida de 2 toques)
- Toque com 2 dedos em simultâneo
- Desenhar um circulo
- Outra: _____

25. Sugestão:

_____

26. Das seguintes opções, qual lhe faria mais sentido usar para apagar a última letra que inseriu? *

Marcar apenas uma oval.

- Toque duplo (1 só dedo, sequência rápida de 2 toques)
- Toque com 2 dedos em simultâneo
- Toque longo (1 só dedo, tocar no ecrã e ficar a pressionar)
- Desenhar um circulo
- Outra: _____

27. Sugestão:

_____

Agradecimentos

Sendo assim, damos por terminada esta sessão do estudo.
Espero que tenha sido uma experiência no mínimo agradável.
Mais uma vez muito obrigado pelo tempo dispendido, a sua participação foi fundamental.
Da nossa parte é tudo,
Obrigado!

Appendix F

Second study's second phase questionnaire

BrailleShapes Questionnaire

*Obrigatório

Dados Pessoais

NOTA: As questões assinaladas com ** não são para perguntar ao participante

1. ** ID Utilizador *

2. Nome *

3. Notas:

Estudo

< REALIZAR ESTUDO >

4. ** Duração do estudo *

Subjective Feedback

Vou agora expor-lhe algumas afirmações às quais gostava que me dissesse numa escala de 1 a 5, se concorda ou não (1 = Discordo Fortemente, 5 = Concordo Fortemente).

5. Eu acho que eu gostaria de usar um sistema baseado neste método com frequência. *

Marcar apenas uma oval.

Discordo Fortemente

1

2

3

4

5

Concordo Fortemente

6. Achei o método desnecessariamente complexo. *

Marcar apenas uma oval.

Discordo Fortemente

1

2

3

4

5

Concordo Fortemente

7. Achei o método fácil de usar (para realizar as tarefas propostas). *

Marcar apenas uma oval.

Discordo Fortemente

1

2

3

4

5

Concordo Fortemente

8. Após o tempo de treino (depois de inserir algumas letras/frases), foi fácil e rápido inserir as letras/frases. *

Marcar apenas uma oval.

Discordo Fortemente

1

2

3

4

5

Concordo Fortemente

9. Eu imagino que a maioria das pessoas aprenderia a usar um sistema com este método muito rapidamente. *

Marcar apenas uma oval.

Discordo Fortemente

1

2

3

4

5

Concordo Fortemente

10. Eu senti-me muito confiante a usar o método. *

Marcar apenas uma oval.

Discordo Fortemente

1

2

3

4

5

Concordo Fortemente

11. A maneira como se insere texto é intuitiva. *

Marcar apenas uma oval.

Discordo Fortemente

1

2

3

4

5

Concordo Fortemente

12. Eu usaria um sistema com este método se tivesse um smartwatch. *

Marcar apenas uma oval.

	Discordo Fortemente				Concordo Fortemente
	1	2	3	4	5

13. Sugestão:

Agradecimentos	Sendo assim, damos por terminada esta sessão do estudo. Espero que tenha sido uma experiência no mínimo agradável. Mais uma vez muito obrigado pelo tempo dispendido, a sua participação foi fundamental. Da nossa parte é tudo, Obrigado!

Appendix G

Second study's script

This script is performed alongside 2 questionnaires/forms
Questionnaire 1: https://forms.gle/umNJeuyd9J35njfF8
Questionnaire 2: https://forms.gle/6g1WUdzMxrj4HPvN6

Take notes of what feels relevant

< Iniciar ambas as aplicações >
< Iniciar cronômetro >

Bom/Boa dia/tarde/noite,

Muito obrigado por ter aceite participar neste estudo, esperamos que corra tudo da melhor forma e que seja uma experiência o mais agradável possível para si.
O meu nome é Pedro Marques, sou estudante de mestrado em Engenharia Informática na Faculdade de Ciências da Universidade de Lisboa, e estou aqui hoje acompanhado pelo Professor Tiago Guerreiro que é o orientador responsável por este projeto.

Antes de mais deixe-me perguntar-lhe se tem preferência que eu utilize algum dos termos "pessoa cega", "pessoa invisual" ou outro que lhe faça sentido?

Este estudo faz parte da minha dissertação que tem como objetivo principal desenvolver e avaliar um método de escrita para smartwatches para pessoas cegas, através do desenho de formas que representam caracteres Braille.
Esta etapa do projeto é uma continuação de uma etapa anterior, onde o nosso objetivo foi recolher o máximo de dados e informação possível focando-se na recolha de gestos/formas desenhadas representativas de caracteres Braille, bem como comentários ou sugestões dos participantes.
À semelhança da etapa anterior, nesta etapa vamos também fazer uma recolha de gestos que representam caracteres individuais, mas vamos ainda recolher frases escritas pelos participantes sendo então este estudo dividido em 2 partes.
Quero deixar claro que ao longo deste estudo, o nosso foco vai estar sempre na recolha de informação e na avaliação da nossa abordagem, sendo que nunca o/a iremos avaliar a si. Para além disso, toda a informação recolhida será tratada única e exclusivamente no âmbito do projeto e para fins académicos. Deste modo, quero pedir-lhe que partilhe todas as observações e sugestões que possa ter, visto que são aspetos fundamentais para a melhor compreensão do problema e por consequência da solução. Quero ainda pedir-lhe que exponha qualquer dúvida que tenha a qualquer momento.

Dito isto, tem alguma questão até agora?

Então antes de começarmos vou só pedir-lhe que leia e assine este formulário de consentimento relativamente à participação neste estudo. < Pedir para assinar consentimento >

ETAPA 1

NO CASO DE PARTICIPANTES QUE NÃO PARTICIPARAM NO ESTUDO ANTERIOR
< Realizar Questionário Até Secção de Tecnologia >

Como dito anteriormente, o objetivo é desenvolver um método de escrita para smartwatches, o que quer dizer que ao longo deste estudo vamos utilizar um. Caso não tenha conhecimento, um smartwatch é um relógio com ecrã tátil que permite realizar algumas das mesmas tarefas que

um smartphone < dar smartwatch e pedir para colocar o mais justo possível de forma a que fique confortável no pulso >. Este smartwatch em particular tem, para além de uma face tátil, dois botões do seu lado direito. O seu ecrã tátil apresenta, apesar de não ser palpável, um rebordo que o limita antes do fim do relógio em si. Este rebordo é pequeno o que acaba por não ter grande efeito no uso do smartwatch, no entanto vou pedir-lhe que tente não começar ou acabar os gestos mesmo no limite do ecrã do relógio <se necessário mostrar demo app> < tentar indicar quais os limites do smartwatch e se necessário a localização dos botões >. Pode colocar os braços na posição que lhe parecer mais confortável para usar o relógio. Antes de prosseguirmos, vou fazer-lhe mais algumas perguntas, neste caso, relativas a dispositivos tecnológicos < voltar para questionário >.

Antes de começarmos a sessão de recolha de gestos, vamos perceber um pouco melhor o que quer dizer escrever em Braille usando gestos ou desenhando formas. Braille consiste num sistema de escrita tátil, em que cada caractere é representado por um conjunto de pontos levantados numa célula, e cada célula Braille é composta por duas colunas de 3 pontos cada. Se pensarmos em unir os pontos utilizados para representar algum caractere, com um traço apenas, passando uma única vez por cada ponto, é possível imaginarmos uma forma aberta. É essa a forma que pretendemos que utilize para inserir os caracteres que lhe pedirmos.

Até aqui, alguma dúvida?

Iremos começar com uma pequena sessão de treino de cerca de 5 minutos para que se habitue ao relógio e a esta nossa abordagem, e de seguida iremos prosseguir para o estudo propriamente dito. Em ambas as sessões vou-lhe pedir para desenhar padrões Braille equivalentes a determinados caracteres. Para o fazer, pode fazer um gesto com o dedo no ecrã, da forma que lhe parecer mais ágil e intuitiva possível, começando no sítio que quiser e com a orientação que quiser, desde que comece e termine a forma dentro dos limites do ecrã.
Vai começar por ouvir uma letra dita pelo relógio e de seguida vai-me dizer que letra ouviu. Após o fazer, eu irei ativar o ecrã e poderá então desenhar a forma correspondente à letra. Quando completar a forma total, vou-lhe pedir que descanse a mão na mesa de forma a saber que terminou. Caso se tenha enganado a tocar no ecrã e queira refazer a letra basta dizer.
Este processo vai ser repetido até passarmos pelas letras pelo menos uma vez.
A qualquer momento esteja à vontade para perguntar caso tenha alguma dúvida.

Vamos então começar a nossa sessão de treino.
< Treino: A B D G K M O R U X Z (repetir se necessário ou usar outras letras) >
< Aprox. 5 min >

Vamos agora dar então início à recolha em si. Tal como na sessão de treino ser-lhe-ão lidas letras através do smartwatch, as quais lhe vou pedir que repita para depois desenhar. O resto do processo é o mesmo, no entanto desta vez iremos desenhar todas as letras do alfabeto inglês e uma forma que será um círculo. Se a algum momento precisar de fazer uma pausa, não hesite em pedir.
Alguma questão?

< Estudo: Todas as letras - 2 Iterações >

Para terminar esta primeira etapa, vou-lhe só fazer algumas perguntas de forma a obter a sua opinião em relação a alguns aspetos, e de seguida terá a oportunidade de, se quiser, deixar os seus comentários em relação a todo este processo.

< voltar para questionário >

ETAPA 2

Nesta segunda etapa, o objetivo é, utilizando os mesmo gestos/formas que utilizou na primeira etapa, que escreva um conjunto de frases aleatórias que lhe vão ser ditas. Tal como na primeira etapa iremos começar com uma pequena sessão de treino para que se habitue a esta nova metodologia, e de seguida iremos prosseguir para o estudo propriamente dito. Em ambas as sessões vou-lhe pedir para desenhar padrões Braille como anteriormente, mas de forma a que componha frases.
Vai começar por ouvir uma frase dita pelo relógio e de seguida vai-me dizer que frase ouviu. Após o fazer, eu irei ativar o ecrã e poderá então escrever a frase. Vai fazê-lo desenhando uma forma de cada vez no ecrã. Para dar um espaço entre palavras, poderá dar dois toques rápidos no ecrã, e para apagar alguma letra, basta ficar a clicar no ecrã até ouvir um som respetivo. Quando acabar a frase, pode dar novamente dois toques de forma a confirmar.
Este processo vai ser repetido até passarmos pelas frases todas.
Se a algum momento precisar de fazer uma pausa, não hesite em pedir.
Alguma questão?

Vamos então começar a nossa sessão de treino.
< Treino: 2 frases aleatórias (repetir se necessário) >
< Aprox. 10 min >

Vamos agora dar então início à recolha em si. O procedimento é o mesmo que o da sessão de treino. Se a algum momento precisar de fazer uma pausa, não hesite em pedir.
< Estudo: 3 frases aleatórias >

< Realizar Questionário 2ª Etapa>

< FEEDBACK SESSION >

Bibliography

[1] Lisa Anthony, Radu-Daniel Vatavu, and Jacob O. Wobbrock. Understanding the consistency of users' pen and finger stroke gesture articulation. In *Proceedings of Graphics Interface 2013*, GI '13, page 87–94, CAN, 2013. Canadian Information Processing Society.

[2] Shabnam Mohamed Aslam and Shirina Samreen. Gesture recognition algorithm for visually blind touch interaction optimization using crow search method. *IEEE Access*, 8:127560–127568, 2020.

[3] Matthew N. Bonner, Jeremy T. Brudvik, Gregory D. Abowd, and W. Keith Edwards. No-look notes: Accessible eyes-free multi-touch text entry. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6030 LNCS:409–426, 2010.

[4] Mauricio Cirelli and Ricardo Nakamura. A survey on multi-touch gesture recognition and multi-touch frameworks. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, ITS '14, page 35–44, New York, NY, USA, 2014. Association for Computing Machinery.

[5] J.-P. Crettez. A set of handwriting families: style recognition. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 489–494 vol.1, 1995.

[6] Nem Khan Dim and Xiangshi Ren. Designing Motion Gesture Interfaces in Mobile Phones for Blind People. *Journal of Computer Science and Technology*, 29(5):812–824, sep 2014.

[7] Tim Dittmar, Claudia Krull, and Graham Horton. A new approach for touch gesture recognition: Conversive hidden non-markovian models. *Journal of Computational Science*, 10:66–76, 2015.

[8] Shirin Feiz and I. V. Ramakrishnan. Exploring feasibility of wrist gestures for non-visual interactions with wearables. In *Proceedings of the 16th International Web for All Conference*, W4A '19, New York, NY, USA, 2019. Association for Computing Machinery.

[9] Brian Frey, Caleb Southern, and Mario Romero. Brailletouch: Mobile texting for the visually impaired. In Constantine Stephanidis, editor, *Universal Access in Human-Computer Interaction. Context Diversity*, pages 19–25, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[10] Fulano, Cicrano, and Beltrano. A paper on something. In *The 7th Conference on Things and Stuff (CTS 2009)*, Lisbon, Portugal, May 2009. Accepted for publication.

[11] Tiago Guerreiro, Paulo Lagoa, Hugo Nicolau, Daniel Gonalves, and Joaquim A. Jorge. From tapping to touching: Making touch screens accessible to blind users. *IEEE MultiMedia*, 15(4):48–50, 2008.

[12] Tiago Guerreiro, Hugo Nicolau, Joaquim Jorge, and Daniel Gonçalves. Navtap: A long term study with excluded blind users. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '09, page 99–106, New York, NY, USA, 2009. Association for Computing Machinery.

[13] Rongrong Huang, Biao Liu, Wei Su, and He Lin. Research on braille music recognition based on convolutional neural network. In *2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pages 837–843, 2018.

[14] Shaun K. Kane, Jacob O. Wobbrock, and Richard E. Ladner. Usable gestures for blind people: Understanding preference and performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 413–422, New York, NY, USA, 2011. Association for Computing Machinery.

[15] P. O. Kristensson and L. C. Denby. Continuous recognition and visualization of pen strokes and touch-screen gestures. In *Proceedings of the Eighth Eurographics Symposium on Sketch-Based Interfaces and Modeling*, SBIM '11, page 95–102, New York, NY, USA, 2011. Association for Computing Machinery.

[16] Mingzhe Li, Mingming Fan, and Khai N. Truong. Braillesketch: A gesture-based text input method for people with visual impairments. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '17, page 12–21, New York, NY, USA, 2017. Association for Computing Machinery.

[17] Yuan Li and Yangyang Li. Design and implementation of handwritten chinese character recognition method based on cnn and tensorflow. In *2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, pages 878–882, 2021.

[18] Mateus M. Luna, Fabrízzio A. A. de M. N. Soares, Hugo A. D. Nascimento, and Aaron Quigley. Braille text entry on smartwatches: An evaluation of methods for composing the braille cell. In *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, EICS '19, New York, NY, USA, 2019. Association for Computing Machinery.

[19] Mateus Machado Luna, Fabrizzio Alphonsus Alves de Melo Nunes Soares, Hugo Alexandre Dantas do Nascimento, Joyce Siqueira, Eduardo Faria de Souza, Thamer Horbylon Nascimento, and Ronaldo Martins da Costa. Text entry on smartwatches: A systematic review of literature. In *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, volume 02, pages 272–277, 2018.

[20] Mateus Machado Luna, Hugo Nascimento, Aaron Quigley, and Fabrizzio Soares. Text entry for the blind on smartwatches: A study of braille code input methods for a novel device. *Universal Access in the Information Society*, 02 2022.

[21] I. Scott MacKenzie and Kumiko Tanaka-Ishii. *Text Entry Systems: Mobility, Accessibility, Universality*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2007.

[22] I. Scott MacKenzie and Shawn X. Zhang. The immediate usability of graffiti. In *Proceedings of the Conference on Graphics Interface '97*, page 129–137, CAN, 1997. Canadian Information Processing Society.

[23] Thamer Hobylon Nascimento, Fabrizzio Alphonsus A. M. N. Soares, Pouang Polad Irani, Leandro Luiz Galdino de Oliveira, and Anderson Da Silva Soares. Method for text entry in smartwatches using continuous gesture recognition. In *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, volume 2, pages 549–554, 2017.

[24] Hugo Nicolau, João Guerreiro, Tiago Guerreiro, and Luís Carriço. Ubibraille: Designing and evaluating a vibrotactile braille-reading device. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '13, New York, NY, USA, 2013. Association for Computing Machinery.

[25] Hugo Nicolau, Kyle Montague, Tiago Guerreiro, João Guerreiro, and Vicki L. Hanson. B#: Chord-based correction for multitouch braille input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, page 1705–1708, New York, NY, USA, 2014. Association for Computing Machinery.

[26] Hugo Nicolau, Kyle Montague, Tiago Guerreiro, André Rodrigues, and Vicki L. Hanson. Holibraille: Multipoint vibrotactile feedback on mobile devices. In *Proceedings of the 12th International Web for All Conference*, W4A '15, New York, NY, USA, 2015. Association for Computing Machinery.

[27] João Oliveira, Tiago Guerreiro, Hugo Nicolau, Joaquim Jorge, and Daniel Gonçalves. 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*, ASSETS '11, page 179–186, New York, NY, USA, 2011. Association for Computing Machinery.

[28] João Oliveira, Tiago Guerreiro, Hugo Nicolau, Joaquim Jorge, and Daniel Gonçalves. Brailletype: Unleashing braille over touch screen mobile phones. In Pedro Campos, Nicholas Graham, Joaquim Jorge, Nuno Nunes, Philippe Palanque, and Marco Winckler, editors, *Human-Computer Interaction – INTERACT 2011*, pages 100–107, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[29] André Rodrigues, Hugo Nicolau, Kyle Montague, Luís Carriço, and Tiago Guerreiro. Effect of target size on non-visual text-entry. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '16, page 47–52, New York, NY, USA, 2016. Association for Computing Machinery.

[30] Fathma Siddique, Shadman Sakib, and Md. Abu Bakr Siddique. Recognition of handwritten digit using convolutional neural network in python with tensorflow and comparison of performance for various hidden layers. In *2019 5th International Conference on Advances in Electrical Engineering (ICAEE)*, pages 541–546, 2019.

[31] Joyce Siqueira, Fabrizzio Alphonsus Alves de Melo Soares, Deller James Ferreira, Cleyton Rafael Gomes Silva, Luciana de Oliveira Berretta, Cristiane Bastos Rocha Ferreira, Igor Moreira Félix, Anderson da Silva Soares, Ronaldo Martins da Costa, and Mateus Machado Luna. Braille text entry on smartphones: A systematic review of the literature. In *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*, volume 2, pages 521–526, 2016.

[32] R. William Soukoreff and I. Scott MacKenzie. Measuring errors in text entry tasks: An application of the levenshtein string distance statistic. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '01, page 319–320, New York, NY, USA, 2001. Association for Computing Machinery.

[33] Caleb Southern, James Clawson, Brian Frey, Gregory Abowd, and Mario Romero. An evaluation of brailletouch: Mobile touchscreen text entry for the visually impaired. In *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '12, page 317–326, New York, NY, USA, 2012. Association for Computing Machinery.

[34] S.N. Srihari, Sung-Hyuk Cha, H. Arora, and Sangjik Lee. Individuality of handwriting: a validation study. In *Proceedings of Sixth International Conference on Document Analysis and Recognition*, pages 106–109, 2001.

[35] Hussain Tinwala and I. Scott MacKenzie. Eyes-free text entry on a touchscreen phone. In *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)*, pages 83–88, 2009.

[36] Daniel Trindade, André Rodrigues, Tiago Guerreiro, and Hugo Nicolau. *Hybrid-Brailler: Combining Physical and Gestural Interaction for Mobile Braille Input and Editing*, page 1–12. Association for Computing Machinery, New York, NY, USA, 2018.

[37] Radu-Daniel Vatavu. Improving gesture recognition accuracy on touch screens for users with low vision. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 4667–4679, New York, NY, USA, 2017. Association for Computing Machinery.

[38] Radu-Daniel Vatavu, Lisa Anthony, and Jacob O. Wobbrock. Gestures as point clouds: A $p recognizer for user interface prototypes. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, ICMI '12, page 273–280, New York, NY, USA, 2012. Association for Computing Machinery.

[39] Radu-Daniel Vatavu, Lisa Anthony, and Jacob O. Wobbrock. $q: A super-quick, articulation-invariant stroke-gesture recognizer for low-resource devices. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '18, New York, NY, USA, 2018. Association for Computing Machinery.

[40] Sylvie Vidal and Grégoire Lefebvre. Gesture based interaction for visually-impaired people. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, NordiCHI '10, page 809–812, New York, NY, USA, 2010. Association for Computing Machinery.

[41] Georgios Yfantidis and Grigori Evreinov. Adaptive blind interaction technique for touchscreens. *Universal Access in the Information Society*, 4(4):328–337, may 2006.

[42] Xin Yi, Chun Yu, Weijie Xu, Xiaojun Bi, and Yuanchun Shi. Compass: Rotational keyboard on non-touch smartwatches. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 705–715, New York, NY, USA, 2017. Association for Computing Machinery.