UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

Human-Powered Smartphone Assistance for Blind People

" Documento Definitivo"

Doutoramento em Informática

Especialidade de Ciência da Computação

André Filipe Pereira Rodrigues

Tese orientada por:

Prof. Doutor Tiago Guerreiro, Prof. Doutor Kyle Montague

Documento especialmente elaborado para a obtenção do grau de doutor

2019

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

Human-Powered Smartphone Assistance for Blind People

Doutoramento em Informática

Especialidade de Ciência da Computação

André Filipe Pereira Rodrigues

Tese orientada por:

Prof. Doutor Tiago Guerreiro, Prof. Doutor Kyle Montague

Júri:
Presidente:

- Doutor Nuno Fuentecilla Maia Ferreira Neves, Professor Catedrático e Presidente do Departamento de Informática da Faculdade Ciências da Universidade de Lisboa

Vogais:

- Doutor Gerhard Weber, Professor da School of Engineering Sciences da Technische Universität Dresden (Alemanha)
- Doutor Hugo Alexandre Paredes Guedes da Silva, Professor Auxiliar com Agregação da Escola de Ciências e Tecnologia da Universidade de Trás-os-Montes e Alto Douro
- Doutor Manuel João Caneira Monteiro da Fonseca, Professor Associado da Faculdade de Ciências da Universidade de Lisboa
- Doutora Sara Alexandra Cordeiro Madeira, Professora Associada da Faculdade de Ciências da Universidade de Lisboa
- Doutor Tiago João Vieira Guerreiro, Professor Auxiliar da Faculdade de Ciências da Universidade de Lisboa (Orientador)

Documento especialmente elaborado para a obtenção do grau de doutor

Fundação para a Ciência e a Tecnologia (SFRH/BD/103935/2014; LISBOA-01-0145-FEDER-030347)

2019

Acknowledgements

About five years ago I started the long journey that my PhD has been. So far, every minute was worth it. I could not have imagined the variety of things I would be given the opportunity to partake. At this point I am not even sure I remember everything I got involved in throughout the years. Below there is a mixture of professional and strictly unprofessional acknowledgements. If you are reading with the sole intent of understanding my research I would like to invite you to skip this section unless you’re in it for some confusing statements which I want to stress in advance, may not be exactly 100% accurate or validated in any way.

First and foremost, I have to thank Tiago, my mentor, my friend, the Socrates to my Plato (no relation with the quality of the work). Similarly, it was Plato that had to materialize the work as Socrates was an enigmatic figure that spent time debating ideas and concepts with everyone particularly his disciples. Tiago mindset, dedication, care for others and hard work had a significant effect on the researcher and individual I am today, and I am grateful for it. Still, future research needs to be conducted to assess whether this effect will produce positive or negative outcomes. It has been a pleasure to work with you and I hope we continue to have the chance to do so. I could not have chosen a better advisor.

To Kyle and Hugo who I first met during my master thesis, I could have not picked anyone better to guide me through this journey. Some of the best times in the last few years was when we had the opportunity to be all together (i.e. all the SOS crew), whether that was for work or other more joyful occasions has a couple of weddings. I owe both of you a great deal, and I cannot imagine any better way to learn than discussing with you. You have been my Yin and Yang throughout this process, two very different researchers that have moulded the way I think about research. Nothing better than hearing three different people telling you three different, and apparently opposite things, and somehow having to figure out how to present all three in one cohesive thought. I hope we continue to work together in the years to come. For better or worse, the work presented in this thesis would not have been the same if you had not been involved.

A side-note thank you to Kyle as well for the time I spent at Open Lab in Newcastle. The three months I spent there were amazing, and you were one of the biggest reasons why. I am not just saying this because I am in my nighties. Thank you, Lesley, for being awesome. To Andy and Roxy thank you for your taking me in, the three months flew by because it was always a blast spending time with you.

During this thesis I have co-advised two master students, André Santos and Leonardo Camacho who I must thank for their contributions to the work presented herein. Leonardo has since started working in the industry, but André Santos has remained working with us and continue to contribute to the research I have been involved in. I have learned a lot with you Santos and we will hopefully continue to design, develop and fail miserably (*cough telephony*) with the solutions we like to build and evaluate.

To my lab roommates throughout the years, Tiago Gonçalves, Diogo Lima, José Coelho, David Costa, Daniel Costa and Diogo Marques thank you for your support and insights on what's to come. A particularly thank you to Diogo Marques from whom I have learned the most. Your insights, advice and challenges have often rocked my boat (in a research kind of way), leading me to strive to learn and question more.

To João Guerreiro who contributed to my master thesis work and was involved in some follow up studies who have led to this dissertation. I am sure we will have more chances to collaborate now that you are back, and I am looking forward to it.

I would like to thank everyone I had the pleasure to work with throughout these years in some capacity, whether that was directly associated with my PhD or not. Particularly Luis Carriço with whom I have collaborated since early on my master thesis. Moreover, I would like to also thank Carlos Duarte, and Manuel João, for the opportunities provided. I have learned a lot working with you in a variety of projects and look forward continuing to do so.

A special thank you to Carlos Bastardo from the Fundação Raquel and Martin Sain (FRMS), your valuable insights and your contributions to the execution of almost every single study presented herein were fundamental. This work would not have been possible if not for Carlos and FRMS support. Thank you, Norberto Sousa, from the Associação Promotora do Ensino dos Cegos (APEC) for always being available to discuss any topic. To both institutions, thank you for your support. I would like to thank all participants for putting up with me, particularly in the preliminary stages with flimsy prototypes, hopefully I will slowly be able to repay your time.

To my host institution, LASIGE, thank you for giving me all the necessary conditions to conduct my research and pursue my goals. To FCT for supporting my PhD for the last 4 years through the scholarship: SFRH/BD/103935/2014 and with LISBOA-01-0145-FEDER-030347.

To you Xana, thank you for the last-minute proofreading marathon. This work is substantially better due to your efforts. You rock!

Lastly, thank you mom and dad for the continuous support, I owe you the most for what I am today. To my siblings and friends thank you for being there whenever wherever.

I cannot end without thanking my wife or else I don't think I would make it to my defence. Thank you for being there at the end of every day. Even now when the end of the day means having to take off your shoes and give you a foot rub because of your dinosaur like feet. If all goes well, soon after delivering this, my wife will be 'delivering' too. A thank you in advance to Ema for putting a hard deadline on this delivery due to her own.

Dedicated to Ema and all my future children as I will not be doing any other PhD and I want to start by being a fair father.

Resumo

Os dispositivos móveis desempenham cada vez mais papéis no nosso dia-a-dia. Consequentemente, têm-se tornado instrumentos essenciais de inclusão e independência. No entanto, há ainda um longo caminho a percorrer e uma grande variedade de desafios a ultrapassar, nomeadamente, na sua acessibilidade para pessoas cegas.

Qualquer utilizador que adquira pela primeira vez um *smartphone* depara-se com um novo paradigma de interacção ao qual se terá de adaptar. Aprender a usar um *smartphone* não é uma tarefa fácil, especialmente tendo em conta a necessidade de utilização de novas aplicações que, com as suas constantes atualizações, exigem uma adaptação contínua.

Contrariamente aos telemóveis mais antigos que tinham um número finito de aplicações e funcionalidades, os *smartphones* têm a capacidade de as expandir através de novas aplicações disponibilizadas através de lojas digitais. A existência de um elevado número de *developers* com liberdade total de desenvolvimento faz com que a multiplicidade das aplicações e a variabilidade das interfaces seja imensa. Esta variedade acarreta, para utilizadores cegos, uma necessidade acrescida de criar um modelo mental para cada uma das aplicações.

Nesta dissertação começamos por caracterizar os desafios que pessoas cegas enfrentam na interação com os seus *smartphones*. Percebemos durante o processo, que têm que frequentemente que pedir ajuda à família, amigos, colegas e, por vezes, até a estranhos para conseguirem ultrapassar algumas das barreiras que enfrentam. Infelizmente, a ajuda nem sempre está disponível, especialmente se considerarmos a disparidade de cada desafio, a rede de suporte de cada um e o local onde se encontram. . Para utilizadores menos confiantes nas suas competências tecnológicas, ou mais receosos das consequências de usos indevidos, esta dependência na aprendizagem, por vezes põe em causa a adopção e o uso dos dispositivos.

Em alternativa à ajuda de outros, um utilizador pode recorrer a materiais de aprendizagem (tutoriais, vídeos, fóruns, manuais). Contudo, a quantidade destes materiais é reduzida e nenhum se adequa ao ensino em contexto móvel. Muitas das aplicações já incluem um tutorial que, numa primeira utilização, guia o utilizador pelas funcionalidades mais

relevantes do sistema. No entanto, estes tutoriais têm um conteúdo limitado e usam metáforas visuais, o que exclui utilizadores com deficiências visuais.

Ajudas em contexto podem facilitar o processo de aprendizagem (Grossman and Fitzmaurice 2010) permitindo ultrapassar desafios sem necessidade de recorrer à assistência de outros. Por exemplo, na *web*, soluções de Perguntas e Resposta em contexto permitem ao utilizador colocar questões a uma rede de voluntários prontos a responder (Chilana, Ko, and Wobbrock 2012). As soluções que tiram partido de trabalho humano são, por natureza, flexíveis na resposta, ultrapassando as capacidades de comunicação e compreensão dos atuais sistemas de ajuda automática. A computação humana tem sido usada em diversas abordagens para providenciar ajuda a utilizadores com deficiências visuais: desde permitir colocar questões visuais tirando fotos (Bigham et al. 2010) à criação de um mecanismo de criação de metadados para melhorar a acessibilidade *web* (Takagi et al. 2008).

Na primeira fase desta dissertação, executámos três estudos com o objetivo de detetar as dificuldades sentidas no uso de *smartphones*, começando pelo processo de transição. Procurámos também compreender os desafios encontrados por utilizadores com diferentes níveis de experiência e perícia. Para tal, realizámos um estudo longitudinal de dois meses onde seguimos cinco utilizadores invisuais no seu processo de transição para *smartphones.* Para perceber qual o grau de prevalência dos desafios enfrentados na variedade de dispositivos e sistemas operativos e se são ou não independentes da experiência dos utilizadores, realizámos workshops com 40 participantes cegos. Durante as sessões identificámos desafios e observámos como os participantes colaboraram e ultrapassaram os problemas. Para complementar a informação recolhida com dados de utilizadores mais experientes, analisámos os conteúdos mais discutidos nas duas maiores comunidades *online* dedicadas ao uso de *smartphones* por pessoas com deficiência visual. Constatámos que os problemas são prevalentes e embora mudem conforme a experiência do utilizador, a necessidade de recorrer a terceiros mantém-se.

Nestes primeiros estudos apercebemo-nos que, quando tinham problemas com os seus dispositivos, a maioria das pessoas procurava a ajuda de utilizadores experientes entre os seus amigos. No entanto, muitos outros não tinham ninguém a quem pudessem recorrer. Um dos objectivos desta investigação é estabelecer uma rede de suporte capaz de dar resposta à variedade de problemas encontrados.

Self-efficacy foi definida por Albert Bandura, como a crença nas nossas capacidades para completar tarefas de um domínio específico. Estas crenças nas nossas habilidades têm impacto em como perseguimos objectivos ou enfrentamos problemas. Para uma rede de suporte ser eficaz, tem de ter um impacto na *self-efficacy* dos seus utilizadores. Para tal, começámos por explorar se a computação humana poderia ser a resposta.

Com base no conhecimento adquirido na caracterização dos desafios enfrentados por pessoas cegas no uso de *smartphones*, nesta dissertação investigámos se, e como, ajudas em contexto baseadas em computação humana podem ser usadas para facilitar essa utilização. A tese proposta é:

A assistência para smartphones através de computação humana providenciada por não especialistas, é eficaz e afecta percepções de 'self-efficacy'.

Primeiro, investigámos a aceitação e a percepção da utilidade de um serviço de computação humana capaz de dar respostas quase perfeitas. Para tal, desenvolvemos um serviço que atua numa camada entre o sistema operativo e as aplicações. Este serviço permite aos utilizadores consultarem e colocarem questões a uma rede de voluntários (serviço de Q&A). Realizámos um estudo onde um voluntário normo-visual, especialista em acessibilidade móvel, respondia às questões colocadas através do serviço. Os resultados deste estudo sugerem que tecnologias baseadas em computação humana só seriam eficazes se os voluntários tivessem acesso ao contexto total do utilizador. Serviços de Q&A são, por natureza, ideais para lidar com problemas simples e inesperados. Eles não são desenhados, ou adequados, para suportar um utilizador que esteja continuamente a executar ou a aprender uma tarefa. Mais, são dependentes da capacidade dos utilizadores exprimirem os seus problemas e, assim sendo, o seu funcionamento assente em voluntários especialistas iria causar problemas de disponibilidade no suporte.

Em seguida explorámos como poderíamos desenhar ferramentas que dependessem apenas de voluntários não especialistas, e que fossem capazes de suportar o utilizador continuamente. Começámos por explorar tutoriais, dado que são desenhados para apoiar utilizadores a realizar uma tarefa na sua totalidade. Para identificar que informações poderíamos obter dos voluntários, realizámos um estudo onde os participantes criaram tutoriais. De seguida, realizámos uma sessão com participantes cegos onde pedimos que seguissem os tutoriais criados. Os nossos resultados sugerem que as instruções criadas por não especialistas normo-visuais e cegos são limitadas em diferentes aspectos, e que

não conduzem a uma aprendizagem eficaz da tarefa. Conseguimos identificar o tipo de informação providenciada pelos voluntários e o tipo de informação necessária aos utilizadores. Estes dois grupos de informação são frequentemente disjuntos.

Este estudo levou-nos a crer na possibilidade de desenhar soluções que suportam o processo de autoria, combinam contribuições, adaptam a apresentação de conteúdos ao perfil do utilizador, reagem ao contexto actual e são capazes de estar em constante atualização. Como conclusão, elaborámos um conjunto de recomendações para autoria e apresentação de tutoriais para invisuais.

Com o conhecimento adquirido nos estudos anteriores desenhámos e implementámos um assistente de tarefas em contexto, para *smartphones*, baseado nas contribuições de não-especialistas. Estávamos interessados em avaliar se a ferramenta criada era capaz de suportar utilizadores eficazmente, e se sim, que impacto essa assistência teria nas percepções dos mesmos. Para tal, recrutámos 12 pessoas normovisuais, sem formação em acessibilidade, para demonstrarem como realizar um conjunto de tarefas. A ferramenta desenvolvida guardava os dados gerados durante a demonstração para mais tarde ser capaz de criar assistentes de tarefas. Numa segunda fase, recrutámos 16 participantes cegos para completarem um conjunto de tarefas, umas com o auxílio da ferramenta desenvolvida e outras sem qualquer assistência. Os resultados mostraram que assistência baseada em computação humana, providenciada por não-especialistas, foi eficaz. Mais, a assistência teve um impacto positivo na taxa de sucesso dos participantes e afetou a percepção de *self-efficacy* dos mesmos. Estes resultados permitem-nos validar a tese proposta.

Palavras-chave: Acessibilidade, Computação Humana, Cego, Smartphone, Assistência

Abstract

Mobile devices are fundamental tools for inclusion and independence. Yet, there are still many open research issues in smartphone accessibility for blind people (Grussenmeyer and Folmer 2017). Currently, learning how to use a smartphone is non-trivial, especially when we consider that the need to learn new apps and accommodate to updates never ceases. When first transitioning from a basic feature-phone, people have to adapt to new paradigms of interaction. Where feature phones had a finite set of applications and functions, users can extend the possible functions and uses of a smartphone by installing new 3rd party applications. Moreover, the interconnectivity of these applications means that users can explore a seemingly endless set of workflows across applications. To that end, the fragmented nature of development on these devices results in users needing to create different mental models for each application. These characteristics make smartphone adoption a demanding task, as we found from our eight-week longitudinal study on smartphone adoption by blind people.

We conducted multiple studies to characterize the smartphone challenges that blind people face, and found people often require synchronous, co-located assistance from family, peers, friends, and even strangers to overcome the different barriers they face. However, help is not always available, especially when we consider the disparity in each barrier, individual support network and current location.

In this dissertation we investigated if and how in-context human-powered solutions can be leveraged to improve current smartphone accessibility and ease of use. Building on a comprehensive knowledge of the smartphone challenges faced and coping mechanisms employed by blind people, we explored how human-powered assistive technologies can facilitate use. The thesis of this dissertation is:

Human-powered smartphone assistance by non-experts is effective and impacts perceptions of self-efficacy.

Keywords: Accessibility, Human-Powered Computation, Blind, Smartphone, Assistance

Content

Figure List

Chapter 1

Introduction

Mobile subscriptions have reached a total of 7.9 billion in 2018[1]. Smartphones already account for 60% of all subscriptions and their market share keeps rising every quarter. Smartphones have become pivotal tools in many facets of our lives: they are our primary channel to access information, the way we stay socially connected, our personal assistants, and our entertainment platforms, to name a few. The inability to operate these devices is likely to have social and professional repercussions; it is therefore paramount to guarantee that they are accessible to everyone. For blind people, these devices also represent a new set of possibilities for assistive technologies: instead of having 10 different devices (e.g. colour identifier, label reader) people can carry one device which does it all and often more.

Since 2009, with the inclusion of VoiceOver, the first built-in screen reader on smartphones, we have witnessed an increase in smartphone adoption amongst blind people. According to WebAim surveys[2], 69% of the inquired screen reader users in 2015 also relied on a mobile screen reader. From those who did, 20% were still using feature phones or restrictive accessibility services on smartphones. There was a considerable share of the visually impaired population that had yet to fully transition to smartphones. In October 2017[3], 88% of the enquired screen reader users reported using a mobile screen reader. The substantial increase was associated with a less diverse sample of participants in the 2017 survey, and with a higher number of experts. Still, only 50.5% of beginners reported using mobile screen readers. Nevertheless, the number of mobile screen reader

---

[1]Ericsson Mobility Report, August 2018 (https://www.ericsson.com/assets/local/mobility-report/documents/2018/emr-q2-update-2018.pdf)

[2] Screen Reader User Survey, July 2015 (https://webaim.org/projects/screenreadersurvey6/)

[3] Screen Reader User Survey, October 2017 (https://webaim.org/projects/screenreadersurvey7/)

users keeps rising. Although some issues persist with the transition to smartphones, others have become more prevalent with the endless stream of new apps and quick update cycles, often with significant changes to use.

Upon the emergence of system-wide accessibility services on mainstream touch-based smartphones, a large part of accessibility research has been dedicated to the improvement of these services for blind people. However, past research has been limited to understanding gestures (Grussenmeyer and Folmer 2017; T. Guerreiro, Nicolau, and Jorge 2008; Kane, Wobbrock, and Ladner 2011; Oh, Kane, and Findlater 2013), text-entry (Azenkot, Wobbrock, et al. 2012; Azenkot and Lee 2013; Bonner et al. 2010; J. Guerreiro et al. 2015; Buzzi et al. 2014; Nicolau et al. 2014; Joao Oliveira et al. 2011; João Oliveira et al. 2011; Rodrigues et al. 2016; Romero et al. 2011) and other specific tasks (Azenkot, Rector, et al. 2012; Bigham et al. 2010; J. Guerreiro and Gonçalves 2014; Nicolau et al. 2015a) in confined laboratory studies. There was no understanding of blind people-s daily experiences with smartphones, how these devices are involved in everyday activities, and if and how challenges are surpassed in an acceptable time frame.

Smartphones are inherently associated with rich visual interfaces and touch screens, both of which present additional challenges. Undoubtedly, layout structure and feedback information are lost when using a screen reader. Consequently, people resist and delay transitioning to a smartphone at all costs, and fear for a time when external factors will force them to do so (Rodrigues et al. 2015). Yet, there seemed to be the unclaimed assertion that mobile devices were accessible.

Self-efficacy has been defined by Bandura (Bandura 2010) as *“people's beliefs about their capabilities to produce effects”*. This belief shapes the way one approaches its goals and faces challenges. However, self-efficacy has remained unexplored in the context of smartphone use, despite its potential impact in user’s autonomy and efficacy.

1.1 Mobile Accessibility for Blind People

Blind people often need synchronous, co-located assistance from family, peers, friends, and even strangers to overcome the different barriers they face (Rodrigues et al. 2015; Rodrigues, Montague, et al. 2017; Grussenmeyer and Folmer 2017). Although this approach tends to solve their issues, it has severe drawbacks. It inevitably creates a dependency on others, damaging their autonomy and self-efficacy. Moreover, help is not always available, especially when we consider the disparity in each barrier, individual support network and current location. Consequently, we found some people give up when struggling with smartphone adoption (Rodrigues et al. 2015), and even stockpile feature phones that are no longer manufactured. People should have the choice to not have to rely on others to learn, adapt and interact with their technology.

In this dissertation, we first focused on understanding how blind people are using smartphones in their daily lives, what challenges they face in their use and what are their go to coping mechanisms. Next we investigated how we could address the issues found. We focused our efforts in supporting user's autonomy, efficacy; and consequently, perceived self-efficacy.

1.2 A Human-Powered Approach

Traditionally help was only available co-located, but now it is often provided exclusive through technology. A human-powered approach enables us to design solutions that leverage the technology to mediate assistance, while having the flexibility to address unpredictable issues at scale.

Currently, services like BeMyEyes[4], Aira[5], VizWiz (Bigham et al. 2010), and others, are how blind people deal with a variety of challenges that occur in their daily lives. These human-powered services allow blind people to request assistance from volunteers, and paid workers, to deal with challenges, such as identifying medicine, reading digital displays or describing someone's appearance (E. Brady et al. 2013). Such services are

[4] BeMyEyes (https://www.bemyeyes.com/)

[5] Aira (https://aira.io/)

designed to support independent living, promoting autonomy. Depending on the complexity of the service provided, volunteers/workers may have to be trained or specialized individuals; which may limit the scale/availability of the service or require significant fees.

It is tempting to look for automatic assistance due to its inherent benefits of availability. However, the challenges blind people face, are often still out of reach to automatic approaches. For example, while OCR and object recognition are slowly becoming a reality through services like Google Lens[6], they still pose challenges for blind people (Gurari et al. 2018; Mahendru et al. 2017; Manduchi, Coughlan, and Ivanchenko 2008).

Technology has allowed people to increase their support network using novel assistive tools. Automatic solutions are, unfortunately, limited to specific contexts and cannot answer the variety of challenges users face. On the other hand, human powered solutions can potentially provide the flexibility needed. With non-expert volunteers we can guarantee the scalability of the solution, given that we design to accommodate the lack of knowledge from contributors. In this dissertation, we investigate if, and how, human-powered technologies can be leveraged to address the open challenges blind people face when interacting with their mobile devices. Furthermore, we will assess the impact such assistance can have in efficacy, users' perceptions of accessibility, and self-efficacy.

1.3 Terminology

In this dissertation, we use the term "*blind people*" to refer to people who are legally blind and have at most light perception. Furthermore, when we refer to screen reader users, it is implied that they are blind screen reader users.

We use the term "*human computation*" and "*human-powered*" interchangeably. Human computation overlaps in many aspects with the term crowdsourcing. Both are used to portray the work performed by people instead of technology, while mediated by it (Quinn and Bederson 2011). However, there is a key difference. Human computation reflects a paradigm where human work is essential to solve problems that technology cannot yet

[6] Google Lens (https://lens.google.com/)

solve. On the other hand, crowdsourcing is applied to every work that can be outsourced to an unidentified larger segment of people (Quinn and Bederson 2011). The term of human-powered access technology represents a particular segment of these solutions that supports access for people with disabilities (Bigham, Ladner, and Borodin 2011).

When we use the term *in-context* attributed to a solution developed, it symbolizes that users do not need to change device or the device state to use the solution. For example, if assistance is provided *in-context* when using an application, users do not need to close or change app in order to receive assistance.

1.4 Hypothesis & Goals

Building on a comprehensive knowledge of the smartphone challenges faced and coping mechanisms employed by blind people, we explored how human-powered assistive technologies can facilitate use. The thesis of this dissertation is:

Human-powered smartphone assistance by non-experts is effective and impacts perceptions of self-efficacy.

To accomplish our goal and validate the thesis we set out the following research goals.

1.4.1 Characterize mobile challenges

Our first studies investigated the smartphone adoption process to understand its challenges and influential factors. Next, we explored the barriers faced at all levels of expertise, identifying long standing issues and developed coping mechanisms. Lastly, we conducted a user study to identify the priorities and relevance of each identified issue.

1.4.2 Augment support networks through human computation

Tackling issues often requires the assistance of others. While requiring help is not a problem itself, not having available support is. Moreover, even if we assume help from others is always close, people will refrain from burdening their peers and friends with frequent requests.

In the current model, users depend on the availability and knowledge of their limited support network. We explored how to take advantage of volunteers with potentially no domain awareness to guarantee availability of assistance. To capitalize on every assistance provided, we investigated how to gather and repurpose the knowledge accrued.

1.4.3 Provide effective self-contained smartphone assistance

Users should be able to overcome every challenge presented by simply interacting with the device. Our goal was to develop in-context solutions that provide assistance, promoting user independence and self-learning. We first developed and explored an in-context question and answer (Q&A) service for any smartphone app. Then, for more complex tasks, we developed an authoring and playthrough service for interactive, in-context tutorials. Finally, we developed a task assistant that we have shown to be effective.

1.4.4 Assess the impact of effective assistance on users' perceptions of self-efficacy

We intended to go beyond traditional metrics and assess if the designed solutions empower its users. Effective smartphone assistance might not be enough if it does not have a positive impact on the user's perceptions of their ability to interact with the device. Human-powered assistance will only be an improvement if not perceived in the same way as co-located assistance from others. Our end-goal was to develop solutions that foster user independence and promote self-efficacy.

With the developed task assistant, we conducted a user study to assess how effective human-powered assistance affected users' perceptions, with a focus on self-efficacy.

1.5 Research Overview

To achieve the aforementioned goals, we first conducted a two-month user study where we followed closely the adoption process of five blind people, when transitioning from a feature phone to a smartphone. Through it, we understood the challenges, concerns and the process of adapting to a new device, which we describe in detail in *Chapter 3 Smartphone Adoption*. To verify if the issues are pervasive throughout the variety of devices, operating systems and user's expertise, we conducted a series of workshops with 40 blind participants at a local institution. During the sessions, we were able to observe the challenges faced, how participants collaborated, and discuss their traditional coping mechanisms. To deepen our understanding, particularly of expert users, at a larger scale, we analysed the contents of the top online community forums for visually impaired smartphone users. We found problems are pervasive, shift depending on expertise, and that the need for the assistance of others is prevalent. Our findings are reported in detail in chapters *Smartphone Adoption* and *Mobile Challenges.*

In our initial studies we observed that most people, whenever they had any problems with their device, would rely on tech savvy individuals within the community. However, there were many others that were not part of any support network and had no one knowledgeable on whom to rely. We set out to augment support networks to cover a wider range of people, while also alleviating the burden that is upon tech savvy people. We started by exploring if human-powered support could be what was missing.

With a characterization of the mobile challenges, we began to investigate the role human-powered technologies could play in assisting with smartphone tasks, in *Chapter 5 Human-Powered Support*. First, we explored the acceptance and perceived usefulness of a human-powered service if blind participants were given close to perfect answers. We developed a system-wide service that provides in-context questions and answers (Q&A). Using this service, we conducted a user study where a sighted expert, fully aware of the user context (e.g. current screen) and current thoughts (i.e. the expert was in the same room listening to the user) was providing answers through the service without any direct communication with the participant. Through the service, participants were able to successfully complete all the tasks. At this point, our findings suggested human-powered technologies could be effective only if volunteers were experts and ‘all knowing’. Questions were often not clear, and some issues could only be understood by observing users’ interactions. Q&A services are also only ideal to deal with simple and unforeseen issues. They are not designed to continually support a user when trying to learn or perform a task, but rather assist when everything else fails. Moreover, having to rely on experts posed a problem for the availability and, therefore, the feasibility of the solution.

We sought to understand how we could design human-powered assistance that would not require experts. To identify which information we could collect from non-experts, we conducted a user study where participants authored tutorials. Tutorials can continually support a user when trying to perform a specific task. Furthermore, they are easily repurposed to be used by everyone since they are task oriented, and it is possible to create them prior to user's needs, as well as in response to them. We conducted a session where we asked blind participants to follow the tutorials created. Our findings suggest that instructions authored by sighted and blind people are limited in different aspects, and that those limitations prevent effective learning of the task at hand. We identified the types of contents produced by authors and the information required by followers during playthrough, which often are not aligned. We provide insights on how to support both authoring and playthrough of nonvisual smartphone tutorials. This study has led us to believe that there is an opportunity to design solutions that mediate authoring, combine contributions, adapt to user profile, react to context and are living artefacts capable of continuous improvement.

We aimed to develop an effective self-contained smartphone assistance that would be able to leverage non-experts. We were interested in understanding the impact effective assistance would have on users' perceptions. Following the insights from the previous study, in *Chapter 6 - Designing a Rich Smartphone Assistant* , we discuss in detail the tool and the following user study. We recruited 12 sighted users with no accessibility training to demonstrate how to perform a set of tasks. With data collected both implicitly and explicitly, the tool generated task assistants. In a second phase, we recruited 16 blind participants which were asked to complete a set of tasks, with and without any assistance. We found the assistance provided had a positive effect on the participants' task success rate. When the assistance was effective, users felt more in control of the device, with reports of higher rates of self-efficacy. Users believed the assistant would enable them to do tasks faster, facilitate first time app uses and the exploration of new tasks.

In the last chapter of this dissertation, *Conclusion*, we revisit our contributions discussing the major results and implications. Additionally, we describe the limitations of the work, and outline a set of directions for future research in the field.

1.6 Contributions

This dissertation shows how to leverage non-experts' contributions to create effective assistance, that has a positive impact in users' external perceptions of accessibility, and internal perception of self-efficacy. We outline how to leverage non-experts to provide tech assistance for blind people and reveal its impact. The research that has led to this thesis provided the following contributions (C):

C1) An in-depth characterization of the smartphone adoption process of blind people. This analysis has unveiled the user perceptions prior, during and post adopting the device. It revealed a set of open challenges beyond the *stereotypical* issues reported in prior research. Results show that users delay transitioning to a smartphone, that the first contact is challenging, as users are confronted with both a lack of support and a set of inconsistencies in how the device/accessibility services behave, and that after 8 weeks users still face challenges (Rodrigues et al. 2015).

C2) A data collection framework to support mobile in-the-wild studies. There is a demand for mobile HCI evaluations in-the-wild, particularly when investigating the interactions of populations with highly variable abilities and diverse needs. During this thesis, we developed TinyBlackBox (TBB), our standalone mobile data collection framework, with analysis and interaction playback tools (Montague et al. 2015). We have successfully conducted a four-month user study with novice blind users, exploring their adoption experiences (Rodrigues et al. 2015) and text-entry performances (Nicolau et al. 2017) using TBB. The tool supported a deeper comprehension of the interactions and habits faced by the users in their day-to-day usage, along with a low-level analysis of their finer-grained interactions with the onscreen keyboard and text-entry behaviours. TBB has also been used by other research groups in the collection of data from people with Parkinson's disease and partially integrated in the AWARE framework (Vega 2019). TBB is open source[7].

C3) A catalogue of the open challenges blind people face when interacting with smartphones and their coping methods. Through a multiple methods approach, we identified and validated challenges locally with a diverse set of user expertise and devices, and at scale through the analyses of the largest Android and iOS dedicated forums for

[7] TinyBlackBox Github (http://goo.gl/VQTrJt)

blind people. We describe the coping mechanisms employed and contribute with a corpus of smartphone challenges for blind people.

C4) A comprehensive set of recommendations for the creation of non-visual smartphone assistance. We characterized which and how information is provided by amateur authors when creating non-visual tutorials. The lessons learned should be of interest to researchers and practitioners, working on the design of solutions that rely on untrained individuals to provide asynchronous technical assistance. We identified the different information required by users during playthrough when following instructions by others. Even though instructions were accurate, it was clear that users required additional assistance that was not present. When following a tutorial, the differences in users' expertise, interaction behaviours and preferences dictate the type of instruction adequate for each user (Rodrigues et al. 2019).

C5) A human-powered nonvisual task assistant that relies on non-expert's content authors. Based on the previous user studies, we developed a smartphone task assistance that relied on sighted authors with no accessibility knowledge to create content. We streamlined the authoring process by asking authors to contribute by exemplifying the tasks. Whenever possible, we rely on automatic data collection to minimize the amount of additional information required during authoring. Our task assistant monitors user interactions and has some basic knowledge of the app structure to ensure consistency of instruction delivery and flexibility of assistance. Our solution can be used to inform future designs in tech assistance.

C6) An assessment of the effectiveness of a human-powered nonvisual task assistance. We conducted a user study where we compared the task success rate between two user groups that performed the task with, and without, the developed assistant. We found the use of the assistant to be more effective than the alternative. Also, we provide insights on how to design solutions to improve assistance effectiveness.

C7) An assessment of the impact of effective task assistance in perceived self-efficacy. We evaluated the impact that a human-powered task assistant had on users' perceptions. We found the assistant positively affects the user's sense of self-efficacy related with their smartphone competence, with apps/devices ease of use and their overall accessibility. Tasks are perceived as simpler when the assistant is available. Future research, and particularly technology developers/manufacturers, should consider how impactful an effective system-wide assistant can be, for the perceived accessibility of their provided solutions.

1.7 Publications

Since I enrolled in my doctoral studies, I have co-authored a book chapter, and 18 peer reviewed papers, all of which in international journals or international conference proceedings. Although all publications played a role in my formation, I outline the ones that are related with mobile accessibility and highlight the ones directly associated with this dissertation.

1.7.1 International Conference Papers

1. **André Rodrigues**, Kyle Montague, Hugo Nicolau, Tiago Guerreiro. 2015. *Getting Smartphones to TalkBack: Understanding the Smartphone Adoption Process of Blind Users.* In Proceedings of the international ACM SIGACCESS conference on Computers and accessibility (ASSETS '15). ACM, NY, USA, 23-32.

2. Kyle Montague, **André Rodrigues**, Hugo Nicolau, and Tiago Guerreiro. 2015. *TinyBlackBox: Supporting Mobile In-the-Wild Studies.* In Proc. of the international ACM SIGACCESS conference on Computers and accessibility (Assets '15), 379-380. [Poster]

3. Hugo Nicolau, Kyle Montague, Tiago Guerreiro, **André Rodrigues**, and Vicki L. Hanson. 2015. *Typing Performance of Blind Users: An Analysis of Touch Behaviors, Learning Effect, and In-Situ Usage*. In Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS '15). ACM, New York, NY, USA, 273-280.

4. **André Rodrigues**, Hugo Nicolau, Kyle Montague, Luís Carriço, and Tiago Guerreiro. 2016. *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). ACM, New York, NY, USA, 47-52.

5. **André Rodrigues**, Kyle Montague, Hugo Nicolau, João Guerreiro, Tiago Guerreiro. 2017. *In-context Q&A to Support Blind People Using Smartphones.* In Proceedings of the international ACM SIGACCESS conference on Computers and accessibility (ASSETS '17).

6. **André Rodrigues**, André Santos, Kyle Montague, and Tiago Guerreiro. 2017. Improving Smartphone Accessibility with Personalizable Static Overlays. In Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '17). ACM, New York, NY, USA, 37-41.

7. **André Rodrigues.** 2018. *Facilitating smartphone use for blind people through a human powered approach*. SIGACCESS Access. Comput. 120 (January 2018), 28-31. [Doctoral Consortium - Newsletter]

8. **André Rodrigues**, Kyle Montague, Tiago Guerreiro. 2018 *Data Donors: Sharing Knowledge for Mobile Accessibility*. In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)

9. Daniel Trindade, **André Rodrigues**, Tiago Guerreiro, and Hugo Nicolau. 2018. Hybrid-Brailler: Combining Physical and Gestural Interaction for Mobile Braille Input and Editing. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM, New York, NY, USA, Paper 27, 12 pages. DOI: https://doi.org/10.1145/3173574.3173601

10. **André Rodrigues**, Leonardo Camacho, Hugo Nicolau, Kyle Montague, and Tiago Guerreiro. 2018. *Aidme: interactive non-visual smartphone tutorials*. In Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '18). ACM, New York, NY, USA, 205-212. [Poster]

11. **André Rodrigues**, André Santos, Kyle Montague, Hugo Nicolau, Tiago Guerreiro. 2019. *Understanding the Authoring and Playthrough of Nonvisual Smartphone Tutorials.* In Human-Computer Interaction – INTERACT 2019 (Lecture Notes of Computer Science)

12. Hugo Nicolau, **André Rodrigues**, André Santos, Tiago Guerreiro, Kyle Montague and João Guerreiro. 2019 *The Design Space of Nonvisual Word Completion.* In Proceedings of the international ACM SIGACCESS conference on Computers and accessibility (ASSETS '19)

1.7.2 International Journal Paper

1. Hugo Nicolau, Kyle Montague, Tiago Guerreiro, **André Rodrigues**, and Vicki L. Hanson. 2017. Investigating Laboratory and Everyday Typing Performance of Blind Users. ACM Trans. Access. Comput. 10, 1, Article 4 (March 2017), 26 pages.

1.7.3 Book Chapter

1. Tiago Guerreiro, Luís Carriço, and **André Rodrigues**. 2019. *Mobile Web. In Web Accessibility: A Foundation for Research*, edited by Yeliz Yesilada and Simon Harper, 737–54. London: Springer London.

1.8 Dissertation Outline

This dissertation is organized in eight chapters. The following chapter reviews the state of art of smartphone accessibility for blind people, how in-context assistance has been leveraged in other context, and lastly, how human-powered access technology has been used to support blind people.

Chapter 3 describes the longitudinal study conducted to characterize the smartphone adoption process of blind people. In this chapter we discuss the challenges faced and coping mechanisms employed during the process. In the next chapter we conduct a detailed analysis of the mobile challenges faced by blind people beyond adoption. To understand if challenges were pervasive throughout different devices and expertise level, we conducted two studies that are detailed in this chapter.

With an in-depth assessment of the challenges people faced, in *Chapter 5 - Human-Powered Support* we describe our efforts in exploring human-powered support to address them. We present two studies that provide a comprehensive set of insights for the creation of nonvisual smartphone assistance. We report our findings on the perceived usefulness and acceptance of human-powered Q&A service, when the assistance is provided by a domain expert with full context knowledge. Next, we describe the limitations and opportunities when we rely on non-experts to create nonvisual tutorials.

In *Chapter 6 - Designing a Rich Smartphone Assistant* we outline in detail our developed human-powered solution based on the findings from the previous chapter. In *Chapter 7 - Evaluating Efficacy and Perceived Self-Efficacy,* we present a comparative assessment of the effectiveness of our human-powered nonvisual task assistant, when compared with no assistance at all. Furthermore, we describe how users perceived the assistance provided, and how it impacted their self-assessed self-efficacy.

Lastly, *Chapter 8 - Conclusion* concludes this dissertation with a summary of its main contribution; an outline of the design implications for design of human-powered nonvisual technology assistance; a discussion of the limitations of this work; and a set of promising research directions.

Chapter 2

Related Work

Smartphones are prevalent in our modern societies. Their capabilities have changed the way we interact, collaborate and engage with other people. As a result, Mobile Human-Computer Interaction research is an active and popular topic.

In this chapter, we outline the field of mobile accessibility for blind people. We provide an overview of the evolution of the approaches taken. Then, we report in detail the research conducted in nonvisual input and identify the existing knowledge gap in a broader understanding of smartphone accessibility.

In this dissertation, we explored the use of in-context human powered assistance. Thus, we portray how in-context solutions have been leveraged in the past to provide assistance. Lastly, we discuss previous work on web and mobile human-powered access technologies. In each section, we discuss the shortcomings of current work, motivating the need for human-powered approaches to support smartphone accessibility.

2.1 Understanding Smartphone Accessibility

Every day, mobile phones with physical keys become scarcer in our markets, particularly those with accessibility features (e.g. screen reader). With the growth of the smartphone market, and the inevitably associated touchscreen, devices have become rich visual interfaces and lost almost all tactile feedback. Researchers soon became aware of the opportunities and challenges smartphones would bring to blind people. The first

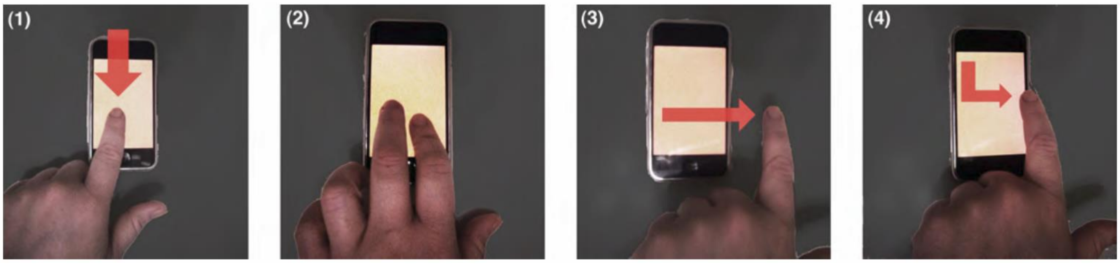


Figure 1- Slide Rule uses multi-touch gestures to interact with applications. (1) A one-finger scan is used to browse lists; (2) A second-finger tap is used to select items; (3) A flick gesture is used to flip between pages of items or a currently playing song; (4) An L-select gesture is used to browse the hierarchy of artists and songs in the music player. (Kane, Bigham, e Wobbrock 2008)

challenges were how to convert the rich visual interfaces to audio ones and how to map direct manipulation to accessible gestures.

In Slide Rule (Kane, Bigham, and Wobbrock 2008), researchers developed and evaluated a set of multi-touch interaction techniques to provide non-visual access to multiple applications (Figure 1). It reported on the results of five gestures: one finger scan, split-tap to select (i.e. use one finger to scan and tap with a second one to insert), directional flicks for additional options, L gesture to scan hierarchical content, and double tap to select. The evaluation revealed Slide Rule to be significantly faster and preferred to the opposing method, a Pocket PC application with Mobile Speak Pocket, where navigation was accomplished with a four-way navigation control pad. In McGookin et al. (McGookin, Brewster, and Jiang 2008), researchers designed and evaluated, in a control lab experiment, a different set of gestures. Participants were able to flick right or left to navigate the items and tap anywhere to select the focused item. The paper outlines several design implications on the accessibility of touchscreens (e.g. “Avoid *localised* gestures or provide touchscreen awareness”).

Smartphones are now the norm and touchscreens are the mainstream input method that accompanies them. As such, researchers took an interest in further understanding gestural interaction for blind people. Kane et al. (Kane, Wobbrock, and Ladner 2011) reported the

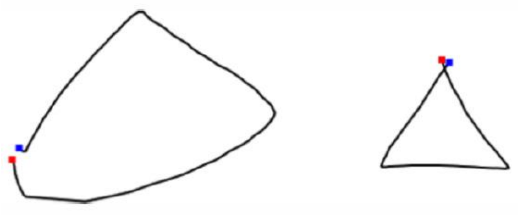


Figure 2 - Two representative versions of a triangle gesture produced by a blind person (left) and a sighted person (right). (Kane et al. 2011).

results of two lab studies to understand the touchscreen gestures preferences and usability by people with visual impairments. To do so, the studies compared the gestures performed by blind people with the ones performed by sighted people. In the first study, users were asked to create their own gestures to perform common tasks (e.g. copy and paste). People with visual impairments created significantly different gestures, they relied more often on two handed gestures and the edges of the devices. In the second study, the two groups performed the same set of gestures (Figure 2). Researchers found significant differences in the size, speed and shape of the gestures. Additionally, blind people had a wider variation in size. Oh et al. (Oh, Kane, e Findlater 2013) explored how to use the sonification of gestures to provide an understanding of the desired shape of a touchscreen

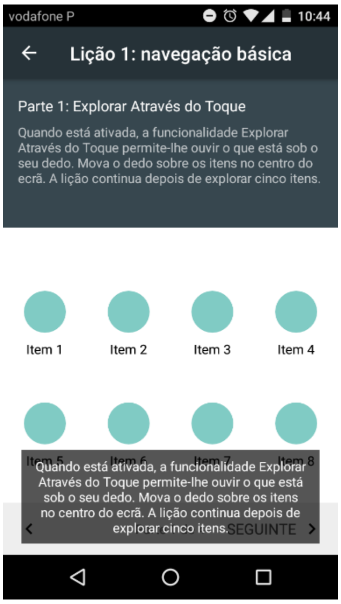


Figure 3 - Talkback first lesson on how to use explore by touch.

gesture. To do so, the user would hear a different sound depending on the direction they needed to move their finger. Even with a greater understanding on gesture interaction, it remained a difficult challenge for blind people to learn how to perform them.

2.1.1 System-Wide Accessibility

In 2009, Apple released the iPhone 3GS, the first smartphone shipped with a built-in screen reader, *VoiceOver*. Users were now able to navigate the entire system with a set of defined gestures highly inspired in Slide Rule (Kane, Bigham, and Wobbrock 2008). They could flick to focus the previous or next item or simply scan the screen with one finger and the content displayed below would be read (Figure 3); to select, users had to double tap. Consequently, smartphones started to become popular amongst blind people. Nowadays, the two major mobile operating systems, Android and iOS, come with screen readers (Talkback and Voiceover) and other accessibility features with very similar behaviours and interactions. These small devices brought forward a set of accessibility tools that were previously available only through expensive assistive technology (Narasimhan, Gandhi, and Rossi 2009). A single device now enables users to read labels through OCR (i.e. Optical Character Recognition), identify colours or objects all by using different apps, many freely available in app stores.

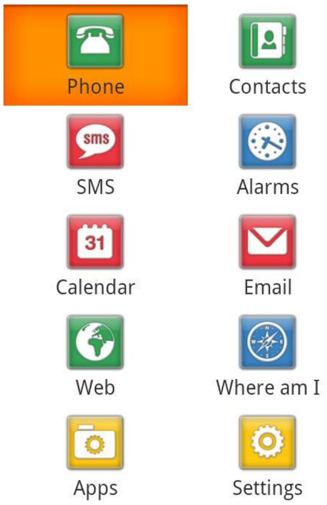


Figure 4 - Mobile Accessibility application interface.

2.1.2 Custom-Made

System-wide screen readers were not the only approach towards smartphone accessibility. Applications suites, such as *Mobile Accessibility*,[8] were developed specifically for visually impaired people. This approach intended to substitute the entire system and limit the user to the set of provided applications (Figure 4). The benefit is that it guarantees a higher degree of accessibility, and particularly simplicity, to the applications provided. Conversely, it has the drawback of restricting the user control over the phone, negating all the benefits provided by the thousands of different applications available in the smartphone ecosystem. Custom and implicitly restrictive assistive

[8] Mobile Accessibility suite for Android, (http://codefactoryglobal.com/app-store/mobile-accessibility/)

technologies can also have the adverse effect of stigmatizing people, that would rather be able to use mainstream applications and devices (Shinohara and Wobbrock 2011).

2.1.3 Adapting Content

In a crossover between custom-made approaches and system wide services, one can adapt how content is rendered and navigated to maximize its accessibility (Rodrigues, Santos, et al. 2017; D. Zhang et al. 2017; X. Zhang et al. 2017).

Zhang et al. (X. Zhang et al. 2017) proposes the use of interaction proxies to allow third party developers to address apps' accessibility issues. In a similar approach for the mobile web, Zhang et al. (D. Zhang et al. 2017) proposed adapting web pages to personalizable template hierarchical structures, enabling users to quickly skim content. However, for a successful adaptation, one must be aware of each application's issues and adapt it to the end user specific needs without compromising any features. Thus, it relies on a limited population of third-party developers. Moreover, apps can be fully accessible and still pose a challenge for the untrained user.

In Rodrigues et al. (Rodrigues, Santos, et al. 2017), we divided every screen in two halves, dedicating the top half to a set of fixed options, emulating the quick access options of tactile keyboards on feature phones. The bottom half was dedicated to navigating the content present on the screen, accordingly to the ordering and filtering preferences defined. The adaptation of every screen to a fixed layout, reduced the complexity of interacting with a smartphone and was seen by participants as a steppingstone in the learning process. The creation of these types of system-wide adaptive services is already possible in mobile operating systems, such as Android[9].

2.1.4 Text-Entry

Text-entry on touchscreens is one of the most tackled problems (Azenkot, Wobbrock, et al. 2012; Azenkot and Lee 2013; Bonner et al. 2010; Buzzi et al. 2014; J. Guerreiro et al. 2015; Nicolau et al. 2014; Joao Oliveira et al. 2011; Rodrigues et al. 2016; Romero et al.

[9] Create an accessibility service (https://developer.android.com/guide/topics/ui/accessibility/service)

2011) in the accessibility of smartphones for blind people. The loss of the physical keys in favour of on-screen virtual QWERTY keyboards hindered the interaction. Users can no longer quickly grasp keys through tactile feedback. Instead, one must wait for audio feedback from the screen reader to hear what is currently on focus. People now have to accurately reach a vast number of small targets; they must rely on their spatial ability to become proficient typists. Several studies focused in understanding text-entry needs on touchscreens (Nicolau et al. 2017; Rodrigues et al. 2016), while others looked for alternative interaction methods, many of which based on Braille (Southern et al. 2012; Romero et al. 2011; João Oliveira et al. 2011; Nicolau et al. 2014; 2015a; Azenkot, Wobbrock, et al. 2012).

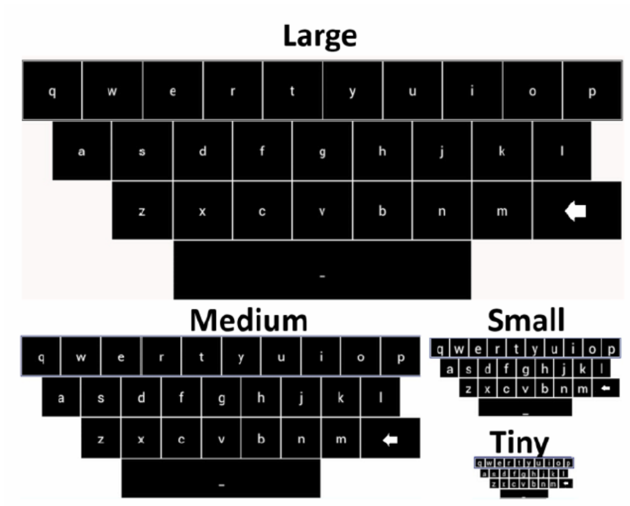


Figure 5 - The keyboard conditions and their relative sizes. Large representing a standard 10' tablet to the tiny representing a wearable smartwatch (Rodrigues et. al. 2016)

Understanding Text-Entry Needs

Text-entry has been a focus of research due to the underwhelming performance blind people achieve with touchscreens when compared to their old keypad phones. Additional efforts have been made to understand and report on the diverse needs of blind users given their variety of abilities (Joao Oliveira et al. 2011). As mobile devices progress to different form factors (Figure 5), it becomes relevant to understand the effects of target sizes. In our previous work (Rodrigues et al. 2016), we established an upper limit between 10 and 15 mm to the benefits of larger targets, and report a significant decrease in performance for targets below 10 mm, with targets of 2.5 mm being considered unusable.

Braille & Other Alternative Methods

In BrailleType (João Oliveira et al. 2011), the touchscreen had six areas representing the dots in a braille cell. To enter a character, users had to tap or drag their finger into the target, and they would hear the respective cell number; to confirm and insert a character, users had to double tap anywhere on the screen. Although, BrailleType was found to be slower when compared to *VoiceOver*, it produced less errors and provided a simpler interface. Similarly, BrailleTouch (Romero et al. 2011) provided the six targets but required users to use multi-touch to select; upon release, the corresponding character would be inserted. The device had to be held landscape, with the back of the device facing the user for the interaction to take place.

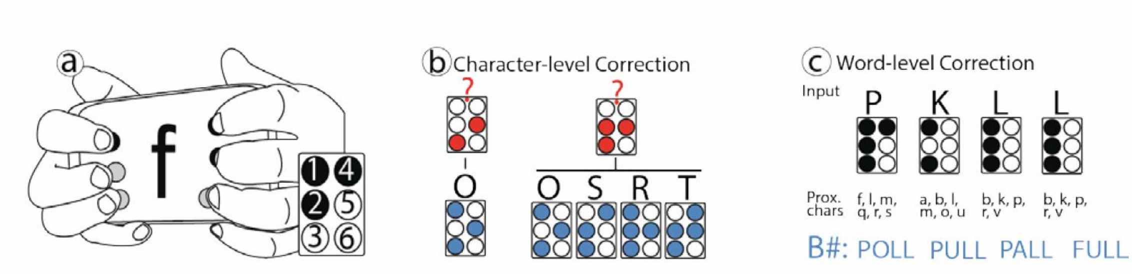


Figure 6 - B# is a novel correction system for multitouch Braille input. (a) The user types the letter 'f'. (b) Character-level correction; the closest characters in terms of Braille distance for 2 unidentified chords. (c) Word-level correction; top suggestions return by B# considering the letters that are at a Braille distance of one from the entered chord. (Nicolau et al. 2014)

With BrailleTouch, participants achieve speeds comparable with physical Braille keyboards, at around 20 words per minute. However, chord-based approaches like BrailleTouch have reported high error rates. With B# (Nicolau et al. 2014), researchers developed a novel chord-based spellchecker addressing the most common chording errors. B# leveraged a distance of similarity (Figure 6) between the chords inputted and chord-level representations of characters, and the words in the dictionary, improving upon current spellcheckers and achieving a higher input accuracy.

Other works have tried alternative interaction methods. With NavTouch (T. Guerreiro, Nicolau, and Jorge 2008), the users used a sequence of gestures to navigate in a layout based around the alphabet (i.e. characters were accessible through a table layout where vowels were at the start of every row). In No-Look Notes (Bonner et al. 2010), the screen is divided into sections containing the letters available on previous 12-key multitap keypad layouts (e.g. one - abc). To enter a character, the user would have to select a section and split-tap; the screen would then present a list with the characters of that section and the user had to split-tap again to select. Guerreiro et al. proposed a novel interaction method that leveraged users' previous QWERTY knowledge (J. Guerreiro et al. 2015). It allowed users to rest their fingers and use multi-touch to entry text by spatializing the audio feedback received.

Although multiple approaches have been explored through laboratory studies, and all found text-entry to be slower than with previous feature phones, we have yet to understand how users first engage with smartphones and what problems they face when doing so, even with the default methods.

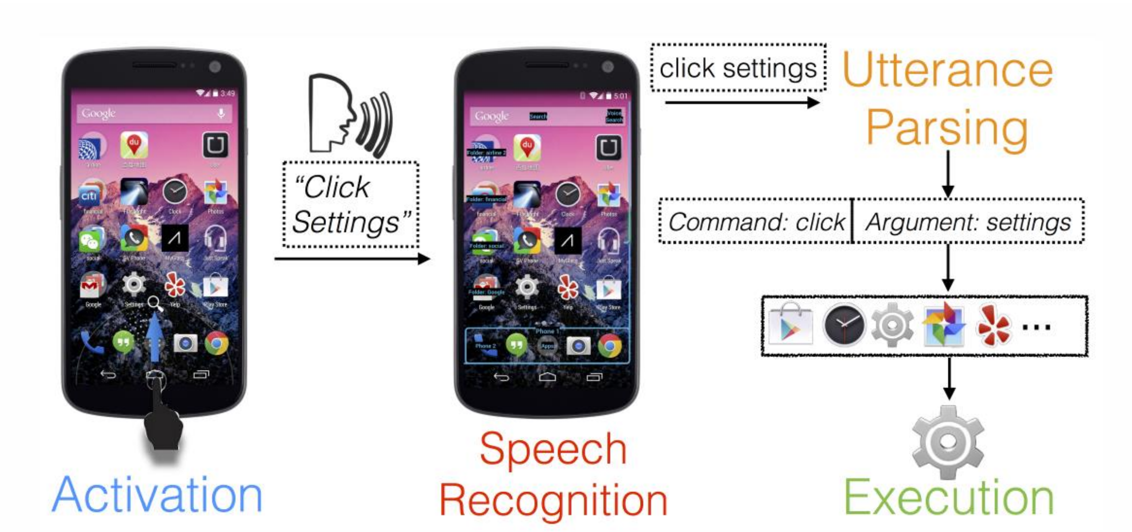


Figure 7 - When activated, JustSpeak records audio spoken by the user, transcribes it into plain text, then parses the text into formal commands, and then finally finds the correct object on the screen to perform the command. (Zhong et al. 2014)

2.1.5 Speech

Speech is an alternative to text-entry methods, but it can also provide users with additional navigation controls. In a survey conducted by Azenkot et al., 90% of the participants reported to have recently used dictation (Azenkot and Lee 2013). Furthermore, visually impaired participants were more satisfied with it in comparison with sighted participants, even though they spent up to 80% of their time editing inserted input. Mobile assistants are becoming more prevalent (i.e. Siri, Google Assistant, Cortana) and all allow the users additional navigational control using voice commands. JustSpeak (Zhong et al. 2014) went a step further, and instead of allowing only a fixed set of commands to be available, it opens full control over the interface (Figure 7). It allows for the user to provide, with a single utterance, any sequence of commands available in the current interface to be performed (i.e. "Open Gmail then refresh"). In 2018, Google released *Voice Access*[10] which enables users to interact with the device solely through voice interaction. Users can issue navigation and gesture commands (e.g. "back", "scroll down"), dictate, and select items based on their textual description or number displayed on the screen. Although *Voice Access* is targeted at motor impaired users who struggle to interact, it can also be beneficial for blind users if integrated with current screen readers.

Still, speech input is limited by the surrounding environment and user context. It provides a solution only for noise-free environments where the user is not concerned with privacy issues such as eavesdropping.

Figure 8 - The system consists of six vibrotactile motors attached to springs and a 3D-printed case. The springs mould to users' hands and dampen vibrations through the device allowing better stimuli discrimination. (Nicolau et al. 2015)

2.1.6 Privacy

Smartphones are inherently more private devices given the amount of sensible information they contain. Blind people face additional challenges when it comes to guarantee their privacy and security. Since it is harder to assess their surroundings, they are more susceptible to shoulder surfing attacks and the simple eavesdropping. Yet, authentication mechanisms have been reported to pose significant challenges to blind people (Damaceno, Braga, and Mena-Chalco 2018). Moreover, users are often unaware of the dangers and the precaution they could take (Azenkot, Rector, et al. 2012). Aware of these issues, researchers have investigated alternative authentication mechanisms (Azenkot, Rector, et al. 2012; Marques, Carriço, and Guerreiro 2015). In both works the

[10]Voice Access, (play.google.com/store/apps/details?id=com.google.android.apps.accessibility.voiceaccess)

tap pattern on a blackout touchscreen is used for authentication. In PassChords (Azenkot, Rector, et al. 2012) the pattern is defined by the set of fingers used in each tap. In Marques et al. the password is defined by the a tapping sequence where both the time pressed and the release time are taken into account (Marques, Carriço, and Guerreiro 2015).

Others have looked at how additional hardware could improve users' privacy, particularly on how one could support inconspicuous interactions beyond authentication (Nicolau et al. 2013; 2015a). In this series of works, the researchers explored how to use vibro-tactile devices to enable blind people to read. In UbiBraille (Nicolau et al. 2013), six vibro-tactile actuators are used to code a Braille cell and communicate single characters, and the actuators were attached to each index, middle, and ring fingers. Similarly, Holibraille (Nicolau et al. 2015a) presents a vibro-tactile device that can be attached to smartphones, allowing users to identify braille letters by resting the fingers on the smartphone case actuators (Figure 8). The results have shown that word recognition is possible through haptics alone by leveraging past braille knowledge.

Previous work has explored, again, in laboratory settings how to improve privacy in current smartphones. However, there is currently a gap in knowledge in how users perceive their security and privacy when interacting with their smartphone in-the-wild. Moreover, we need to understand what the fears and expectations around smartphones by unexperienced users are, and how it might impact their use of the device.

2.1.7 Understanding Challenges

Through interviews and participant diaries, Kane et al. assessed the challenges faced by people with visual and motor impairments when using mobile devices in the early days, when smartphones had physical buttons (Kane et al. 2009). Leporini et al. conducted an expert evaluation of VoiceOver on a mobile device with three accessibility experts, one of which blind (Leporini, Buzzi, and Buzzi 2012). The identified issues were used to create an online survey answered by 55 blind participants, corroborating the findings. Participants reported issues with text being hard to write, some interactive elements being hard to distinguish, and focus issues on forms. Pal et al. used surveys and interviews to ask blind smartphone users what their devices enabled them to do and how they were taking advantage of these capabilities (Pal et al. 2017). The paper went on to argue for the need to understand human agency within the necessity of adopting technology. This research provided insights about real-world technology usage by individuals with disabilities. For example, the first contact with AT technologies in low-mid incoming countries primarily happens in non-profit organizations. Currently, the data gathered has been limited to a particular time-window, dependent on self-reporting and/or on the reflective capabilities of participants/experts to elaborate on their experiences. Moreover, when we take into consideration that one of the challenges blind people face when interacting with content is "not knowing what they do not know" (Bigham, Lin, and Savage 2017), we realize the need to go beyond reflective approaches.

In recent years, there have been projects that review the current state of smartphone accessibility for blind users (Damaceno, Braga, and Mena-Chalco 2018; Grussenmeyer and Folmer 2017). These already reference some of the work presented in this dissertation. Both identify a list of challenges gathered from past research for visually impaired people. Grussenmeyer et al. with a focus on touchscreen accessibility (Grussenmeyer and Folmer 2017) and Damaceno et al. on mobile devices (Damaceno, Braga, and Mena-Chalco 2018). While some challenges have been identified in prior work, so has the need for a real-world assessment of the usage of mobile technologies (Grussenmeyer and Folmer 2017).

2.1.8 Discussion

Anecdotally, we observed that several blind people continue to use their older feature phones and fear for a keyless future (Buzzi et al. 2014). There is still a considerable number of users that have yet to transition to smartphones. For others, interacting with these devices can still be confusing and challenging to master.

Research in smartphone accessibility has been limited to laboratory experiments. Moreover, the studies presented had users engage in artificially created tasks (e.g. perform a swipe, write the phrase X, perform task X) for short periods of time to assess the proposed hypothesis. Current research may be providing a superficial view of the challenges users face when interacting with smartphones. Most of the research has also been focused on text-entry and alternative input mechanisms. However, there was a knowledge gap in the understanding of the adoption process and the day to day issues blind people face when interacting with their devices. Still, none of the identified works has considered the challenges users face when transitioning to smartphones; how challenges evolve or differ with user expertise; identified coping mechanisms; or established challenges' relevance.

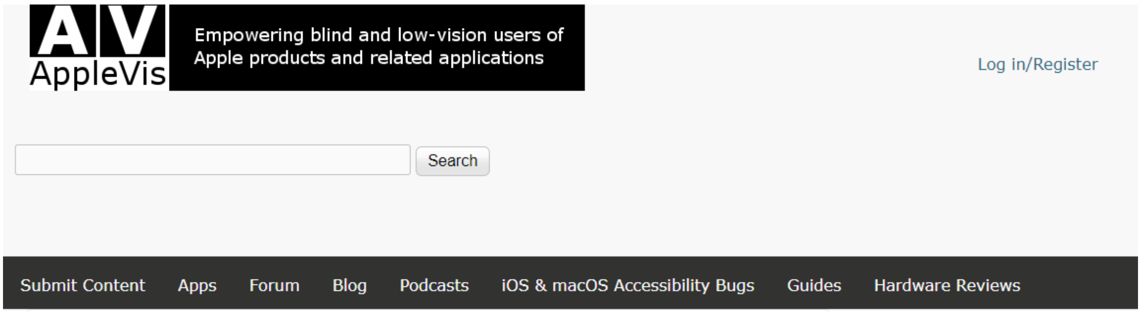


Figure 9 - AppleVis webpage menu with a variety of feature, from forums to podcasts on Apple products accessibility.

2.2 In-context Assistance

When faced with a challenge, users can look for answers online. Online assistance comes in a variety of forms: blogs, forums, tutorials, videos and podcasts (Figure 9). Except for the forums, the other media provide static information and rely on the user ability to find and follow the content. There are large community forums dedicated to smartphones for blind users (e.g. *Eyes-Free*[11], *Applevis*[12], *Viphone*[13]), where people can create and discuss topics freely. These online assistance methods depend on the user ability to switch between contexts to find an answer or even rely on a secondary device (e.g. desktop). To find help, people need to know how to describe the problem at hand.

In-context assistance has been reported in past research on desktop applications to have performance benefits (Chilana, Ko, and Wobbrock 2012; Grossman and Fitzmaurice 2010; Fernquist, Grossman, and Fitzmaurice 2011; Hagiya et al. 2015; Hailpern, Reid, and Boardman 2009; Wang et al. 2014). Yet, in the smartphone ecosystem we have mostly overlooked its relevance.

[11] Google Group for Blind Android Users, (https://goo.gl/uiib1F)

[12] Community-Powered Website for Blind Apple Users, (https://www.applevis.com/)

[13] Google Group for Blind Apple Users, (https://goo.gl/g4YiQZ)

2.2.1 Static Assistance

In a photo manipulation application, researchers investigated quick contextual access to documentation and short videos (Grossman and Fitzmaurice 2010). Tooltips were successfully integrated in the participants' workflow, and after a week, users were performing the tasks significantly faster than in the control condition. Chilana et al. developed LemonAid, a tool that provided contextual Q&A for web applications. Users could ask, answer and browse questions by selecting interface elements (Chilana, Ko, and Wobbrock 2012). LemonAid's selection-based interaction was able to retrieve results for 90% of the queries and, for 57%, a relevant result was within the top two. Although the previous solutions were not motivated by visual impairments, they show how in-context assistance can be leveraged. In Hailpern et al. , help and hints were added in-line with the content on the Gmail web application, allowing the screen reader to quickly access help within-context (Hailpern, Reid, and Boardman 2009).

Figure 10- User interface without (above) and with (below) a stencil. (Kelleher et al. 2005)

2.2.2 Interactive Assistance

In-context assistance can be more than just enriching the interfaces with static help content; it can provide interactive learning artefacts (e.g. tutorials). In-context interactive tutorials can improve the user ability to quickly learn a new task. Kelleher et al. investigated an interaction technique for presenting in-context tutorials (Kelleher and Pausch 2005). The technique, now common in smartphone onboarding tutorials, uses an overlay to obscure the non-relevant content and restrict user interaction (Figure 10). In the user study, participants were 26% faster at completing the tutorial than with the baseline condition (i.e. paper tutorials). Interactive tutorials can be applied to a variety of contexts and population. Hagiya et al. reported a text-entry tutorial for older adults that detects errors and provides instructions to correct them (Hagiya et al. 2015). It detected when the user was taking too long to type the next letter/word and provided instructions simultaneously through voice, text and finger animations. The tutoring system increased typing speed by 17% and reduced errors by 59%. With the popularization of graphical user interfaces (GUI) in the 90's, researchers (Weber et al. 1994) explored how to train blind users in this new paradigm of interaction. First, by relying on tactile graphics and models of the GUI, followed by using a text editor designed to introduce a variety of interface elements through an interactive step by step tutorial. The work argued for the need of flexible tutorial systems to enable users to reap the benefits of GUIs.

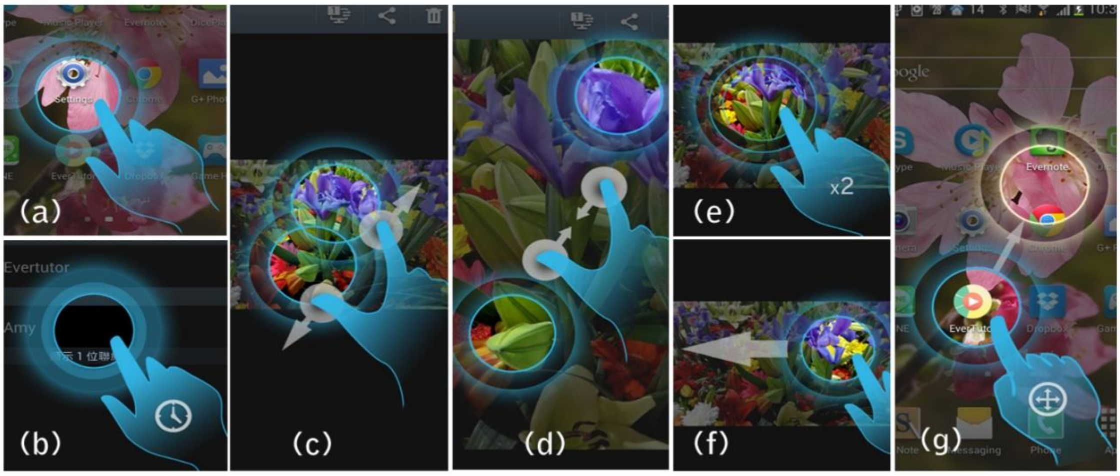


Figure 11- The overlays for touchscreen gestures in EverTutor. (a) Touch (b) Long press (c, d) Pinch open/close (e) Double touch (f) Swipe (g) Drag. (Wang et al. 2014)

Tutorials can be designed to be engaging experiences to boost user performance. Li et al. report on gamified in-app interactive tutorials for AutoCAD first time users. Users reported higher engagement, had 10% higher completion rate and were 20-76% faster completing the task (Li, Grossman, and Fitzmaurice 2012). In Fernquist et al., the system guides users through the interface, providing assistance while sketching (Fernquist, Grossman, and Fitzmaurice 2011). Using a step navigation dialog, in each step the user is shown how, where, and when to change settings and when and where to draw. Yet, interactive tutorials should not restrict the user to following the steps. In Lieberman et al. the authors argue that at different points during the tutorial, users might wish different levels of assistance (e.g. let me do it, show me how, guide me through it) (Lieberman, Rosenzweig, and Fry 2014). At each step, the user could delve into the particularities of the step and freely navigate between steps.

On smartphones, with EverTutor (Wang et al. 2014), researchers have investigated how to broaden the reach of interactive tutorials by allowing the creation of system-wide tutorials from user demonstration. The system records low-level touch events, detects gestures and identifies on-screen targets using computer vision techniques. When a tutorial is played, it uses overlays with visual metaphors, and an obscuring overlay to convey the next target and gesture in-context (Figure 11); additionally, it prevents users from performing incorrect steps. Unfortunately, it requires a modified smartphone with special access privileges, and relies solely on visual feedback and computer vision for target detection. Thus, it is not adaptable to dynamic content or accessible to blind users.

2.2.3 Discussion

Providing in-context assistance can facilitate users' learning process and provide them with the support to tackle challenges. Web Q&A in-context solutions, like LemonAid (Chilana, Ko, and Wobbrock 2012), provide users with the ability to ask questions that are made available to anyone. In smartphones, in-app assistance has become more common, with many applications adopting an on-boarding approach guiding the user on their first interactions. They incorporate many of the findings reported in the interactive tutorials. Unfortunately, they are limited in their scope (i.e. supporting only first usage) and rely on visual metaphors to guide the users (e.g. transparent overlay occluding all the irrelevant content), making most inaccessible to blind users. Furthermore, developers are solely responsible for the creation of these assistances, without any feedback from the users.

Providing in-context assistance that adequately responds to the users' needs is a demanding task. Solutions that solely rely on the developer are doomed to lack in coverage and availability without incurring in significant costs. Demonstration based tutorial creation can be leveraged to expand the pool of assistance.

2.3 Human-Powered Access Technology

Human-powered solutions can provide the flexibility needed to support users in a variety of contexts and subjects. For example, in the past, it has been leveraged to provide video relay services to facilitate the communication between deaf and hearing people, or to create a support network for Alzheimer disease caregivers (Bateman et al. 2017). In this section, we describe how this type of solutions has been previously leveraged on the web and on mobile devices to the benefit of people with disabilities.

2.3.1 On the Web

With each passing day, access to web content becomes a commodity and accessibility becomes imperative. However, a lot of the online content remains inaccessible. One common issue blind people face are the CAPTCHAs (i.e. Completely Automated Public Turing Test to Tell Computers and Humans Apart) that many websites require to guarantee a human user is behind the interactions. Unfortunately, many of these CAPTCHAs are based on inaccessible visual questions (Holman et al. 2007). Through a human-powered approach, *Webvisum*[14] provides CAPTCHA image solving. Moreover, it enables users to tag and share page enhancements through its web browser add-on.

When facing a barrier on a website, users can usually contact the developer to address it. However, some developers might not have the time, expertise or be currently in charge of supporting the website (Lazar, Dudley-Sponaugle, and Greenidge 2004). It can take a long period of time before an issue is solved, if at all. Moreover, it relies on the user ability to describe the problem and the developer capability to extract the underlying cause of it. Human-powered approaches provide the perfect features to address this issue.

[14] Webvisum, (http://www.webvisum.com/)

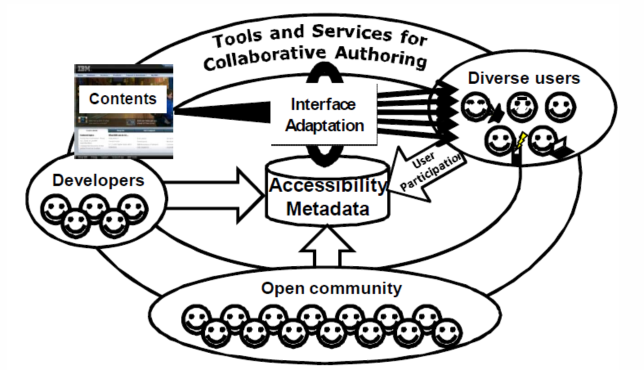


Figure 12 - Concept of Social Accessibility Approach. (Takagi et al. 2008)

The Social Accessibility project (Takagi et al. 2008) gives blind users the ability to report problems to a network of sighted volunteers (Figure 12). Volunteers are notified whenever an issue is reported, and can use the tool to improve the website accessibility for all by adding the requested metadata (e.g. image labelling, document structure). After 20 months of deployment, the project had 350 volunteers contributing, that created about 19,000 metadata over 3000 web pages (Sato et al. 2010), revealing a promising future for human powered approaches reliant on volunteers.

In Martins et al. researchers aimed to provide a prosthetic memory for people with dementia(Martins et al. 2014). In their approach, they relied on a private social network to enrich and validate the data that was collected automatically from the user smartphone or retrieved from the web. Similarly, SocialMirror (Hong et al. 2012) leverages a trusted online social network composed of friends, family and professionals, to allow young adults with autism and caregivers to seek advice. Given the sensibility of the issue, using a social network brought additional challenges. Caregivers were concerned with the safety and privacy of the solution. Furthermore, conflicting advice from multiple caregivers aggravated tensions between the individual and his or her caregivers. When developing human-powered solutions one must consider the validity from contributions.

2.3.2 Mobile

Smartphones brought forward new possibilities for assistive technology. Commercial applications such as *TapTapSee*[15] and *BeMyEyes*[16], reveal the opportunities provided by these devices when coupled with a human-powered platform. Both apps allow users to take photos, *TapTapSee* uses computer vision to identify objects and text, while *BeMyEyes* allows the user to ask a question. In *TapTapSee*, crowd workers are used when

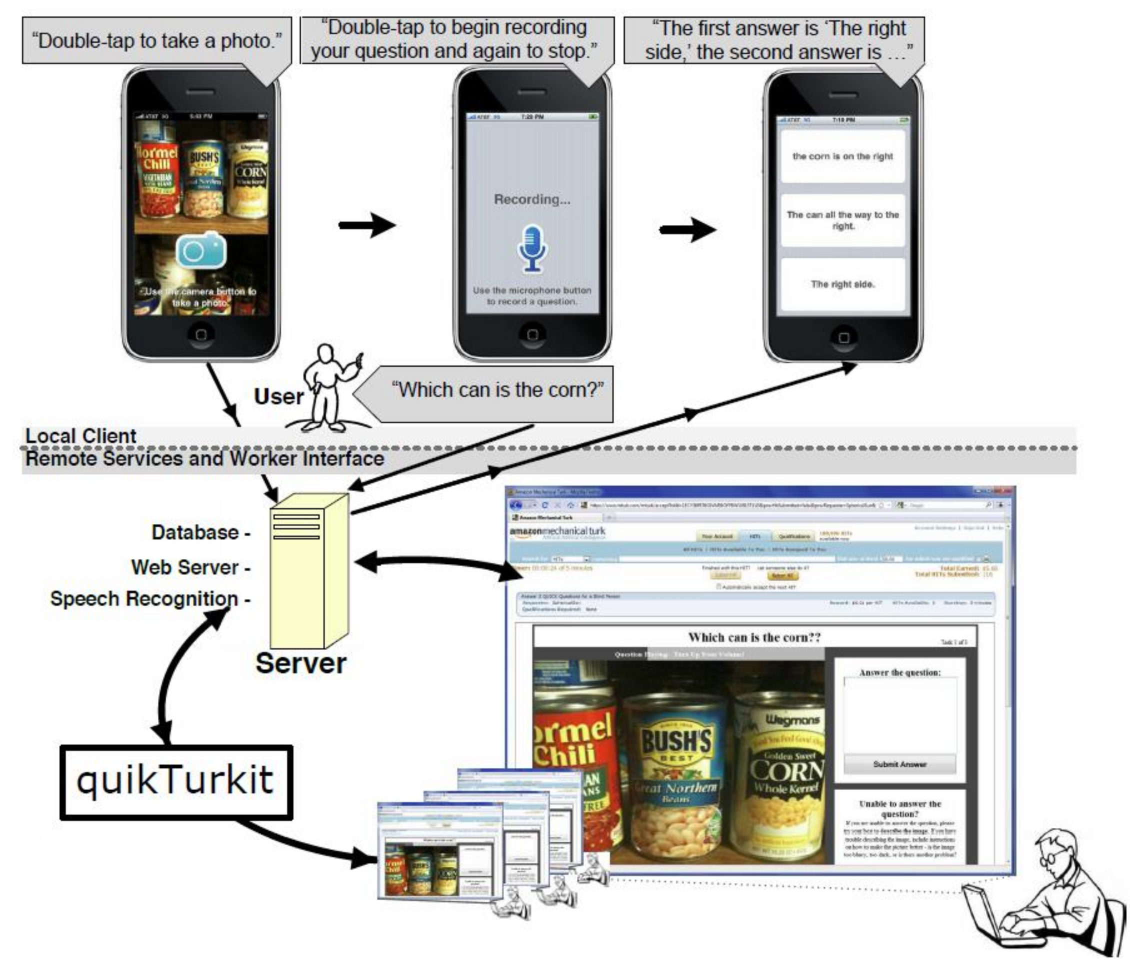


Figure 13 - The VizWiz client is a talking application for the iPhone 3GS that works with the included VoiceOver screenreader. VizWiz proceeds in three steps—taking a picture, speaking a question, and then waiting for answers. System components include a web server that serves the question to web-based workers, a speech recognition service that converts spoken questions to text, and a database that holds questions and answers. quikTurkit is a separate service that adaptively posts jobs to Mechanical Turk in order to maintain specified criteria (for instance, a mininum number of answers per question or a pool of waiting workers of a given size). (Bigham et al. 2010)

---

[15] TapTapSee, (http://taptapseeapp.com/)

[16] BeMyEyes, (http://bemyeyes.com/)

the app cannot automatically recognize the object. With *BeMyEyes*, users are connected to a volunteer to ask questions (e.g. which of these cans is the tomato soup).

Preceding its commercial look-alikes, VizWiz (Bigham et al. 2010) provided blind people with the ability to ask questions to crowd workers, by recording a question and taking a photo (Figure 13). To address the issues these approaches face in timing, VizWiz recruited workers in advance, and asked the same questions to multiple people, thus achieving near-real-time responses. Lasecki et al. reports on the benefit of using multiple crowd workers to provide users with a continuous conversation about the user video stream (Lasecki et al. 2013). Crowd workers were able to answer faster and more accurately to the users' visual questions.

In VizMap (Gleason et al. 2016), volunteers map indoor locations by recording videos. Crowd workers are then tasked with labelling the objects found in each key frame, thus creating a query-able 3D model of the building. In the proposed work, users will then be able to take a photo to locate themselves in the environment and learn about the objects in their surroundings. VizLens (Guo et al. 2016) combines computer vision with human computation to tag interfaces of the real-world. Users take a photo of the desired interface; crowd workers are then able to describe each of the interface elements. After an interface is tagged, VizLens, through computer vision, can recognize the interface and provide users with feedback, similar to a mobile screenreader (i.e. reads what is beneath the finger in the real world while pointing the camera at the interface).

2.3.3 Discussion

Human-powered technology can be leveraged to create powerful assistive technologies. These platforms can be built around crowd workers (Bigham et al. 2010; Gleason et al. 2016; Guo et al. 2016), volunteers (Takagi et al. 2008) or even people we know (Martins et al. 2014). Each of the workforces has advantages and disadvantages in matters of privacy, security and knowledge. These approaches create an always available network of assistance for a specific problem, giving users the autonomy to overcome challenges on their own. As we have seen, there are advantages in the availability, coverage and the accuracy of responses. Some approaches (Takagi et al. 2008; Chilana, Ko, and Wobbrock 2012) build a knowledge base with each question asked. As the knowledge base grows, it reduces the workload on the respondents and improves response times. We aim to leverage the benefits of a human-powered approach to facilitate smartphone usage by blind people.

2.4 Summary

In this chapter we detail the current state of the art regarding mobile accessibility for blind people, in-context assistance, and the use of human-powered technology to support people with disabilities. We identified the knowledge gap in past research, regarding the adoption and everyday challenges blind people face when using smartphones. There appears to be an opportunity to leverage in-context assistance to provide support, where currently there is none. The past work in other domains have shown how it can be used to the advantage of its target users. Lastly, we highlight past work in human-powered access technology, particularly the ones developed for blind people. We believe the combination of these types of solutions can be fruitful to facilitate smartphone use and promote self-efficacy.

Chapter 3
Smartphone Adoption

The advent of system-wide accessibility services on mainstream touch-based smartphones has been a major point of inclusion for blind and visually impaired people. Ever since, researchers aimed to improve the accessibility of specific tasks, such as text-entry and gestural interaction. However, few works aimed to understand and improve the overall accessibility of these devices in real world settings.

In this dissertation, we go beyond the state of the art by leveraging a multiple research method to characterize the challenges faced by blind people. We report, in this chapter, how in early 2015 we sought to understand the adoption process of newcomers, understanding their concerns, barriers, support mechanisms and evolution. We conducted a twelve week in-the-wild longitudinal study with five novice blind users. The study included pre-adoption and weekly interviews, weekly controlled task assessments, and in-the-wild system-wide usage.

Our results show that mastering these devices is an arduous and long task, confirming the users' initial concerns. We report on accessibility barriers experienced throughout the study, which could not be encountered in task-based laboratorial settings. Finally, we discuss the role others play during the adoption process and highlight the need for better support tools.

In this chapter we contribute with **C1) an in-depth assessment of the smartphone adoption process of blind people** and **C2) a data collection framework to support mobile in-the-wild studies.**

3.1 Real-world Use and Adoption

Prior studies have investigated real-world technology adoption and use by people with disabilities. Using interviews and participant diaries, Kane et al. explored the accessibility challenges faced by people with visual and motor impairments when using mobile devices (Kane et al. 2009). Anthony et al. analysed *Youtube* videos of people with physical disabilities interacting with touchscreen devices, which were supported with online surveys on technology use (Anthony, Kim, and Findlater 2013). This approach captured the unique and interesting ways in which people have augmented or crafted solutions to support their interactions with touchscreen devices. Furthermore, the study highlights real-world scenarios and interaction contexts from individuals with motor impairments. Similarly, Naftali et al. (Naftali and Findlater 2014) conducted in-the-wild case studies with four people with motor impairments to explore the impact of environmental context on their mobile interactions, using a combination of interviews, participant diaries, and contextual session observations.

Commonly, in-the-wild user studies do not allow the researchers to obtain objective performance measurements of device interactions, as previously seen within laboratory evaluations. One exception is the work by Montague et al. who conducted a four-week in-the-wild user study involving participants with motor-impairments, using a custom-built game to capture touchscreen interaction performance measurements (Montague, Nicolau, and Hanson 2014). However, this approach was limited to collecting interaction data within the custom-built game - overlooking the interactions that participants were making with other device applications.

In the current study, we go beyond the state of the art on understanding technology use and adoption. Using a mixed-method approach, we relied on quantitative and qualitative data. We collected qualitative weekly data, quantitative weekly controlled assessments, and in-situ device interaction and usage data. This information allowed us to understand how users learn and evolve.

3.2 Supporting Mobile In-The-Wild Studies

To support in-the-wild studies we developed TinyBlackBox (TBB), an open-source system-wide data collection solution for Android mobile devices. TinyBlackBox, is a standalone Accessibility Service, built for the Android 4.0+ OS. Once installed and activated, TBB will continuously run in the background of the OS, capturing the user’s device interactions system-wide.

TBB scrapes application data, including page layouts and interface elements – these are represented in a DOM tree structure, revealing information about the nesting of interface elements. TBB also records all of the interface interactions e.g. clicks and swipes made within applications. In addition to recognizing interface clicks, TBB provides overwritten touchscreen drivers. This enables TBB to receive the sub-gesture touch *begin*, *move* and *end* interactions, as typically recorded for touch modelling, and gesture analysis (Montague, Nicolau, and Hanson 2014).

We ensure the security of user data by encrypting the log files locally on the device before they are transmitted using HTTPS protocols. TBB supports parameterized levels of data encryption i.e. *encrypt all text*, *only user generated content* such as personal SMS and emails or *encrypt nothing*. We believe it is vital that the user is not only aware of what data is being recorded, but also entirely in control. In future versions, we plan to include application specific encryption rules.

We have integrated TBB with Google Cloud storage to aggregate log data from multiple participants while the study was live. TBB will attempt to synchronize with the cloud storage when the device has an active WIFI connection, at least 40% battery remaining, and the device is inactive or charging. Prior to uploading log files, they are compressed for network performance and to minimize cloud storage costs.

In addition to transmitting log files, TBB periodically pings the cloud storage servers with a status report. We use this to verify that TBB is functioning correctly and that the participants are using the devices regularly, reducing the need to conduct field assessments of the devices and software.

While TBB is a standalone data collection service, we recognize that there is often the need for study specific information to be gathered. To support this need, we have also developed a small library that can be embedded into third-party applications and services, providing APIs to facilitate communication with the standalone TBB service - enabling additional data to be captured alongside the existing TBB data. Using this approach, we created an external logger, BlackBoxBuddy, to capture contextual information such as the user's location (from Google location services, which leverages GPS, WiFi and Cell towers) when interacting with the device.

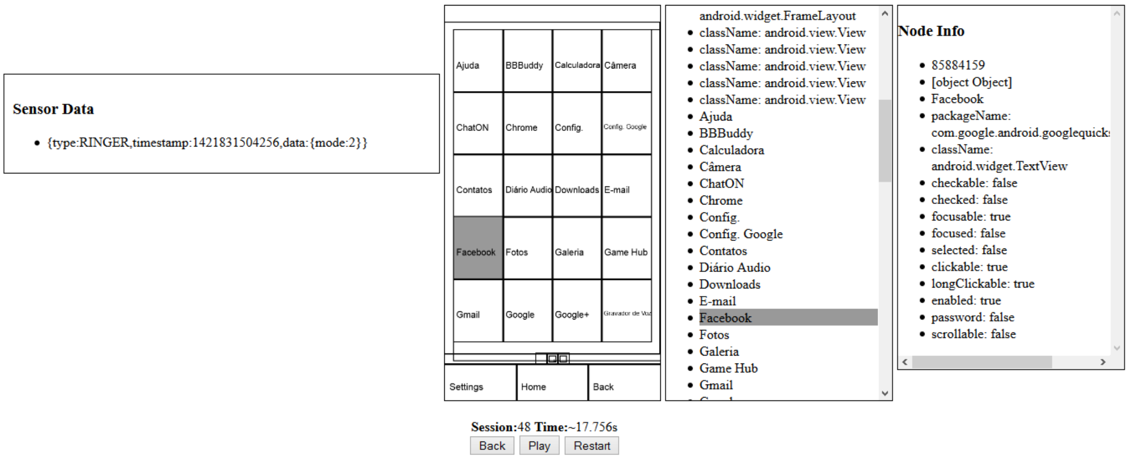


Figure 14 - Tiny Black Box interaction player. App drawer with interface element details.

Finally, we have developed a suite of analysis tools to support data extraction, visualization and playback. Since TBB captures the device interactions into individual text files, we have created scripts to rebuild the logs into a local MySQL database. In this form, it is possible to extract high-level statistics, such as application usage behaviours. To assist the deeper inspection of individual device sessions, we created TBB Player, an interactive web-based playback tool (Figure 14). TBB Player renders a wireframe of the screen content that was visible to the user and replays their touchscreen interactions in real-time. Researchers can pause and rewind interactions for further inspection. This tool is particularly useful for investigating gesture interactions that are context-specific, as each step of the device interaction is captured.

We have successfully conducted a four-month user study with novice blind users, exploring their adoption experiences (Rodrigues et al. 2015) and text-entry performances (Nicolau et al. 2017; 2015b) using TBB, which we describe in the next section. TBB supported a deeper comprehension of the interactions and habits faced by the users in their daily usage, and a detailed analysis of their finer-grained interactions with the onscreen keyboard and text-entry behaviours. Moreover, it enabled us to ask the relevant questions during our weekly interviews that were prompted by the data. TBB is open source and available[17].

[17] TinyBlackBox Repository (http://goo.gl/VQTrJt)

3.3 User Study: Adoption of Smartphones by Newcomers

In this study, we focused on uncovering and understanding the smartphone adoption by blind people. We identified the concerns, expectations, barriers, text-entry behaviours, support mechanisms and external influences participants had. In this dissertation, we describe in-depth the adoption process of blind users with a focus beyond text-entry, which is outside the scope of this thesis. Thus, for a full analysis of the typing performance in laboratory and in-the-wild refer to (Nicolau et al. 2015b; 2017).

Table 1 - Participant profile, where YB (Years of Blindness) and UD (Use per Day)

ID	Age	Sex	YB	Old device	UD	Features
P1	55	M	52	Nokia C5	5-7	Calls, SMS
P2	34	F	11	Nokia E52/ Nokia 3230	>10	Alarm, Calls, SMS
P3	51	M	25	Nokia N70	2-4	Alarm, Calendar, Calls, SMS
P4	23	M	9	Nokia E66	>10	Alarm, Calendar, Calls, SMS
P5	23	M	5	Nokia C5 / Nokia E65	>10	Calls, SMS

3.3.1 Research Questions

We aimed to answer the following research questions:

- What are the main concerns and expectations of novice smartphone blind users?
- How well do current systems support the novice user in a first contact?
- What are the barriers to smartphone adoption by novice blind people?
- What role did external influences play in the adoption process?

3.3.2 Participants

Five participants with visual impairments, four males and one female, took part in our user study. Participants' ages ranged from 23 to 55 (M=37.2, SD=15.2) years old. They were recruited from a local social institution, and all participants were legally blind as defined within our IRB approved recruitment. None of the participants owned or used smartphone devices; however, they were all experienced desktop screen reader users. Table 1 provides further information about participants, including their technology usage and experience.

3.3.3 Apparatus

Participants were each provided with Samsung S3 mini touchscreen smartphones running the Android 4.1 operating system. We enabled the Talkback screen reader and pre-installed our data collection service, TinyBlackBox (TBB).

3.3.4 Procedure

We conducted an eight-week user study, with briefing and debriefing sessions. During the longitudinal study, we collected in-the-wild data and had weekly sessions and assessments. We extended the in-the-wild data collection period another four weeks for the text-entry study reported in (Nicolau et al. 2017).

Below we describe in detail each of the study phases: 1) pre-adoption interviews and background data, 2) introduction session with Talkback tutorial, 3) in-the-wild device usage, 4) weekly sessions with researchers, 5) post-study interview.

Pre-adoption Interview and Background Data

We met with the participants at a local social institution for blind people. All sessions were audio recorded to a maintain fluid conversation flow, and to allow for data analysis afterwards. Participants first completed a background questionnaire, provided details of their existing mobile phone, device usage, and prior experience with touchscreen interfaces. All participants owned and used feature phones, as shown in Table 1. Moreover, none had previously used a touchscreen smartphone.

During the pre-adoption interview, participants were asked to discuss in more detail their current mobile device, particularly what they liked or disliked about it. Then they were asked how they felt about smartphones and touchscreens, what their expectations were, i.e. things they would like to do with the device, how long they thought it would take to learn how to use it, and challenges or concerns with using smartphones. Finally, they were asked about their existing support network, i.e. friends or family that could assist them in learning how to use the smartphone.

To capture a quantitative baseline of mobile device usage, we collected data from the participants performing a set of basic tasks on their former phones: check the time; add a contact; call a contact; call a number; answer a call; read a new SMS; reply to an SMS (no text); create a new SMS conversation; open target app, and complete text-entry trials.

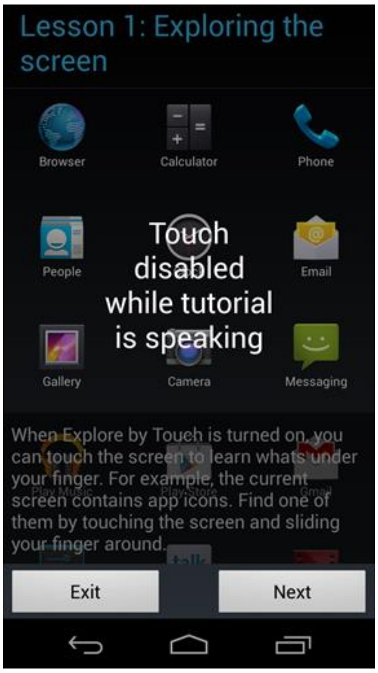


Figure 15 - Talkback Tutorial first lesson.

Talkback Tutorial and Basic Tasks

Participants were first introduced to the form factor of the smartphone, including its features - i.e. how to turn it on and off, the volume controls and touchscreen. We then enabled the Talkback accessibility service and started the Talkback tutorial (Figure 15). Participants were then given the device and asked to follow the tutorial instructions. At the time, Talkback tutorial was composed of four lessons: 1) Explore the Screen; 2) Scrolling through Lists; 3) Context Menus and Reading Granularities. Each lesson introduced at least one new gesture. Whenever participants had doubts, they could prompt the researcher to explain verbally how to accomplish the task.

Participants were then guided by the researcher through all **basic phone tasks**: 1) check time; 2) add a contact; 3) call a contact; 4) call a number; 5) answer a call; 6) check received text message; 7) send message in existing conversation; 8) Send a new message creating a new conversation; 9) go to <App>. During the tasks we asked participants to think aloud and talk about their experiences while learning to use Talkback and the device. The sessions were video recorded and the TBB framework captured all interactions.

In-the-Wild Device Usage

To capture the participants' adoption experiences, we asked each participant to replace their existing feature phone and use the provided smartphone as their primary device. We assisted the participants with installing their carrier SIM cards and transferring their device contacts onto the new device. Participants were informed of the TBB data collection service and provided examples of the data that would be collected. The in-the-wild usage of the adoption phase was scheduled to run for three months. While it was not expected to cover the participants' full learning experience, i.e. from novice to expert, it allowed enough time to trace the adoption process and conduct additional controlled assessments of key behaviours through weekly sessions.

Weekly Session

We met each participant once a week for 30 to 60 minutes to take performance measurements of the nine basic tasks, 20 minutes for text-entry trials and conduct ten-minute interviews.

First, participants were asked to perform the same nine basic phone tasks described in section *Talkback Tutorial and Basic Tasks*. During these tasks, the researcher was observing and taking notes.

After a short break, participants performed text-entry trials. They finished once they written 10 phrases or spent 20 minutes writing. Each trial contained one sentence comprised of five words, with an average size of 5 characters, and a minimum correlation with language of 0.97. We developed an experimental application that would select the trial sentences from a written language corpus. The application randomly selected the sentences for the session to avoid order effects and captured transcribed sentences and completion times.

During the weekly interviews, we prompted participants to discuss their experiences throughout the week, describing any challenges or concerns they may have had and informing us of new installed applications or activities they tried that week with the device. We relied on the data collected from TBB to inform the interview, and had questions related with the tracked events (e.g. "*You haven't used the device this past few days, did anything happen?*"). Before finishing every session, we gave participants the opportunity to request assistance as they would with family or friends. We would only provide verbal responses to avoid training users on device interactions and usage.

Post-Study Interview

After the eight-week period of weekly sessions, we conducted the post-study interview with the same procedure as the pre-adoption. Interviews lasted from 15 to 30 minutes. We asked questions related with their answers in the pre-adoption interview, such as *did their concerns and expectations come to fruition? Did their personal opinions change? Would they make the transition again? What would they do differently? Would they recommend the device to others? What advice would they give to novice users?* Finally, *what are the challenges looking forward?*

3.3.5 Data and Analysis

A total of 50 interviews were conducted: 5 pre-adoption, 40 weekly and 5 post-study interviews. We followed an iterative coding process, where two researchers independently created codebooks. Each researcher coded the same two adoption interviews and two weekly interviews, after which the codebooks were refined and merged. Using the merged codebook, the researchers proceeded to analyse five weekly and two adoption interviews, which lead to further refinement of the codebook and finally a Cohen's kappa agreement of k=.85 (SD=.08) was achieved.

The results from the qualitative analysis of the interviews were complemented with observations, log data from the weekly tasks and log data collected in-the-wild. We gathered a total of 7175 in-the-wild sessions from the five participants. A session starts with an activation of the screen and ends when the device goes to standby.

Dependent Measures

Text-entry performance was measured by analysing trials' input stream (Wobbrock and Myers 2006). We report on words per minute (WPM) and total error rates.

3.3.6 Findings

In this section, we present our findings structured by the phases explained in the procedure: Pre-Adoption, First Contact, Weekly Evolution, and Post-Factum. Weekly

Evolution comprises the data collected from the weekly session tasks, interview, and in-the-wild data.

Pre-Adoption: Concerns and Desires

More than a phone

Feature phones are slowly disappearing from the marketplace, forcing blind users to, sooner or later, make the transition to smartphones. These devices are now mainstream, and having one is seen as a gateway to social inclusion.

"I have nothing against my current phones, but I would be more included in society and in today's technology". (Participant 5 – P5)

One of the many benefits of smartphones is their ability to provide richer communication channels (e.g. email, *WhatsApp*, *Facebook*). Smartphones can also be the solution to some of the current inaccessibility problems users face every day with standard out of the box technology.

"[With a smartphone] I can have access to the Internet, email... and to be able to listen to music. I like my MP3 player, but it is extremely hard to control I have to keep asking for help" (P2)

Concerns

While participants were usually aware of the benefits of owning a smartphone, they considered them too expensive and viewed them as a luxurious item. Participants associated it with the use of Internet, and therefore attributed its cost to the retail price plus the hidden costs of internet data. Participants believed using such "luxurious" devices would have security risks associated, particularly in public transportation.

None of the participants was aware of how touchscreens were accessible, nor how they were able to interact with it. The lack of knowledge triggered several sentiments of self-doubt. Participants were afraid of performing actions unwillingly, or that using a touchscreen meant having to rely on their spatial ability.

"I will need to target where I know something is rather than feeling where it is" (P5)

“I mean, I cannot see the keys, so how am I going to be able to select the letters? How will I be able to tell where they are?” (P2)

The lack of tactile feedback on touchscreens led blind users to feel smartphones were not for them. They believe smartphones were adapted to fit their needs after development and, therefore, would inherently have accessibility problems.

Participants reported conflicting views of the device, between and even among themselves. They expected to have problems and less control, while also expecting to be able to do more. They were simply unaware of what to expect:

“I don't know how accessible touchscreen applications are” (P2)

Expectations

Doubts on the interaction method and thoughts on the inaccessibility of the device resulted in participants expecting a difficult transition that required the assistance of others. However, expectations on adoption time highly depended on individual needs. While one user expected to take months to feel comfortable with the device, others believed it would take only a few days or weeks. One participant was committed to the change and needed to make the transition in a day:

“I have to learn it the day I get it. People will keep messaging and calling me. If I don't answer, people will start to think I am dead [smiling]. I have to at least answer my mother!” (P2)

Most smartphones come out of the box with a screen reader available. Thus, participants had higher expectations for it, as opposed to their feature phones where the screen reader was an external software that had to be installed. They expected applications to be inherently more accessible, again, contrasting with their belief that smartphones would not be as accessible as their old feature phones.

The Daunting First Contact

Talkback tutorial was problematic for all users; only one participant was able to successfully complete it. During the tutorial, the touch interaction was locked while the lesson was being explained, or when participants took too long to perform the intended action. Users were not given the freedom to explore, nor to complete the task at their own pace. Often users would get locked out and had to listen to the lesson description again, it was infuriating. Users felt frustrated since they weren't in control of the flow of the lesson and their learning experience.

The tutorial was unable to distinguish successful from unsuccessful tasks. Participants did not get direct feedback of the gesture they had just performed, just the resulting action. For example, when the tutorial asked for users to swipe to the next item to learn about sequential navigation, users would try to swipe; independently of whether the gesture performed was recognized as a tap or a swipe, an element would be focused on the screen, and the feedback provided was about it. The tutorial assumed the user had performed the correct gesture and move on to the next step of the lesson. As a result, none of the participants understood how to swipe to go to the next element since all their first attempts at a swipe were not congruent with what touch recognizers expect of swipes.

Moreover, most of the tasks did not have an intended target. Lessons would ask the user to, for example, navigate a list but would not provide a target for the interaction, nor provide any feedback on how the list was affected. As an example, the *List Lesson* asked users to select an option. Users would tap anywhere on the screen, and go to the next step of the lesson, having no idea what they just did other than that they tapped the screen.

Gestures

The last two lessons from the *Talkback* tutorial required users to perform L based gestures to open the local/global context menus. Four out of five participants were unable to consistently open the menus even after 15 attempts. They struggled to perform the gesture fast enough, in the correct area, with the correct shape, and, even when they successfully did so, they struggled with context menus. When they opened these menus, the interaction method changed. They now had to focus an item by dragging the finger around a circular area, and lift their finger to select the intended option. The participants were confused as to why the interaction method suddenly changed.

Figure 16 - Samsung S3 mini with two capacitive buttons with no feedback

Physical Cues

Touchscreens lack physical cues. As such, participants attempted to use whatever they could to facilitate screen exploration. P4 *noticed that the shortcuts to Phone, Messages and App Drawer were located in the bottom of the home screen, above the physical home button of the device*. Consistently throughout the weekly sessions, whenever he had to select one of these apps, he would first locate the home button and go from there.

Participants struggled with the capacitive buttons (i.e. back and settings), as both were located on the bottom of the device and had no physical cues, making them indistinguishable from the touchscreen (Figure 16); moreover, they provide no audio feedback and activate on touch. Participants ended up inadvertently pressing the back button when using the edge of the device as a physical cue.

Interaction Method

Participants were introduced to two different exploration methods and three different selection methods in under two hours.

Selection methods include 1) double tap to select any focused option; 2) on lift to input text and to select options in the context menu; 3) on press/touch on the physical home button and the two capacitive buttons. During the session, participants were mixing some of these methods: double tapping to input text or trying to select options by lifting. Some of them asked why there would be different selection methods between functionalities.

Exploration methods include: 1) *Explore by Touch*, where users can tap or drag their finger to focus the different option underneath; and 2) Sequential navigation, by swiping left and right to go to the previous/next element. After the *Talkback* tutorial, participants relied solely in dragging their fingers, rather than swipe gestures. Only when participants felt the need to scan all options on the screen, and asked how they could achieve it, did the researcher explain once again both exploration methods. This illustrates how ineffective the tutorial was at explaining the different interaction possibilities.

Text-Entry

Even though text-entry was one concern exposed in the pre-adoption interview, participants were able to input text after a brief explanation of the interaction method of virtual keyboards.

Learning and Evolution

Guided by the qualitative analysis of the weekly interviews, and the data collected in-the-wild and from the text-entry weekly assessments, we present the following themes: Barriers, Text-Entry, Concerns & Insecurities, Influences, Positive Experiences and Attitude. Each theme provides an insight into the adoption process supported by observations, data from weekly tasks, and in-the-wild device logs.

Barriers

During the weekly sessions, users continuously reported barriers which only changed in form over time. Participants were constantly evolving, tackling new challenges as they overcame others.

Gestures. Difficulties with performing gestures directly affected participants' ability to operate the device, especially in the first week. They struggled even with the simpler gestures (e.g. double tapping). The underlying causes varied from accidental touches, timing issues, and unrecognized taps.

Some of the basic features had unnecessary complex gestures. For example, unlocking was composed of a double tap followed by a two-finger slide. Participants struggled with these gestures. During the first three weeks, we observed participants performing several attempts to unlock their devices. In the most severe case, P1 was *unable to consistently* *unlock in the first 3 attempts* throughout the study, from week one to week eight. As the weeks went by, the basic gesture problems were observed less frequently during the weekly sessions and were far less mentioned during the interviews.

Table 2 - Weekly task table per participant where it's represented the first week from which participants started to consistently be able to perform the tasks.

Task	P1	P2	P3	P4	P5
Check time	W1	W1	W1	W1	W1
Add Contact	W6	W4	-	W6	-
Call Contact	W2	W4	W3	W4	W8
Call number	W2	W3	W1	W1	W1
Answer Call	W1	W1	W1	W1	W1
Receive Text Message	W6	W1	W1	W1	W5
Send Message	W1	W1	W5	W1	W3
Send new Message	W7	W5	W6	W3	W7
Go to <App>	W1	W1	W2	W3	W3

Physical Cues. The capacitive buttons continued to be a problem during the first three weeks. Participants would inadvertently press them, get lost in the interaction, and struggle to recover the previous state.

Interaction Method. Even with this study's small sample, we saw participants opting for different exploration methods, and maintaining the same style throughout the study; as soon as they found something that worked.

Those that initially struggled with gestures used almost exclusively *explore by touch*, only relying on swiping to the next option when they could not reach the intended target in any other way. Only one relied solely on sequential navigation.

Selection methods continued to confound users: during the first couple of weeks participants would double tap the capacitive buttons as if they were focused interface elements, which lead to the same consequences as inadvertently pressing them.

Scrolling. To scroll lists without relying on swipe gestures, users had to perform a two-finger slide inside the list bounds to drag it. Participants faced several challenges: 1) where to position their fingers, and 2) understand how finger movement affected the scrolling. For instance, some participants inadvertently opened the status bar on the top of the screen while trying to scroll through the contacts. As a result, tasks involving the contact list (e.g. call a contact and send new text message) were the most problematic (Table 2). Participants tended to rely on the dial phone in the first weeks while learning to manipulate lists.

"I never know how much I move when I slide, I don't know if I skipped a few options or not" (P1)

Awareness. Smartphone applications are quite different from the ones in feature phones. Many of the initial barriers can be attributed to the mismatch between participants' expectations (i.e. mental model) and the actual interfaces. As an example, *P1 was expecting to have the different folders of Inbox/Sent/Drafts in the SMS app rather than a conversation paradigm where all messages are grouped by contact.*

Applications keep the previous state depending on how they were closed and how they were implemented to resume when reopened. These inconsistencies when navigating the same options made it difficult for participants to build a mental model of the application, and all its dynamic components.

"One thing I noticed is every time I try something, I always find things different, and since I can't find regularities I cannot learn [how to do it]" (P1)

Smartphone interactions are accompanied by specific audio feedback. During the initial weeks, users had no understanding of what these audio cues meant, and therefore struggled to understand interface states. P5, *in forms, kept pressing the edit box over and over again to get to the keyboard, and it took him a while to realize the sound he was hearing meant the keyboard was already opened.*

Accessibility Compliance. After the first four weeks, barriers shifted from interaction challenges to feature- or application-driven. Once participants delved into new applications, they started to face traditional accessibility issues, such as finding buttons with no description, forcing them to create coping strategies.

> *"I found the send button is always changing number, I have to remember that it is the button next to the text box" (P2)*

Assistance Required. Although relying on others is probably one of the most effective coping mechanisms, help is not always available. In a similar situation to the one described above, the same participant had no assistance at hand and stated:

> *"I still can't use Endomondo, because instead of start and end, the buttons are numbers and I don't know which is which" (P2)*

Some barriers are difficult to surpass without assistance, which can have dramatic consequences. P3 and P4 mistakenly changed the language of the device and, from that point onwards, were unable to recover. Even when others are available, they might not be able to help; particularly sighted people who are unfamiliar with screen readers. In P3's case, his family couldn't solve the issue; he was not able to use the device for the remainder of the week.

Users and helpers can also collaborate to tackle the issues. P1 was struggling inserting the password to access his home network and his daughter was unable to navigate the device with the screen reader on:

> *"I went into the Wi-Fi configuration screen just like I learned, and then I asked my daughter to select the network and insert the password". (P1)*

Feedback. Audio feedback sometimes does not respond as users would expect, and therefore breaks their trust in its accuracy. The screen reader had speech delays, due to buffering effects. In some instances, users only received audio cues of their interactions, but no immediate feedback on what was focused:

> *"We are searching for the buttons and we just hear poc, poc, poc, but it doesn't say anything" (P3)*

This mistrust in the system led users to blaming it, whenever something did not react as they expected. However, in many cases, the issue was the gesture had been incorrectly performed, or they had a misinterpretation of how the system would behave. The delay between interaction and feedback made users take adaptive measures. P3 and P4 sped up the *Talkback* voice; P2 *started waiting for feedback before doing sequential interactions*.

Coping. When confronted with a barrier, users tended to have one of four approaches. First, they would 1) ask for help, if help was available. If it was an interface or application problem they had not faced before, on that same app, they would try to solve it by 2) rebooting the device. When possible, they would 3) search for a solution using other devices, or simply 4) perform the task without relying on the smartphone.

Text-entry

According to laboratory results, participants achieved an average typing speed of 4 WPM and 4.7% total error rate after eight weeks of usage. Although performance keeps improving after eight weeks, learning rate is slow (0.3 WPM per week). Previous research has shown similar results (Azenkot, Wobbrock, et al. 2012). Regarding real-world performance, input speed is on average 1.5 times faster.

Edit text. Corrections are still time consuming and inefficient. None of our participants used cursor-positioning operations throughout the study. It seems that these actions are only expected to be used by expert typists, preventing novice users to do fine-grained corrections.

Feedback and touch input. Findings suggest that many errors can be due to a mismatch between speech output and touch information. Participants reported how the feedback to their actions while inputting text was not immediate, and thus caused them to either be slower than they needed to be, or to make additional errors.

Concerns & Insecurities

This theme aggregates all the reports of the users' concerns when using the device, including their reported insecurities and privacy issues.

During the first three weeks, users had difficulties controlling the screen reader and felt uncomfortable while doing so. Users stressed how insecure they felt in public transportation. This caused users to avoid using the device in public places, some even after the eight-week period.

"What I like the least is the fact that I feel like a complete idiot looking for the options and then not being able to use them" (P2)

Users had reservations regarding sharing information with new apps. They were afraid that by sharing their location, or having their calendar synchronize, somehow others would be able to tell their location/schedule without their permission.

As newcomers to the smartphone paradigm, they were not familiar with registering in apps through their Google or Facebook accounts: when they opened applications for the first time that warranted for a registration, the participants were weary, and did not proceed before confirming with a friend that they had the correct app installed.

Influences

This theme contains how users were influenced by outside sources and how they influenced others. Other smartphone users can directly affect the way users interacted with their device by sharing their experiences, giving tips, and suggesting applications. Most will undoubtedly help, but when users are misguided it can result in a negative learning experience. P3 was told about a gesture he could do from an iPhone user. During the following weekly session, P3 tried to do it with no success due to it being specific to iPhone.

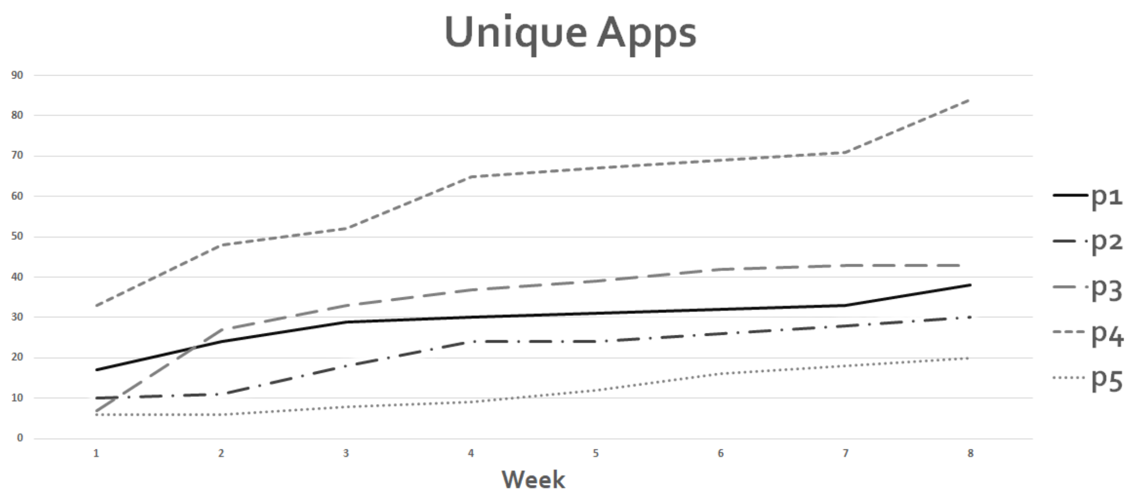


Figure 17- Cumulative number of unique applications visit during the eight weeks by each participant.

Sharing experiences can also be what users need to bootstrap their device usage. P4 was by far the most adventurous explorer: in the first week, he used 33 unique applications (Figure 17). During the first weekend with the device, he explored it extensively with a more knowledgeable friend. Knowing other smartphone users can greatly promote the discoverability of new applications. Other participants were impacting how others surrounding them perceived smartphones.

"My mother never liked new technologies, she got used to phones really slowly. She found my device funny and enjoyed playing with it. She is even considering buying one for herself." (P2)

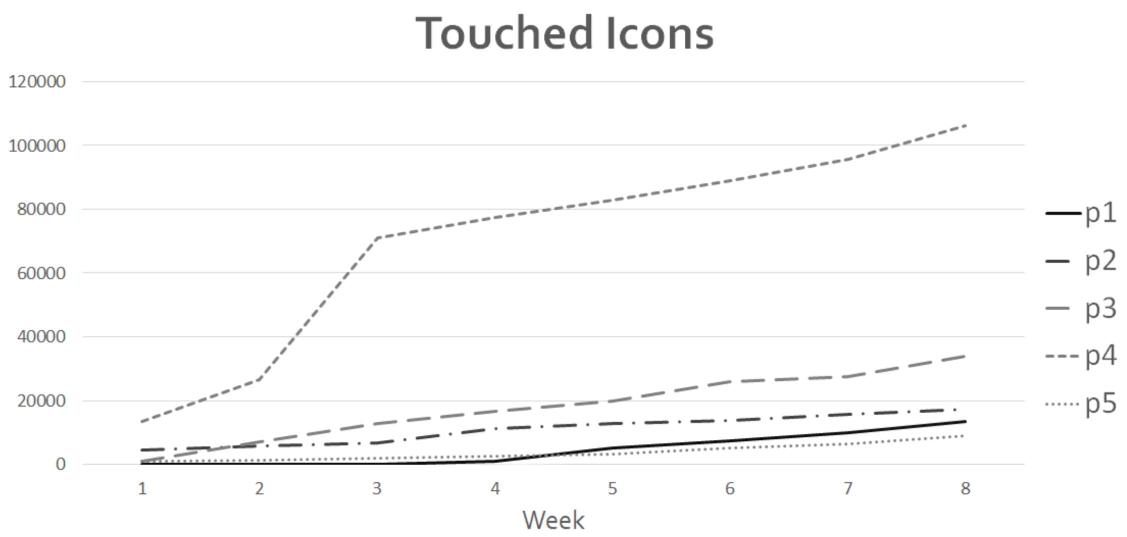


Figure 18 - Cumulative interaction time during the eight weeks by each participant in seconds

Positive Experiences

In this theme, we gathered reports of users' success stories. Through them, we can have a glimpse into their evolution. Users found positive experiences wherever they could, and accordingly to their proficiency with the device. Each had very different experiences, as one can assess looking at the weekly tasks, in Table 2. Users who struggled the most, reported positive experiences associated with basic tasks or exploration. P5 was one of the users that used the device the least, and struggled the most (Figure 18). By the eighth week, the user's positive experiences were only related with being able to perform basic tasks.

As weeks went by, positive experiences shifted to more complex, app related tasks. P1 complemented his PC use with the device, and now checks for emails on the phone before reading them on the PC. To P4, the smartphone was becoming much more than a phone. He kept installing TV apps, games and educational applications; he was particularly happy with his social apps. P3 found great joy in utility apps such as *CamFind* (i.e. users take pictures to identify objects).

“CamFind is very useful around the house. With 2 blind people is great. I like the device more with each passing day” (P3)

Users found joy when they accomplished something without any external help, even if it was supposedly simple, such as setting the phone to vibrate.

Attitude

This theme gathers the attitudes of the participants about the device’s usefulness, and participants perceived self-efficacy during the study.

When successful experiences were sparse, they started to report frustration and doubt of their self-efficacy. Consequently, users did not feel in control of the device, and avoided using it in public spaces. In one extreme case, P1 went as far as warning his family about his inability to control the device to avoid misunderstandings if they were unable to reach him.

“Pretend I don't have a phone” (P1)

Although all users were eventually able to complete all tasks at least once in the eight weeks, they believed their performance would never be the same as with their feature phones, due to it being inherently harder. During the study, users set self-goals, and with each passing week users tried to address their previous barriers, sometimes coming back with success stories.

"It’s hard to handle the post navigation in Facebook” (P1 Week 4)

"I can now navigate between the posts, but I still can't read the full text. I still have to learn" (P1 Week 5)

Post-factum: Perceived Challenges and Benefits

In the post-interview, we were interested in understanding how they perceived their adoption and how they would improve it. Did their concerns come to fruition? Were their expectations met? What was important and what would they advise others to do? What are the next goals?

Concerns

The participants that expressed concerns on using the device in public, continued to feel this way. Two participants (P1, P3) felt that this device was worse as a phone, it was more difficult to use, and slower. Not all users shared this concern. P2 and P4 saw smartphones as being better, even for the most common tasks:

"In my old one, the only thing I'm faster is searching for a contact, even writing a text I am faster here". (P4)

P5 was the user who struggled the most, and in his pre-adoption interview thought he was learning a device not designed for him. The experience with the device changed his view on smartphones:

"My opinion has changed. They aren't that hard, as everything else we just need to learn to use them" (P5)

Expectations

Most underestimated how much time it would take to get used to the device, and how much of a challenge it would be. Some felt that they were used to the device after two weeks (P4), while others stated that they still were not completely comfortable (P3). P2 has met her initial expectations and was happy she was able to send messages in the first week. User's ability with one of the basic tasks did not translate quickly into the others. In P2's case, her prowess with the messages did result in a quick adoption of all features (Table 2).

Looking back

Looking back, users believed dedication to be one of the keys to fast adoption. All except one participant believed they could have benefitted from a gradual transition, like using a secondary phone whenever they were struggling or pressed for time. One believed it would have slowed down his adoption.

"I know that if I was forced to use just one I would have done things quicker" (P5)

During the first weeks, it was crucial to have someone whom to rely on, not only for assistance, but also as an assurance if something went wrong. Participants wished they had more people with which they could have shared their experiences. They believed they could have benefited from collaborative exploration with others, and from knowledgeable blind smartphone users.

One of the reasons participants had to rely on external help was the lack of support material they could rely on. One common request was the creation of a manual with a step-by-step guide of the basic tasks. Participants felt that the device was still not accessible to all blind users. While they would recommend the device, they wouldn't recommend it to everyone.

To participants, their device became more than a communication device. The new applications and features became part of their daily life, playing a role in supporting their independence.

Two users believed they made a trade-off between more features and easiness of use. They don't believe they will ever be as fast as with their old feature phones when performing the same task. But still, they clearly see the advantages.

"It allows you so much more than a regular phone. We can use it almost as well as our old ones. I believe it's a trade off in favour of the smartphone" (P1)

3.3.7 Discussion

Based on our findings we are now able to answer our research questions.

People fear what they do not know and expect the worse.

Pre-smartphone adoption blind users had two major concerns: safety and fear of the unknown. They perceived having a smartphone as an additional security risk, due to the unwarranted attention it brings. This insecurity was deeply felt in the initial weeks, where even users that used their previous phone on a public setting, did not anymore.

Since they had yet to interact with the device, they feared they would be unable to use it, due to the lack of physical cues and the assumption that they would have to rely on their spatial abilities. Their fears were exacerbated by thinking smartphones were inherently not designed for them, and as such, they would have to adapt to an inaccessible device - an opinion they no longer share.

Most expected a faster and easier adoption, they struggled more than they anticipated. From the starting point most believed they would never perform better than with their current phone, but felt they had to adapt mainly due to the market pressure.

Current *getting started* mechanisms are not effective.

The Talkback tutorial was not enough for users to even get started. They struggled with some of the gestures presented and, even after eight weeks, we still saw no reports or data where they used complex gestures.

During the tutorial, they were even misguided into thinking they had performed the correct gesture, when in fact they didn’t. The swiping gestures to navigate sequentially were not learned during the tutorial, which caused users to struggle during it, and in the following weeks with the exploring of applications.

Using just *explore by touch* is a demanding task that heavily relies on the user’s memory and spatial ability. It is crucial to improve the tutorial session in order to reinforce how, when, and which gesture is more appropriated to the task.

The tutorial presents users with the basic gestures but makes no effort in providing feedback on how they affect the underlying interface. We complemented the tutorial with a guided session through the basic tasks. Although users were able to perform all the tasks, most felt they learned too much in a short span of time.

From just basic tasks, users had to learn three different selection methods (lift, press and double tap) and learn where to apply each. Moreover, they had a multitude of new interfaces to learn. As such, when looking back, many users had the desire for a manual that they could rely on. There is a need to provide better learning mechanism for blind users on mobile devices.

The struggles of a new paradigm, the mental model mismatch.

Throughout the weeks, we saw how the different barriers evolved. In the initial weeks, users mainly faced gesture related issues. The biggest hurdle was understanding the new paradigm they were in. Suddenly, users couldn't find any consistency in the applications. They no longer have one single path between point A and B, but any number of ways to do the same thing. Applications now have states: depending on whether you press back or home they close or go to the background and resume in different ways.

This, mixed in with every application behaving differently to the *back* button, led to users struggling to learn new interfaces and to repeat tasks. We need better methods to convey this new paradigm to users.

In the early days of graphical user interfaces (GUI), we faced similar issues when we transitioned out of text-based command line interfaces. With the introduction and eventually spread of GUI, blind people were forced to adapt to this new paradigm. Weber et al. (Weber et al. 1994) recognizing the issue, at the time, proposed the use of tactile representations of GUIs, followed by the use of an app that would function as an interactive tutorial, guiding the user through most common scenarios and interface elements on the Windows operating system. Smartphone getting started tutorials seem to mimic the approach of training users on the underlying interface, but ineffectively.

The assistance of others plays a crucial role

Users felt that without someone whom they could rely on to help, they would possibly have had an insurmountable barrier. During the first weeks, users relied on their friends and family to help them with configurations, app installations or simply understanding the basic interfaces. The primary coping strategy was asking others for help.

Looking back, users wished they knew people that were using the device, so they could share experiences and help each other. Adopting a smartphone should not be a near impossible task without live assistance.

3.3.8 Limitations

This study was performed with only five participants, all of which started using the same device model, and operating system. Despite trying to minimize the researchers influence in the adoption process and overall learning, we cannot discard the impact the initial and weekly sessions had. Nevertheless, the problems would only increase if participants were not able to receive assistance.

3.4 Summary

The findings presented herein, motivated the work presented in this dissertation. Although the issues portraited are from 2015, many are still relevant and applicable to the current panorama. One exception are the findings portraying the *pre-adoption concerns and desires*. In 2015, smartphones were not pervasive in Portugal, and even less so amongst blind Portuguese. Thus, there was a general lack of experience with the technology and a set of assumptions that have since changed. Nevertheless, they characterize the concerns of the time. The first contact with smartphones has also progressed since 2015, with significant improvements to the first screen reader lessons, although they still appear to not be enough.

In this chapter we presented a detailed understanding of the smartphone adoption process of a blind person. We uncovered theirs struggles, motivations and accessibility issues, from the very first contact with the device and throughout the initial weeks of usage. The accessibility issues raised, and their evolution should be taken into account.

Others can play a huge role in the adoption process. They can be the propellers for application exploration or simply someone to share an experience with. Users will not always be able to resort to others for assistance. We saw the current tools are not enough to support continuous learning and can even provide the wrong stimuli to novice users.

Chapter 4

Mobile Challenges

In the previous chapter we described how we sought to understand the adoption process of newcomers (Rodrigues et al. 2015). The study prompted us to question 1) how and if long term users are dealing with the initial barriers, and 2) if the findings apply at a larger scale with a variety of devices and operating systems.

In this chapter, we present two studies performed with the target population, novices and experts, using a variety of methods targeted at identifying and verifying challenges and coping mechanisms. Through a multiple method approach, we identify and characterize challenges locally with a diverse set of user expertise and devices; and at scale through the analyses of the largest Android and iOS dedicated forums for blind people. We contribute with a catalogue of smartphone challenges for blind people, and a discussion on a set of directions for future research to tackle the open, and often overlooked, challenges, some of which we tackle in the next chapters.

To understand the extent of the challenges found across a larger user base and expertise level, we conducted a series of workshops locally in Portugal. We report on the challenges and coping mechanisms of 42 blind screen reader users. A summary of this study has been presented in (Rodrigues, Montague, et al. 2017).

To capture the challenges without any researcher intervention, and understand how solutions are provided to a larger and more diverse population, we analysed discussions from the largest English speaking Android and iOS dedicated forums for blind people. In contrast with our previous studies, where we had a limited number of users who had fully adopted the device and all its features, challenges discussed by forum users suggest a high proficiency with smartphones - especially from those providing answers and support.

Through the aforementioned studies, we identified and verified multiple causes for the difficulties blind people face when interacting with smartphones. In this chapter, we contribute with **C3) an in-depth assessment of the open challenges blind people face when interacting with smartphones and their coping methods.**

4.1 User Study: Newcomers, Novices and Experts Workshops

We sought to verify if the problems reported in the previous chapter were representative of the challenges faced by blind people at large. If so, we were also interested in understanding how they were being addressed. We recruited blind screen reader users, with different expertise and devices. To observe what challenges naturally occur during use, and understand what and how people cope, we conducted a series of semi-structured workshops, where participants led part of the session based on their questions.

4.1.1 Participants

We had a total of 42 blind participants, 23 males and 19 females, with ages ranging from 25 to 79 (M=51.8, SD=14.0) years old. Participants were recruited through social media, word of mouth and through a local social institution for blind people. Users were asked about their smartphone proficiency when registering for the workshop: 17 did not own any smartphone nor had any previous experience with it; 18 had a device but considered themselves novice, as they were only able to perform simple tasks; and, the remaining eight considered themselves experts.

Table 3 - Workshop sessions conducted.

Session	**Participants**	**Nr Sessions**
Android Novice	26	3
Android Expert	1	1
iOS Novice	9	2
iOS Expert	6	1

4.1.2 Apparatus

The workshops were conducted in a room provided by the local institution. All participants were asked to bring their smartphones to use during the workshop, resulting in a wide variety of brands and models. We also provided a smartphone to all participants that did not own one. The Android devices used the default Talkback screen reader, while iOS devices relied on VoiceOver. Sessions were video, and audio recorded.

4.1.3 Procedure

At the beginning of the sessions, participants completed a short questionnaire related to their smartphone usage and general demographics. We offered two types of workshop, one for the newcomers and novice users, and the second for experts. During the sessions, we engaged participants in a group discussion. These sessions lasted between one hour and a half and three hours, depending on group size and questions. Participants guided most of the session, as they were free to ask questions and collaborate during the sessions. We had separate sessions for Android and iOS with a total of seven sessions (Table 3). For all sessions, we had a team of four researchers available, with experience in mobile device accessibility, who assisted the sessions, guaranteeing a ratio of 2-3 participants per researcher. Each researcher was also responsible for taking notes and observations.

Session for Newcomers and Novice Users

The session was divided in two parts. First, participants were guided through basic smartphone and screen reader behaviours, then participants took the lead and were supported individually, or in groups, depending on their questions and on what each desired to learn.

In the first part, participants started by learning about smartphones and the differences to feature phones. Afterwards, participants were guided on how to perform gestures and basic tasks (e.g. navigating the screen, using the contacts app, writing text). Since in the previous chapter we found the last two lessons of the Talkback tutorial to be too demanding, in the first session we conducted, users completed only the first two lessons, *Exploring the Screen and Using Lists*. During this first session, it was clear participants were struggling even with the first two lessons. Consequently, we dropped the tutorial and we offered personalized instructions to each participant.

Figure 19 - A focus group from one of the Android Novice Session

Although this was the basic structure for the workshop, as the level of expertise varied greatly amongst newcomers and novice users, some led the session from the beginning, skipping the first part and learning about more advanced features (e.g. adding a contact or deleting messages). In the iOS workshop, participants were not newcomers and could already perform some of the basic tasks (e.g. call a contact). We started the session by introducing VoiceOver Practice. Participants could freely perform gestures and when a gesture was recognized (e.g. "Touch - Select item under your finger") its function was read aloud. VoiceOver Practice allowed us to introduce and explain gestures they were not familiar or struggled with. Afterwards, participants led the session and learned to perform tasks they were not familiar with but wanted to learn (e.g. add a contact).

Session for Expert Users

We had two expert sessions, one with a single Android user and a second one with six iOS users. Participants were expected to come up with doubts regarding the usage of their device, thus sessions were fully led by participants' interest. The Android session was a one on one session, since we only had a participant. In the iOS session, we engaged the six participants in a focus group discussion. Participants exposed questions, and anyone could contribute based on their experiences. This approach allowed us to observe how participants convey their knowledge when co-located. Moreover, due to the range and specificity of the questions, input from other expert users proved to be crucial to understand the underlying issues and be able to quickly address them. After all questions were answered, we inquired what types of barriers they face during usage, how they tackle them, how they learned to use the device as newcomers, and if/how they assist others.

4.1.4 Findings

We report on the qualitative data collected during the sessions through the researchers' perspectives. We conducted a thematic analysis to '*identify, analyse, and report patterns (themes) within data'* (Braun and Clarke 2006). We first start by identifying the basic codes on the researchers' notes of the workshops, and progressively iterated and discussed the identified themes with the four researchers that participated in the sessions. When the noted observations referred to participants' comments, we relied on the audio recordings to transcribe said comments. When behaviours were described that required further assessment, we relied on the video recordings.

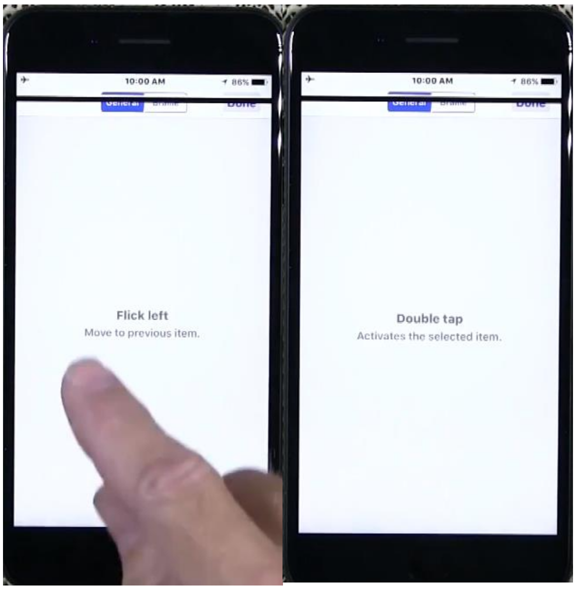


Figure 20 - iOS VoiceOver gesture training

Getting Started

Prior to the workshop, some of the participants reported they had tried to use smartphones, only to give up on the process for being too cumbersome. The current mechanisms to support blind newcomers on mobile devices are insufficient. Although Talkback on Android provides a tutorial to get started, it frustrates and misguides users as we had reported previously (*Chapter 3 - Smartphone Adoption*). In the iOS session, users started in the VoiceOver Practice (Figure 20) which enabled them to understand how their actions were being interpreted by the gesture recognizer. However, by design, VoiceOver Training requires users to already be aware of the gestures available or have someone tell them to try specific gestures. We observed the same problems reported in the previous chapter: smartphone adoption challenges do not appear to be device or operating system dependent. Currently, picking up a smartphone and start using it without assistance is a difficult task with little to no native support.

Discoverability

Only one iPhone participant was a newcomer, yet none of the novice participants was aware of some of the basic gestures they could perform on the device, such as swiping left/right to the previous/next item. Most relied exclusively on tapping the screen, instead of flicking or dragging their finger through it, until they eventually found the option they were looking for. Participants thoroughly discussed issues they experienced with smartphones. For newcomers, their problems were related with touchscreen interactions and simple gestures. However, the cause of their struggles was related to a lack of understanding on how the underlying interfaces were behaving. Participants reported that often they were unaware of the available options.

"In one app I had no way of sharing to Facebook. When I pressed More Actions *nothing happened. What I found out afterwards, when I asked a friend, was that the option was there but it was not yet on the screen. I had to scroll on a new window that appeared*"

All participants reported issues with smartphones, independently of expertise level and device. However, expert users focused more on application-specific issues, such as using advanced features:

"*I am not able to listen to music from my Dropbox in offline mode*"

We observed the same behaviours in Android users: discoverability challenges do not appear to be device or operating system dependent, and are one of the main challenges users seem to face.

Independent and Community Learners

Participants strongly rely on others to surpass challenges, often asking for help from people they consider to be technology experts. We found that users informally created communities that relied on the same specialist; two of them were present in our workshops. They were tech savvy, autodidact, and highly motivated to learn about technology. They regularly read blogs, forums, and mailing lists about assistive technologies, and even contact developers to report bugs and request features. Several participants in the workshop relied on them to cope with daily problems. They provided assistance through a variety of channels (e.g. calls, SMS, Skype) and often about the same issue, but to different people. They reported how their expected availability, given their role in the community, has strained their more intimate relationships.

Sighted Assistance

For some issues, the only possible solution was asking for help from a sighted friend (e.g. screen reader started speaking in a foreign language). However, participants discussed some situations where help from sighted friends and family was challenging due to their unfamiliarity with screen readers. All but one participant mentioned how they preferred to be helped by screen reader users.

"Often the problem is not them [sighted users] not knowing how to solve the problem, the problem is not knowing how to explain to us how we can solve it".

Although sighted people are seen as valuable sources of assistance, most of them are oblivious to the challenges of screen reader users. They usually know the steps needed to accomplish a task but are unaware on how to perform them using accessibility services.

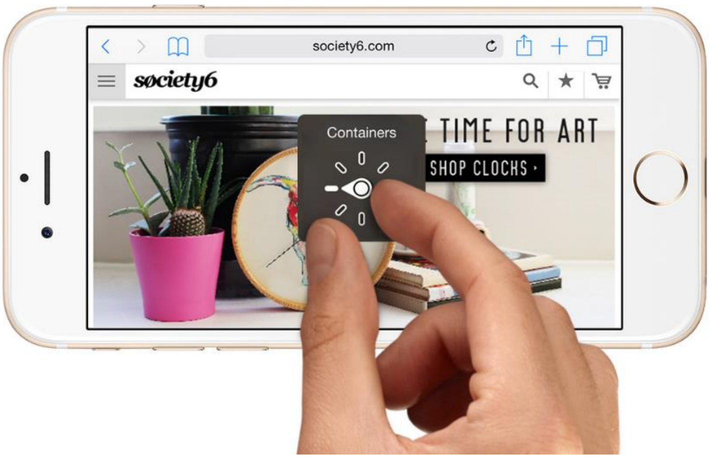


Figure 21 - Rotor Gesture (from https://support.apple.com/en-au/HT204783)

Detailing Instructions

During the workshops, experienced users would often help by guiding others step-by-step, while doing the actions on their own devices and waiting for others to finish each step. For gestural interaction, some participants went further and performed the gesture on the back of the other users' hands. Nevertheless, it was clear that people preferred an active learning approach rather than giving their device to others. The level of detail given can be fundamental to a successful instruction.

About the VoiceOver rotor gesture (Figure 21): *"To use the rotor, rotate two fingers on your iOS device's screen as if you're turning a dial. VoiceOver will say the first rotor option. Keep rotating your fingers to hear more options. Lift your fingers to choose an option."* Although this is an already rich description one of our participants mentioned how it was not enough:

> *"Two fingers? Which fingers? Rotate to what side? Rotate how exactly?"*

The participant carried on and mentioned how he learned that day how to effortlessly perform the gesture, when one of the researchers described it to him.

> *"Use your thumb on the screen and then with your index finger you are able to rotate similarly to a compass; when you hear the option you want to select, lift your finger".*

Before the user heard this description, he was using his index and middle finger to rotate at the same time and was unable to use it. It is important to highlight that the researcher was only able to provide this instruction due to the participant's previous remark.

4.1.5 Discussion

One lingering question from the previous chapter was whether problems existed at different levels of expertise and with different devices. In this workshop, we observed 40 blind people with fifteen unique devices, two different operating systems, and different levels of expertise. The observations, comments, and questions suggested that although barriers can become easier to surpass, they continue to exist. With this study we verified, at a local level, that the issues identified previously are pervasive in the community.

Awareness challenges, mental models, and accessibility issues

Challenges are frequently related to a lack of awareness of the surrounding options - triggered by the difficulties in establishing a complete mental model. When starting to use smartphone apps, participants struggled to find consistency among and within applications. Applications no longer have a single sequence of steps to move from point A to point B, but several alternatives to achieve the same goal. They can also have workflows and loops that may be difficult to recognize.

Some of the problems with understanding patterns and workflows resemble the ones reported in web accessibility with inter and intra page navigation (Vigo and Harper 2013). However, traditional solutions may not work for smartphones: screens don't have unique links or any identifiers, navigational breadcrumbs are not used or practical given the small screen real estate, and back functions are not standardized across apps or OS. Moreover, applications are created by different developers, which makes it even harder to find consistency among apps' interfaces. The issue is exacerbated as users explore more of the device and third-party applications. It becomes highly likely for them to be confronted with inaccessible content, from elements hidden to navigation, to unlabelled ones. Additionally, although we found no instances of this issue during the workshops, mental models may be disrupted by app and OS updates.

The need for assistance does not disappear

Different devices come with different characteristics, but the fundamental problems remain. The support mechanisms on smartphones are not enough, and to surpass the challenges users often resort to external help. Unfortunately, others are not always available, and online help (i.e. checking forums and reading tutorials) is only reported to be used by a few tech savvy experts. Furthermore, our findings suggest this is causing issues beyond smartphone accessibility, with newcomers abandoning devices and experts having their relationships strained due to tech assistance.

4.2 User Study

Community Forums

In the expert sessions of the previous study, we observed that the experts who were providing assistance within the community were also the ones that relied on technology, like forums and online guides, to overcome issues. Likewise, many others around the globe do the same. To capture a wider range of problems and understand how solutions are presented in an environment without any researcher intervention, we relied on online ethnography (Postill and Pink 2012). We analysed the top content in the largest Android and iOS dedicated forums for blind people. The data collected allowed us to understand how knowledge is shared in-the-wild in an online community.

4.2.1 Data and Analysis

In this study, we started by selecting one forum of both major mobile operating systems (OS). Secondly, we analysed the top thread titles and selected threads with content relating to barriers and assistance for further inspection. We then performed a thematic analysis (Braun and Clarke 2006) on the selected content, with a focus on understanding the type of barriers and how solutions are provided. Themes were discussed and iterated within the research team.

Dataset

Although there are multiple relevant online communities in both OS, we selected the ones with the higher number of threads. Viphone[18], the iOS representative, is a Google group with over 41989 threads. Eyes-Free[19] is an Android Google group with over 27898 threads. Google groups is a service from Google that allows the creation of discussion groups. The content can be created and consulted online in a forum-based interface or as an email-based system. When relying on an email client, the number of views is not counted towards the thread statistics. Thus, we used the sum of the number posts and views in each thread to select the highest-ranking content.

To discard outdated discussions, we limited our search to threads with content from January 2015 onward. The data was collected on January 2017. One researcher coded the title of the top 100 threads of each forum. All titles coded as *Doubt* (i.e. question about something), *Problem* (i.e. specifies an issue), *Guide* (i.e. app, feature or device guide), *Getting Started* (i.e. mention getting started) and *Request* (i.e. request for information) were selected for further inspection.

Then we analysed the first message and discarded all threads that did not discuss smartphone applications, devices or features, resulting in 48 selected threads (i.e. 19 Android and 29 iOS). Most of the discarded threads were coded as *Announcements* (i.e. release announcement app/product). The selected threads have 45 unique authors, 2502 posts and 10968 views. Posts per thread range from 2 to 660 (M=52.12, SD=100.98) while number of views range from 9 to 4173 (M=224.14, SD=601.29).

Data Analysis

We conducted a qualitative analysis of the 48 threads. We used a combination of inductive and deductive coding; two researchers coded a set of 10 threads independently. Then the two researchers discussed, iterated and merged the two codebooks. The final codebook was used by one researcher to analyse the remaining 38 selected threads. Threads had as many as 660 messages and often, after a number of posts, the discussion either shifted to

[18] Viphone Google Group (https://groups.google.com/forum/#!forum/viphone)

[19] Eyes-Free Google Group (https://groups.google.com/forum/#!forum/eyes-free)

irrelevant content or repeated arguments. Thus, to prevent analysing unrelated content we stopped coding a thread if three sequential posts were marked as *Other content*. To reduce the amount of repeated content, after the first 10 posts we started assessing if the next three posts contained novel content; if not, the analyses of that thread was stopped. We coded a total of 524 individual posts (M=10.91, SD=6.47 posts per thread).

4.2.2 Findings

In this section, we present the two overarching themes that led the data exploration: *Barriers* and *Solutions*. In *Barriers* we identified the challenges described by users; in *Solutions* we report which and how they are provided. In this section, we use "users" when referring to the forum users and "*Authors*" to indicate thread authors.

Barriers

Barriers are depicted in 41 threads (85%). The remaining threads are open discussions where *Authors* request other users' opinions about a device, feature or application.

Text-entry

As expected, we found multiple instances on the top relevant content discussing text-entry problems. Users are slower entering text in touchscreens when compared with physical keyboards. Consequently, users look for alternative means to write, from QWERTY to Braille virtual keyboards or physical keyboards.

> *"iOS Braille is so much faster for me and I get much less errors."*

However, speed is not the only challenge users face. Currently, editing, copying and pasting text is cumbersome. Users wished they found a simple way to enable them to manipulate text as they are accustomed to in other systems.

> *"What bothers me is the clunkiness of the system for selecting and editing text."*

Gestures

We did not find any gesture related issues, with the exception of the required gesture to answer a call. The underlying challenge was not how to perform the gesture but rather how to interact with the interface presented. However, many of the described issues in other threads may be caused by gesture difficulties that the user is unaware of. For example, not finding an element on the screen, due to an incomplete exploration, may be caused by users skipping elements during exploration.

Visually rich interfaces

A rich visual interface allows a sighted user to quickly understand the structure of the content. On the other hand, blind users have to scan through all of the interface to create a mental model. Moreover, the navigation using a screen reader is affected by the underlying interface grouping. Therefore, for an efficient navigation, users must create a mental model that incorporates the interface structure in each application.

"Remember! If you move your finger to the top of the screen, into a different area of the screen, you will "automatically!" go back to the default navigation/group level. So, do keep that in mind."

Structure and correct labelling of elements is essential for the accessibility of an app, but it does not guarantee it. There is information that is lost when using a screen reader. When first playing Minesweeper, a sighted user will quickly understand the rules and objective of the game through the visual cues provided. In a fully accessible Minesweeper application, a user that is unfamiliar with the game will struggle to understand it.

"I've never played a mine-sweeper game before and so, when playing their accessible mine sweeper, I didn't have a clue what to do"

Discoverability

Application features need to be clear and discoverable. We found multiple reports of users unable to perform an action in an application, not due to content that is impossible to access, but due to a lack of discoverability, corroborating the first study findings.

"Somehow the Alarm volume got set at a very, very high loudness level. I can't seem to find a way to tone it down."

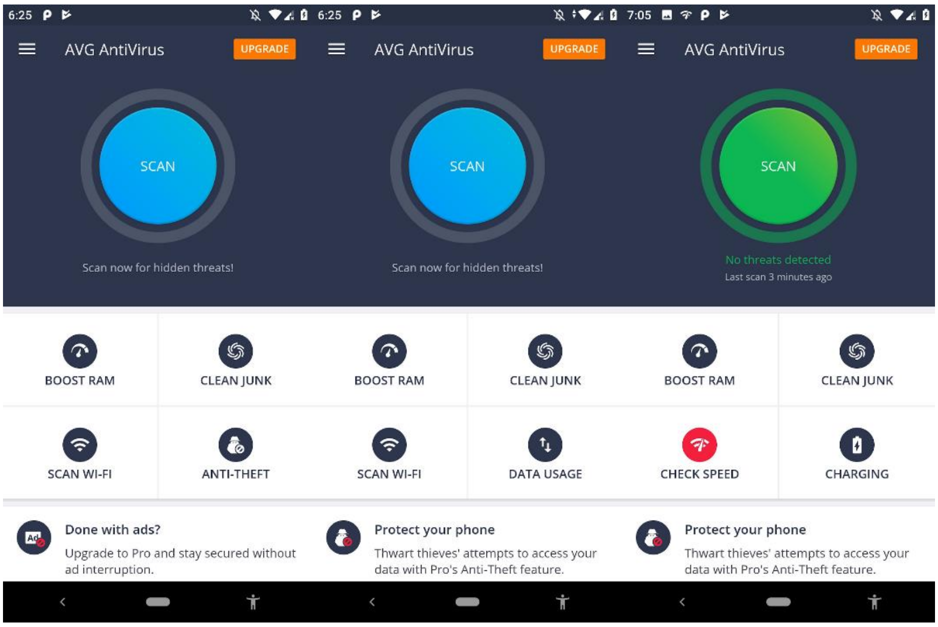


Figure 22 - Example of dynamic content in the app AVG Antivirus. The menu options change every time we open the app

Dynamic content

We found that one of the reasons users struggle to create accurate mental models is because many mobile applications have dynamic content (Figure 22). Apps where repetitive interactions cause different outcomes disrupt users mental models, similar to what happens when the back button behaves inconsistently (Rodrigues et al. 2015). A common case of this behaviour is when users are faced with an ad during their interaction.

"Then, without me appearing to do anything, I seemed to be hurled, into the world of vehicle and personal accident insurance. Where in the wide world that came from, I haven't the foggiest."

Updates and Versions

We confirmed what we posited in the previous study. Updates can, and do, further disrupt users' mental models and prevent users to take advantage of all features. Smartphone OS and applications are updated frequently (Figure 23), which can lead to user segregation with respect to the versions available or installed on devices. Individual versions may possess different features and requirements.

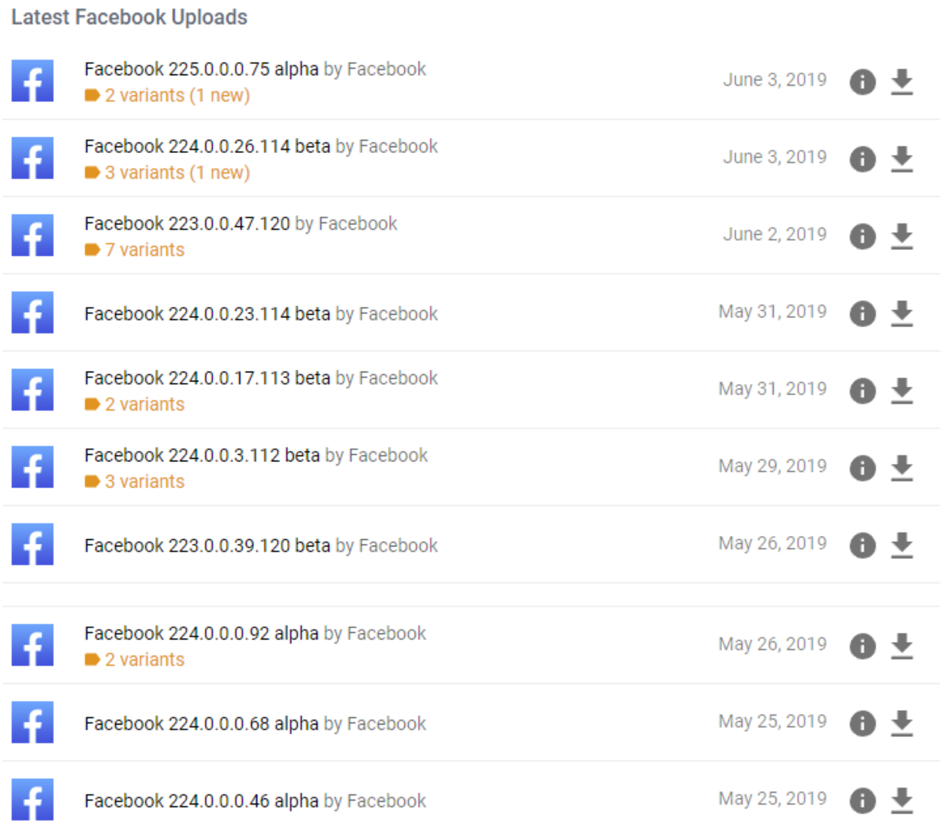


Figure 23 - List of the updates to the Facebook app in the span of 10 days.

"They don't let me buy it!!! Probably, Google sees my Galaxy s3 as an "old Samsung device", not knowing it is now running 5.1.1... "

Updates often bring new features and behaviours. Unfortunately, frequently, the changes cause new bugs and barriers to appear. When there are screen reader or OS updates it potentially affects the whole user experience with the smartphone.

"I've upgraded to 5.0.2 in my Moto G 2nd gen, and now, I'm facing audio cut offs. Talkback is not speaking properly, and audio quality has also been decreased. "

Efficiency

Smartphone interfaces are designed to be visually appealing and provide fast direct manipulation. Some tasks are inherently associated with the underlying interface structure and item location. When using a system-wide screen reader that modifies every interaction, these tasks may become cumbersome or even impossible to perform.

"I can go into the app drawer and tap and hold on an app to add it to the home screen, but this seems an awfully long way of doing things and I still can't add a widget or shortcut which is pretty frustrating. So what am I doing wrong here? I'm so confused and most of all frustrated."

Sighted Assistance

As described in the previous study, we found reports of non-screen reader users not being able to use the device due to its different interaction method. Thus, it can prevent them from providing effective assistance.

"In my experience, even just having VoiceOver on, makes the phone fairly difficult to use for a sighted person, since the VoiceOver gestures are very different from the way a sighted person would normally interact with a touch screen device."

Solutions

In every thread where a barrier was presented, other users contributed to the discussion by sharing their experience, providing solutions, and clarifying features. However, not all problems were solved. In some cases, all the author was advised to do was to contact the developer or manufacturer due to the nature of the barrier. In this section, we report on the different types of solutions and how they were presented to users.

Proficiency dependent instructions

When a user had the knowledge to solve a problem, they would often guide the author to the solution. When providing the answer, the user would assume a certain degree of proficiency of the author. If the author was experienced, the instructions given could be presented at a high level just by describing the overall steps that needed to be taken.

"It is called NuPlayer, and you can disable it in the developer settings."

In cases where the author was inexperienced or asked for guidance, each step was presented in greater detail, including the required gestures/navigation and the expected feedback.

"OK, first you must enable developer options. Go to Settings, About Phone, then tap on the build number about seven times. (...)"

Location information

Unexpectedly, most steps do not contain relative or absolute position of the items. Users are simply instructed to find item <x> in screen/layout <y>. Of the 82 references to guiding, only 12 contained a location reference. When location is provided, it is usually absolute or relative to the always present keys (e.g. home, volume). Absolute locations are provided to items close to the edges of the device (e.g. bottom left, top right). We only found one instance of an absolute location not associated with an edge (i.e. *"check in the middle of the screen, towards the top*") and one of relative location associated with a virtual button (i.e. *"find the 'audio switch' on right of the 'dial pad switch'*").

Sighted People Required

Unlike our previous studies, we only found two instances where sighted assistance was required. In a particular case, a user was inquiring about a phone and its out of the box experience. Unfortunately, that device required sighted assistance to enable the accessibility options.

"The out of the box experience, is not quite so good though, as you can't independently set it up."

The second instance was an unexpected use for a sighted assistance. Users asked the author if he could reach out to a sighted person so he could provide greater detail about the problem he was facing.

"If you have enough vision or have a sighted person available, can you let us know if the VoiceOver visual cursor is also skipping various elements as well?"

4.2.3 Discussion

We analysed the content of two open online communities dedicated to smartphone accessibility. Therefore, we do not have any demographics or information about each individual. However, we know users can communicate in English, and that when faced with a barrier relied on online communities to overcome it. Even without assessing the content of each request these users are at the very least proficient enough with either smartphones or computers to browse and query for questions online. In contrast, in our previous studies, users reported the need for co-located assistance, the lack of its availability and their high dependency on others. In this study, the majority of users tried to address the topic by discussing possible solutions and by guiding the thread author to a successful outcome without external help. Only when the problem seemed to be unmanageable did users suggest getting external assistance (e.g. retail store, co-located sighted assistance).

Instructions are concise, neglect location and app structure

Users provided concise instructions, guiding the user with high level instructions when possible. When more detailed instructions were required, either by author's request or due to an assumption on the author's expertise, users provided step-by-step instructions providing the target element description. Although many of the problems seem to stem from the lack of understanding of the underlying app structure, or how to reach elements efficiently, instructions do not contain location nor an overview of the app layout. This can be accredited to the forum being dedicated to visually impaired users, thus the visual representation and element location can be hard, if not impossible, to understand. Moreover, there are two diverse navigation methods that each user can rely on, thus what would be a helpful app overview for a linear navigation might not be for an *explore by touch* approach.

Lost in constant updates and dynamic content

Many of today's mobile apps rely on dynamic content for a variety of reasons, from ad revenue to keeping the user updated with the latest news. However, many changes in-context and content are strictly announced through visual stimuli. The underlying cause can be undetectable by blind users. Users often feel a loss of control that can even be.

Table 4 - Smartphone Challenge Catalogue

Category	*Challenge*
1. Gestures & Navigation	1.1 Perform a specific touchscreen gesture
	1.1.1 Learn a new gesture
	1.1.2 Perform a gesture consistently correct
	1.2 Perform unintentional gestures
	1.3 Scroll through lists
2. Awareness	2.1 Find the desired option
	2.2 Be aware of the available options
	2.3 Be aware of supported gestures
3. Mental Model	3.1 Get lost when using an app
	3.2 Unexpected navigation triggered without performing an action
	3.3 Return to the first screen of an app
	3.4 Understand content structure
	3.5 Understand features and their effects
	3.6 Adapt to an update
	3.7 Overwhelmed by features
4. Feedback	4.1 Smartphone is unresponsive
	4.2 Screen reader is slow
	4.3 Not knowing the meaning of a specific sound
	4.4 Screen reader is too verbose
	4.5 Overlapped feedback
5. Text	5.1 Text input is slow
	5.2 Editing text
6. Security	6.1 Login in apps and websites
	6.2 Unlocking the device
7. Accessibility	7.1 Labels are inadequate or inexistent
	7.2 Unreachable elements
8. Hardware	8.1 Configuring external devices to connect to the smartphone
	8.2 Using capacitive buttons
9. Sharing & Assistance	9.1 Consistently reproduce an issue
	9.2 Others do not know how to use my device
	9.3 Be aware of what others are doing when handling my device
	9.4 Restrict others access to one's personal device
	9.5 Find knowledgeable assistance
10. Knowledge Void	10.1 Know which device to purchase
	10.2 Find fully accessible apps

disorientating. As a side-effect, it can cause users to struggle to create accurate mental models. Several of the most popular threads analysed were discussions around the impact of bugs and issues of the updates, to either applications or operating systems. These constant updates and revisions to both function and interface only escalate the difficulty of creating mental models, which is already a demanding task to accomplish

4.3 Outlook on Smartphone Accessibility

Smartphone accessibility is currently in a dichotomy of states. We can consider smartphones accessible because accessibility services (e.g. screen readers) allow users to reach every piece of content, and many applications are fully accessible in the sense that every element is identified. Moreover, we have people that have been able to adopt the device and become proficient users, able to tackle almost any issue thrown at them. However, for many, the current state of affairs is not enough.

People struggle to adopt the device and continue to face challenges at every level of expertise. Moreover, two blind people using the same interface can have a vastly different experience, depending on the interaction behaviours, ability, and desire to go through a trial and error process. It is now evident by these studies the lack of support people have when starting to use the device, and how much they have to rely on others to overcome the frailties of the process. For many without an established network of support, this leads to either never transition to a smartphone or forever remain a user of basic phone tasks. There are currently no support mechanisms that support users to evolve their smartphone usage.

Once a user goes through the adoption process, the challenges do not cease, they simply morph. Newcomers struggle to establish mental models and are overwhelmed by the variety of behaviours, interfaces, and feedback mechanisms. For experienced users, new apps and updates create the need to adapt, rediscover and commit entire new interfaces to memory in order to efficiently use their smartphone. None of this would be a problem if all it took to adapt was a quick glance. However, the equivalent for blind people is a long process of hearing everything in the screen and possibly interacting with a few elements to assess the changes and their impact. Another consequence of the variety of interface layouts and behaviours is the lack of awareness of available options that plagued users of all expertise levels.

The number and diversity of challenges users have to overcome leads them to different coping methods. Some neglect the device and only rely on it for simple tasks. Others are still carefully handling old feature phones in an effort to prevent the frustration and self-doubt on one's ability that comes with struggling to operate a mobile device, especially one that for all intents and purposes can be accessible. The social pressure of knowing people around them were able to adapt can be damaging when they first start and struggle to learn. The most common coping method is relying on friends and family whenever a challenge appears, which is far from ideal. For the few that are tech savvy, online help from community forums and other platforms (e.g. *Youtube*) resolves part of their issues, with the remaining requiring persistence - having the desire and ability to learn through trial and error.

Despite all efforts in mobile accessibility, smartphones are still not accessible to all. People face a variety of challenges that prevent some from taking advantage of their device features and make others dependent on their support network.

In this work, we identified and verified a large set of challenges experienced by blind people (Table 4). We reported insights about the current coping mechanisms.

4.4 Multiple-Methods Approach

We presented three user studies (i.e. two in this chapter and one in the previous) where we relied on a variety of data sources and collection methods to have a holistic view of the challenges faced (and coping mechanisms employed) by blind people, when interacting with smartphones. In the previous chapter study, we relied on a mixed-method approach that allowed us to analyse, in great detail, the evolution that our five participants went through. In this chapter's first study, we reached a wider participant pool and realized that locally, the majority of users were newcomers or novice. We understood that the learning process never stopped, and some remained at a novice level despite years of experience due to the lack of support mechanisms. We also identified a select few that take advantage of all the device has to offer. The second study led us to question the challenges faced by the user group of expert tech savvy users. We became interested in understanding how these users addressed their issues. In this chapter's second study, we focused on understanding the issues depicted in the top online community forums, giving us yet another view of the challenges, and coping mechanisms that can be effective for a segment of the user group.

Had we limited our research to a single laboratory or in-the-wild study, we would only have a fragmented view of the problem, capturing only snapshots of the challenges faced. Without multiple methods, we would not have captured the full spectrum of the smartphone challenges from newcomers to experts; and would have no understanding of how challenges morph and how different users tackle the same issues differently. When understanding users' behaviours and challenges with technology, every study raises questions that should be pursued. Through a multiple method approach we were able to delve wide and deep into the challenges blind people face when interacting with smartphones. The approach allowed us to understand the adoption process, the variety of expertise, the challenges, the coping mechanisms, the different impact of different stakeholders, and users' knowledge progression.

We urge researchers that are exploring challenges in technology adoption and usage to take a multiple method approach, providing a holistic view that encapsulates all types of users' challenges, coping mechanisms and identifying the different impact factors, and thus, possible areas for improvement.

4.5 Summary

In this chapter we identified a set of open challenges. Several of the identified issues have been previously reported and heavily researched (i.e. inadequate labels (Takagi et al. 2008), input text (Rodrigues et al. 2016; Nicolau et al. 2017; Azenkot and Lee 2013) learning and performing gestures (Oh, Kane, and Findlater 2013)). However, many others have been largely overlooked or under explored (i.e. getting lost while using an app, unawareness of supported gestures, inability to find desired option). Below we discuss the open challenges that have guided the work conducted during this thesis:

4.5.1 Learning and Discovering

Although there has been some work on how to learn and perform gestures on touchscreens, it appears the challenges are not yet surpassed. Gestures are not discoverable nor easy to learn based on some of the given descriptions, and users have no fallback mechanism to rely on, except for their support network. We can start to explore how gesture discovery, and practice, can be embedded into everyday interactions, how we can track performance, hint at corrections, adapt recognizers or even develop entertainment apps whose sole goal is gesture practice and discovery.

Awareness of the available options was one of the most reported challenges. Past work in multiple audio source (J. Guerreiro et al. 2015; J. Guerreiro and Gonçalves 2014) has explored how to augment awareness in text input, and when passively consuming news. There is an opportunity to leverage user’s ability to segregate multiple audio sources and explore novel audio navigation techniques that can be applied to other tasks.

4.5.2 Adapting Mental Models

The fast paced, iterative, nature of the devices' OS and apps bring additional complexities. Interfaces can be radically reinvented with each new OS update or subsequent version of an app. This means that users must also continuously rediscover and adapt their mental models of interaction to maintain their current level of expertise. However, there has been limited work exploring the design space of assisting blind people with these challenges on mobile devices. For example, how we can inform users of app interaction changes that are relevant for their usual routines. There is an opportunity to explore novel methods that guide, inform, and promote smartphone usage on a continuous basis beyond first steps.

4.5.3 Enable Sharing and Peer Support

We found that not only is the adoption process a long and arduous task, but there are limited tools to guide and support users through this process. The status quo relies on user’s persistence and ability to seek aid from others to overcome challenges. Regardless of how well a technology is designed, there will always be times when additional peer support is needed. Through our workshops and interviews we heard tales of family and loved ones tirelessly answering phone calls, and requests for assistance and guidance to everyday tasks - with individuals being asked the same question repeatedly.

We've seen innovative solutions that leverage crowd workers to assist people with disabilities in overcoming daily challenges (Bigham et al. 2010). While these works have demonstrated that small micro-tasks and contained questions can be easily handled by crowd workers, there is still a need for longer, more engaged and curated support such as one-to-one walkthroughs and tutorials.

More experienced users were able to leverage existing technologies (i.e. asking questions in on-line community forums). However, many others were unable, or found the information inaccessible or incompatible with their device configuration (Hardware, OS and App versions). There is an opportunity for new forms of assistance that don't require the user to take additional complex steps, that are tailored to their individual needs, and are readily available when needed.

Chapter 5

Human-Powered Support

In the two previous chapters we characterized the smartphones challenges blind users currently face. The studies revealed the current lack of support mechanisms, the need to rely on others, and posited the use of novel human-powered approaches to tackle the identified issues. In this chapter, we explore if, and how, human-powered technology can be leveraged to fill the current gap in assistance.

Nowadays, assistance provided by others is becoming pervasive among many of our software's and web applications. For instance, open source projects, have a long history of leveraging public community forums, mailing lists and even social networks to enable users to help each other (Lakhani and von Hippel 2004). In the previous chapter, we have highlighted how this type of assistance is currently out of reach for less tech savvy users. Past work on crowdsource assistance (Chilana, Ko, and Wobbrock 2012) has discussed how the current solutions rely on user's ability to ask the right questions, and find the right content. Currently, users are limited to relying on others for help or searching online for answers. Both are cumbersome, take the user out of the context the problem rose in, and rely on the user being able to portray the issue, often not producing any result.

In this chapter, we start by exploring the perceived usefulness and acceptance of human-powered assistance in smartphones when provided by a domain expert with full context knowledge. In an exploratory study, six blind participants performed a set of tasks with access to the developed in-context Question & Answer (Q&A) service. Initial perceptions showed positive, and promising, results related to in-app support and self-organized learning. However, the study also alerted for the difficulties felt by the expert, that without full context knowledge, would not had been able to help. A Q&A service is also limited to helping the users when they have a problem. They are not designed to continually support a task.

Tutorials, on the other hand, are created with the intent to guide users through a task. They can be ideal to allow users to get started with the use of an app and guide them through what the developers deemed relevant. However, they tend to not be accessible and to be limited in coverage. In the remaining part of this chapter, we sought to understand how to design human-powered tutorials.

To maximize the number of possible tutorial *Authors*, it is essential to look beyond just app developers, and support other users to contribute with assistive content. We conducted a user study exploring the authoring and playthrough of nonvisual tutorials, where untrained individuals are at both ends of the technology. First, we conducted an authoring session with five blind and five sighted participants where we asked them to create tutorials for four smartphone tasks. To evaluate the quality of the created tutorials, we conducted a playthrough session with 12 blind participants. Participants were asked to follow the tutorials, while interacting with the researcher whenever they needed additional assistance. We found that instructions by sighted people were more concise and often had misleading target information, due to the challenges of converting visual references to accurate textual descriptions. Even though blind instructions were accurate, it was clear, in both tutorial types, that users required additional assistance that was not contemplated in the instructions provided.

In this chapter we contribute with **C4) a comprehensive set of insights for the creation of nonvisual smartphone assistance**.

5.1 User Study: Hint-Me! A Human-Powered Q&A Service

Prior research in crowdsourcing has allowed users to identify and locate objects in the real-world through visual questions using a smartphone (e.g. VizWiz by (Bigham et al. 2010)). Following a similar approach, the Social Accessibility project (Takagi et al. 2008) provided a collaborative metadata authoring mechanism to enhance webpage accessibility. In Chilana et al., authors developed and provided a Q&A in-context service for web applications (Chilana, Ko, and Wobbrock 2012).

Based on our previous findings, and inspired by these prior works, we developed Hint Me!, a human-powered service that allows blind users to get in-app smartphone assistance. Using Hint Me! as a design probe, we conducted an exploratory user study with six blind participants to elicit their perceptions on the usefulness, and acceptance of human-powered networks for smartphone support.

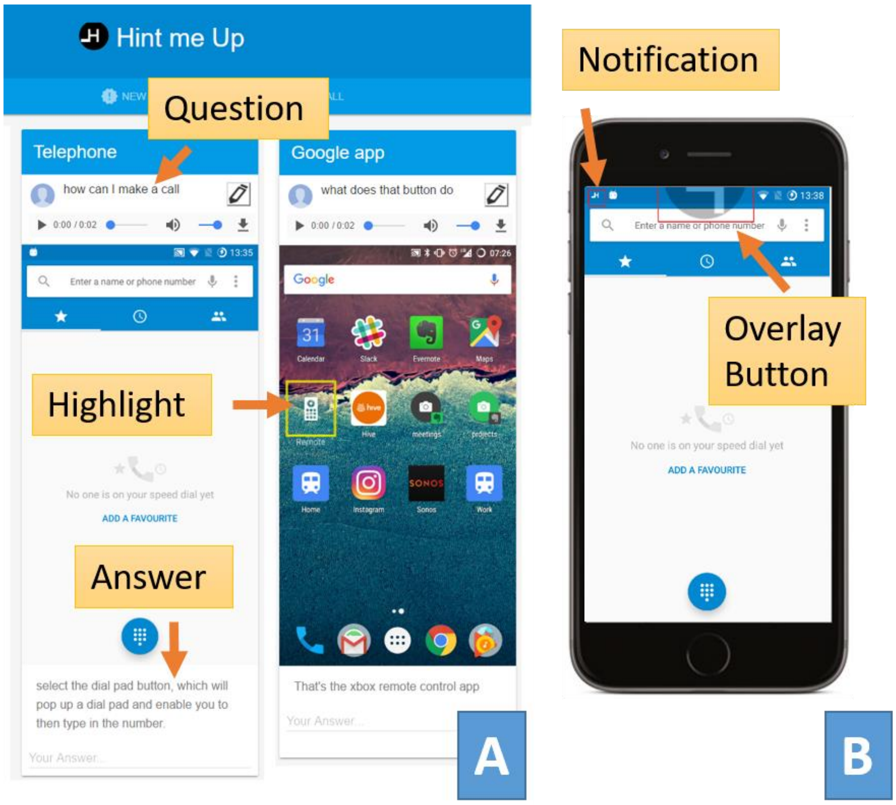


Figure 24 -A) Volunteer web app. It shows two answered questions, one with a specific element of the interface highlighted. B) Hint Me! with the always available button on the top of the screen, and a notification showing the user he received an answer.

5.1.1 Hint Me!

Hint Me! is an integrated Android service that enables users to connect with a support network of people willing to provide assistance. With it, every question is linked with the context it was created in. Using an overlay button, users may quickly browse existing questions and answers associated with the app, current screen, or with a particular element they select. Moreover, since it is an independent service, it is available system-wide in any mobile application. Hint Me! supports, among others, workflow guidance (e.g., how to perform an action), layout description, labelling, and learning workarounds for inaccessible content.

5.1.2 Creating a Question

Users can ask questions, within any application, through an ever-present quick launch overlay button (e.g. Figure 24 - B). Questions are recorded and their text is extracted relying on Google Voice Recognition technology (Figure 25). Additional information is collected to enable us to present the question to volunteers, augmented by in-app contextual information (e.g. Figure 24 - A). In order to provide context, Hint Me! automatically captures a screenshot of the device alongside all element's details (e.g. alternate text, text, position, dimensions), creating the DOM tree structure of the interface and enabling its re-creation; in case the question relates to a specific element of the interface, users can select it in order to be highlighted.

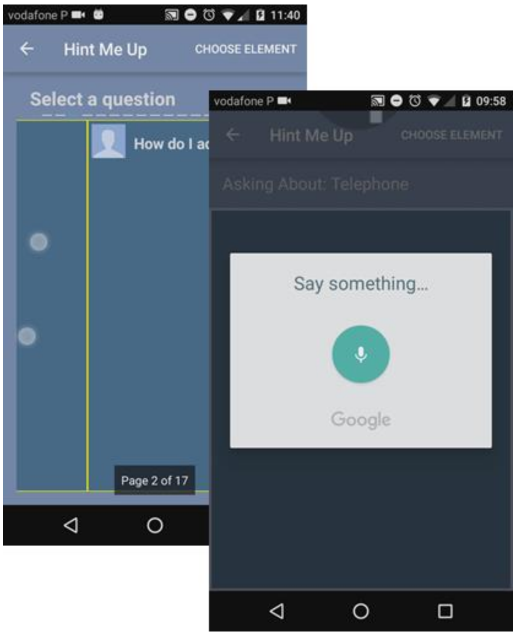


Figure 25 - Hint Me mobile app, browsing answers and asking a question.

5.1.3 Getting an Answer

Hint Me! gradually builds a shared knowledge-base with questions previously asked. People who volunteer to provide answers can edit the question for clarity or to correct errors from the speech conversion. Questions only become available when they have been validated and answered. When an answer is submitted, the author receives a notification.

Users can browse through all Q&A associated with their current context (Figure 25), or through their asked questions. Within their current context, Q&A will be filtered according to their current app or screen. For example, if the user has Facebook open, then only questions that were asked in that particular app screen are shown. Additionally, users can select an interface element to navigate content specifically associated to it.

Users can select an answer to pin it to the Hint Me! overlay button. Long pressing the button announces the answer, enabling users to follow long and hard answers without the need to memorize them or switch context.

5.1.4 Participants

Using Hint Me! as a probe, we conducted user study in an institution for visually impaired people where we recruited 6 participants, if which were 3 females. Their ages were comprehended between 31 and 62 years old (M=45.7; SD=12.6). All had previous experience with smartphones: P2 a month, P1 and P3 a year, P4 two years, P5 and P6 over three years. All were legally blind and screen reader users.

Table 5 - Possible tasks per application.

Application	***Task***
Facebook	Check <John's> friend profile
	Check nearby locations
	Share a photo from your photo gallery
	Share a video from your gallery
WhatsApp	Send a message to <John>
	Call <John>
	Send a photo or video to <John>
	Send your location to <John>
	Create a group conversation with <John> and <Mary>
Du Speed Booster	Release memory
	Optimize
	Check which apps have the access account permission
Spotify	Find the artist Amália and play one of her songs
	Find and listen to the album *Puro* from Xutos e Pontapés

5.1.5 Apparatus

We used a Vodafone Smart Platinum 7 smartphone running Android 6.0 with Talkback and Hint Me!.

5.1.6 Task Design

We selected four apps from the top Play Store apps, discarding apps from the same category. The apps selected were Facebook, WhatsApp, Du Speed Booster, and Spotify. Participants completed four tasks, two in two different apps per session. Applications were counterbalanced between sessions. Tasks were created based on the Play Store descriptions and participants were asked to do the ones they were less familiar with. Here we present only the tasks that were performed at least once from the pool of the created tasks (Table 5). The goal of this study design was for users to try and accomplish tasks they were not familiar with, with the support of a human-powered service.

5.1.7 Procedure

The study was divided in two sessions, each lasting one hour: 1) posing questions; and 2) browsing existing content. The content generated in the first session populated Hint Me! with Q&A derived from the users' needs. One researcher acted as the volunteer through a web-app (Figure 24 - A), which represents the optimal scenario of a volunteer being an expert user. The researcher had previous knowledge of the tasks and was able to listen to the user interacting with the device. Although our focus was to understand the perceptions of the end-users receiving assistance, we also report on the volunteer experience.

The first session started with a preliminary demographic and experience questionnaire, which included questions about the experience with the selected tasks and apps. After a brief explanation of the objective of this research and the study, the participants were briefed about Hint me!. Participants were guided through its features and were encouraged to ask questions. Then participants performed four tasks, two in each of the selected apps. Participants could resort to Hint Me! when they felt they could benefit from it. Each task started with the researcher reading aloud the task description; participants could, at any point during the task, prompt the researcher to repeat.

Participants could only ask questions through Hint Me! in the first session. When a question was submitted, the researcher used the volunteer web-app to provide an answer. In the second session, participants had to exclusively rely on browsing questions. Therefore, the only tasks that were assigned for participants to complete were ones that had been completed by other users during the first session. After the second session, we conducted a semi-structured interview to investigate the perceived usefulness and acceptance of Hint Me! and its underlying approach.

5.1.8 Findings

In the first session, a total of 21 questions were created (e.g. *"How is the page organized?"*, *"How can I reach the artist since I cannot find him in the list?"*, *“Which button is the optimize?”*), and each participant asked at least two. In the 24 tasks of the second session, Hint Me! was opened 18 times and 16 answers were consulted, with all participants relying on them at least once.

Two researchers inductively created two codebooks from a set of three interviews. The codebooks were discussed, iterated, and merged. Then, researchers coded three other interviews independently and reached a Cohen's Kappa agreement of k=0.67, which can be considered from fair to good (Fleiss, Levin, and Paik 2013). Below, we detail our findings, anchored to the four main identified themes, followed by the experience report from the researcher that acted as a volunteer.

From Aid to Self-Organized Learning

All participants reported positively showing interest in installing Hint Me! in their devices. P1 and P4 felt that having direct answers to their very specific questions was the most useful feature, allowing them to surpass many of the barriers previously encountered. Hint Me! was a safety net to explore new applications. It was described as a learning tool that would give users autonomy to fully control their devices.

"For example, nowadays I don't use the Internet on my phone. But if I had access to Hint Me I would have started using the Internet already. I am sure." P2

"[with Hint Me!] we have more autonomy in using the device because we are not dependent on others to tell us how something is done" P4

Interestingly, P2 and P4 felt that *Hint Me!* allowed them to learn without the dependency on others, although we explicitly told them that someone would be answering their questions online. These comments suggest that *Hint Me!* and other human-powered technologies have the potential to reduce the social barriers associated with asking for help, and increase agency and self-efficacy.

"Sometimes people don't have the time to explain to us [how to do things]. If I had this service, I wouldn't need to bother other people, I would just do them [the questions] here" P2

P6 explicitly valued the active learning approach; it is the users who perform the actions by learning and following a set of instructions rather than having someone taking over and performing the task for them.

“I like this does not work as remote assistance; people have access to an image but can't control the device." P6

Participants identified several scenarios where the tool would be helpful. Four participants mentioned *Hint Me!* could be useful when exploring new apps or after an update. The system was also seen as a tool to report and deal with malfunctions or interface elements that had unexpected behaviours.

“[I see myself using this app] mainly in an app that I am using for the first time, or maybe after an update, when new features are made available. Or there can even be a bug which already has workarounds available" P1

P6 saw Hint Me! as an in-app training tool, rather than a questioning app:

"It is useful to describe the app, its structure and layout. It helps. A lot of blind people do not have a mental model [of the app] and can't do things easily - oh it's on the centre of the screen or a little more to the right - they don't have that mental picture" P6

On the other hand, *P*2 focused on using Hint Me! to surpass basic accessibility problems, such as mislabelled or unlabelled buttons:

"I recently installed news apps and some of them are not accessible at all. With this app I could understand which button to press to get to certain sections" P2

Questioning vs Browsing: A Trade-off

Participants identified value in both being able to create a new question and browsing previous stored knowledge. However, when asked about the foreseen usage of the system, they revealed different perceptions and preferences, namely regarding the way they would retrieve knowledge.

"I think I would check the database first. This way, I wouldn't risk making a question that was already asked. If I couldn't find it, I would then add one more question". P5

"*It is always easier to ask a question if the answer comes right away; if it is about the app's layout, I would search for an [existing] question, because that question was probably made, and it would be faster to search rather than ask a new question; if it is something that probably no one asked before, it's easier to ask." P6*

Other participants reinforced time of response as being relevant in their foreseen operation of the system. Time was not the only reason for a browse first approach; other participants leaned towards it as they had doubts about their ability to accurately formulate a question.

Anonymity and Answer Quality

We asked participants about whom they would send their questions to, particularly between unknown volunteers, close friends, or their broad social network. Participants preferred directing the questions to the volunteer group, choosing anonymous communication. The main reasons were related to not overburdening their family and friends, and to the limited knowledge that this closer group may have. There is a common belief that the group of volunteers would be more qualified, both at the application and accessibility level.

"[closer] people are not aware of accessibility (i.e., Talkback), probably they will not be able to help much" P1

However, sharing their in-app information with unknown people was considered a possible issue, where additional contextual information is required.

"I'm not sure what the screen capture shows. I think it would be important for us to understand how much of it is being captured." P5

Asking questions to close people was considered useful when sensitive information was involved.

"If it had [personal] information, [...], I would be more comfortable asking someone I know. But if it was «what is this button in this app, where personal information is not shown», in this case I wouldn't mind asking a broader group of people." P6

People disliked the possibility of having questions posted in their Facebook. They do not feel it is private nor safe. Similar findings were reported in Brady et. al. (E. L. Brady et al. 2013) where blind people were reluctant to use social networks for visual Q&A (i.e. a question associated with a photo) due to the perceived social cost.

When asked about sighted or blind volunteers, participants reinforced volunteers should be experienced with accessibility services to ensure useful responses.

The Role of the Volunteer

There is often a mismatch between the visual information and the output of the screen reader. Without using a screen reader or having additional contextual information, the volunteer would not have been able to answer all questions. One example was a mislabelled option when searching for an artist on Spotify, the first step involved opening a menu incorrectly labelled as *"Go Up"*, instead of *“App drawer”* or *“Menu*”. In some cases, even layouts with correct labels may not be enough to incite a clear answer (e.g. duplicate labels).

Without rich contextual information, sighted people will struggle to provide clear answers. On the other hand, there are questions that are only trivial to sighted volunteers (e.g. *"How is this page organized"*, *"What is this button?"*).

5.1.9 Discussion

Participants showed interest in using Hint Me! to learn at their own pace, thus removing the need to rely on the availability of others. Still, they expressed concerns on how to ensure the quality of the answers and their own privacy. They saw different usages for questioning and browsing, from addressing an issue to finding features or workarounds they were unaware of. The following implications derived from their desires and concerns, and the insights on the volunteer role.

Enable Self-Organized Learning. Facilitating smartphone usage is not just about overcoming challenges; it is also about promoting serendipitous discovery of new features. Assistance should allow users to have control on content consumption in order to learn at their own pace.

Support the Workforce. For a successful assistance, the human supporter must be provided with enough information to become domain competent and aware of the communication needs of the end-user. As such, we should compensate the potential mismatch between the user's experience and the volunteers, augmenting their understanding of the user's context and doubts (e.g. leveraging DOM trees to portrait the information available to the screen reader).

Gather Knowledge. Technologies that rely on human input should not waste contributions in single use, but instead iteratively build a shared knowledge base. Moreover, we must look for opportunities to pre-emptively generate knowledge (e.g. describe layout structure) enabling better coverage and availability.

Nurture Knowledge. The variety of mobile devices, applications versions, and frequent updates demands for continuous re-assessment of the gathered knowledge validity. Applications and services are constantly morphing, if assistive services do not take countermeasures, the knowledge will quickly become outdated and possibly prejudicial to users seeking guidance.

Respect Privacy. Smartphones are inherently private and hold personal data. Human-powered approaches must provide users with control over what they share and with whom, awareness of what is being shared, and selection of supporter-group based on information sensitivity. Alternatively, we should find novel ways to take advantage of context by removing all private and identifiable information.

5.1.10 Limitations

With the guarantee of the quality of the answers, we were able to understand the potential of the approach. However, it limited our understanding of the issues the users face with answers of variable quality. Moreover, we relied on a screenshot of the user interface to provide context, restricting the volunteer role to sighted people.

5.2 User Study: Authoring and Playthrough of Nonvisual Smartphone Tutorials

In our previous study, we rallied for support tools that promote self-organized learning. Moreover, we unveiled some of the challenges present in relying on sighted people, with no accessibility knowledge, to support blind people in using their smartphones.

While there seems to be an opportunity to develop tools that enable and foster the provision of assistance by volunteers (i.e. the crowd), there is a need to better understand how those tools can be designed to be effective. In this user study, we examine how blind and sighted people provide instructions in this context, and how they support blind people in performing smartphone tasks (or fail to do so).

We sought to understand how people provide instructions to others, knowing that the end-user is a blind person. In line with capturing the broader set of possible authors, we conducted authoring sessions with two user groups: blind and sighted people. We recruited sighted smartphone users with no prior screen reader experience, and experienced blind smartphone users. We asked participants to create four tutorials for different tasks. Participants were made aware that the intended audience were blind users. Sighted participants were given a set of tips (discussed in a following section) that were discussed with two accessibility instructors. Instructions given by the two groups allowed us to identify the information that we can gather to be leveraged by interactive tutorials. However, we did not know whether the instructions created were enough for people to be able to successfully follow them, and if they were not, what was missing.

In a preliminary study (Rodrigues et al. 2018), we had 11 blind participants following the tutorials created with a playthrough prototype. At every step, participants would hear the instruction followed by the screen reader announcing the target they needed to find and select in that step. However, only 30% of the tutorials were successfully completed; participants struggled to follow instructions and it became clear that having pre-recorded in-context instructions (plus step target) as the sole assistance would not suffice. However, we did not know what was missing for users to be successful.

To understand how to design effective tutorials, in a playthrough session, we again, exposed the content created to a new group of blind participants, and allowed them to ask additional assistance to the researcher. Similar to the question-asking protocol introduced by Kato (Kato 1986), the researcher acted in place of an ideal interactive tutorial. The protocol was designed to have an expert coaching a user with the system. During the process, users could ask questions that would help understands needs in-context, identify information needs, difficulties and how users perceived the system. The approach has been previously identified for its potential in uncovering learnability issues (Grossman, Fitzmaurice, and Attar 2009). We adopted the approach, and in addition, we observed and analysed the interactions between participant, smartphone, tutorial and researcher. We were particularly interested in understanding the limitations of the instructions provided, what were the problems they caused and how one could complement them to enable users to complete the tutorials. Thus, we investigated how to deal with previously identified challenges by understanding the interaction and content needs of end-users.

5.2.1 Nonvisual Tutorials

We chose to explore the authoring and playthrough of nonvisual tutorials, as they enable the continuous support of more complex tasks, assistive content can be created pre-emptively, and possibly (even implicitly) by users' interactions - two key advantages over Q&A services.

It is commonplace for applications and OS to have onboarding tutorials that guide users through their core features, thus supporting initial learnability of the system. Developers typically create tutorials for what they believe to be the most relevant features, which can result in limited coverage of assistance. They also often overlook nonvisual accessibility, relying solely on visual metaphors to guide users (e.g. overlay to obscure content, animation depicting required gestures). Therefore, onboarding tutorials are not always accessible to blind people. Recognizing the existing challenges of mobile nonvisual interaction, there is an opportunity to leverage the benefits of in-context, and always available, help provided by tutorials. Our research explores the feasibility and the requirements for tools that enable the authoring and playthrough of effective nonvisual smartphone assistance at scale. We believe that only then we will be able to support users in both initial and extended learnability, as described in Grossman et. al (Grossman, Fitzmaurice, and Attar 2009).

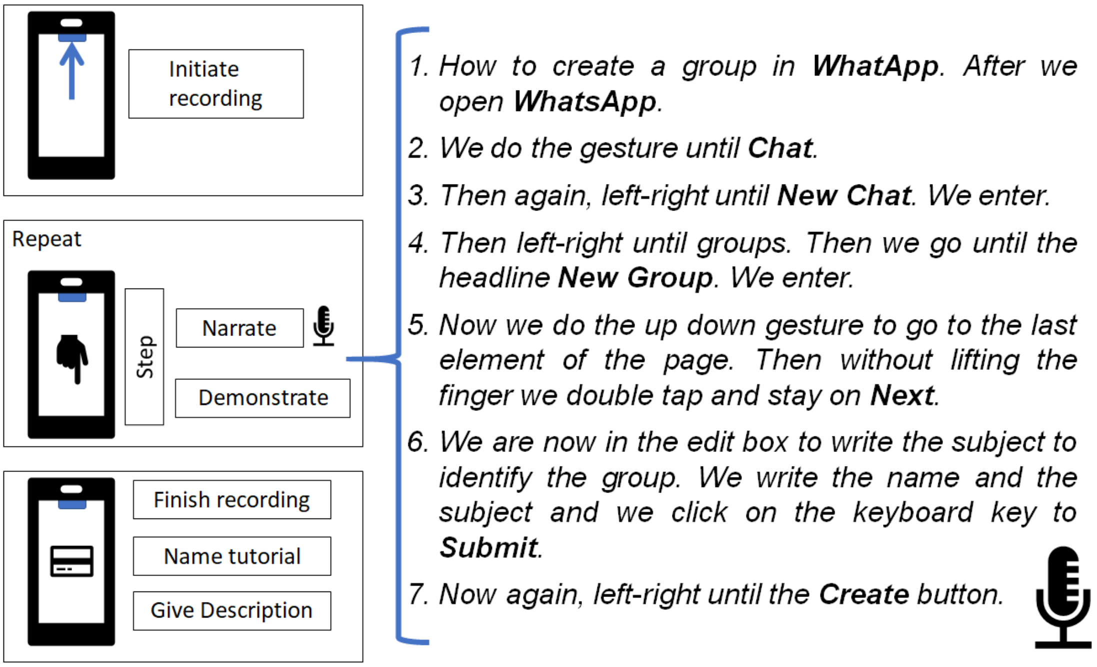


Figure 26 - Example of an authored tutorial

5.2.2 Authoring Tutorials

Users created tutorials while performing the task; first they described a step, then they performed it. Upon finishing, users were asked to name and provide a description for the tutorial. Tutorials were segmented by each selection (e.g. 'Contacts', 'John') and associated with the respective audio snippet; this constitutes what we refer to as a step. An example of a created tutorial is depicted in Figure 26. To record the tutorials, we developed an Android tool that allowed us to audio record authors and detect the steps performed to complete a certain task. The tool was designed to be unobtrusive to user interaction and usable, with and without screen readers. We purposelessly asked participants to demonstrate the task while recording it, as authoring through demonstration can be an effective teaching approach (Wang et al. 2014; Lafreniere, Grossman, and Fitzmaurice 2013).

Table 6 - Task description

ID	Application	Task
TT1	SimpleNote	Delete an existing note
TT2	SimpleNote	Share existing note on WhatsApp
T1	Settings	Clean data from the Calculator app
T2	Messages	Forward an SMS
T3	WhatsApp	Create a group chat
T4	RadioNet	Add a station to favourites

5.2.3 Task Design

Participants created tutorials for six tasks (T) during the authoring session (Table 6). Two were training tasks (TT) and were designed by the research team. To minimize the differences in difficulty between the tutorials, all tasks could be completed with six selections. Three of the tasks were doubts previously asked to members of the research team by blind people. T1 was added by the research team, as an OS task, that could also be completed with six selections.

5.2.4 Tips for Accessible Tutorial Authoring

Sighted participants were informed the tutorials were going to be used by blind people. However, some people are not aware of how screen readers work and go as far as not knowing smartphones can be accessible to blind people. In synchronous assistance, people can ask questions and explain their requirements. On the other hand, for assistance provided through technology (e.g. tutorials, Q&A (Rodrigues, Montague, et al. 2017; Takagi et al. 2008)) there is an opportunity to inform helpers of the user requirements. To this end, we had a session with two blind IT instructors where, together, we devised a description of a screen reader and set of tips to provide to sighted authors:

The tutorials you will be creating today will be used by blind people. Nowadays, smartphones come with screen readers that enable blind people to interact with touchscreens. When active, they change how users interact. Taps now focus elements and read it using text-to-speech technology. Using left or right swipes changes the focus to the next or previous element. To select, instead of a tap, users double tap. When creating a tutorial please remember the following tips:

1) Do not reference visual elements (e.g. tap the green arrow);

2) If possible, indicate the textual description of the elements;

3) Indicate the functionality/purpose of the elements;

4)If possible, indicate the element location;

5.2.5 Participants

For the authoring session, we recruited five sighted participants with no previous screen reader experience. Ages ranged between 19 and 23 (M=20.8, SD=1.64) years old, three Android users and two iOS, experience between 3 and 4 years. We also recruited five blind participants, ages ranging between 25 and 51 (M=38.8, SD=9.49) years old, three iOS users and two Android, experience between 5 and 11 years, and two were IT instructors. Experienced users were chosen because of their knowledge and because people often rely on them to overcome challenges, as reported in *Chapter 4 - Mobile Challenges*. We considered users to be experienced if they had a smartphone for over four years, and were able to perform the following list of tasks: place/receive calls, send/read emails/messages, install new applications, configure accessibility settings, browse the internet, use communication apps (e.g. Messenger, WhatsApp, Skype) and assistive applications (e.g. BeMyEyes). In the following sections we will refer to authoring participants as *Authors*.

For the playthrough session, we recruited 12 blind participants, ages ranging between 29 to 59 (M=49.58, SD=10.36) years old, six Android users and six iOS, experience with smartphones between three months and four years. None of the participants took part in the first session. Participants had a wide range of expertise, with three participants meeting the requirements to be experienced users, and two novice users, that only placed/received calls and received messages. One participant had previously forwarded a message on Android and three others on iOS devices. Additionally, one previously created a WhatsApp group. In the following sections we will refer to playthrough participants as *Consumers*.

5.2.6 Apparatus

We used a device running Android 7.1.2. and Talkback, the default screen reader. In the authoring session, participants were invited to use headphones to prevent recording the screen reader feedback. All applications were made available a priori on the device home screen. For the playthrough session, a laptop computer was used to control the audio instructions given to the participants during the tasks. We controlled for concurrent feedback, only providing the next instruction when the screen reader was silent, pausing/starting when needed.

5.2.7 Procedure

In both sessions, participants were informed the purpose of the study was to understand how interactive tutorials might facilitate smartphone use. Then, participants completed a brief demographics and smartphone usage questionnaire.

Authoring Session

Authors were recruited in advance and given the list of tasks at least one day before meeting with the research team. They were asked to become acquainted with the applications and the tasks if they were not already. Participants were tasked with creating six tutorials. Sighted users were also presented with the introduction and set of aforementioned tips. Prior to creating a tutorial for each task, users were instructed to explore and perform the task. For the first training tutorial (TT1), participants were guided through the creation process. Participants were informed that each step should start with an explanation of the step followed by its demonstration. Participants were then asked to create a tutorial for TT2. All participants successfully created a tutorial, thus completing the training phase. The order of the remaining four tasks was randomly chosen. Participants started every recording from the home screen. Although every task could be completed with six selections, participants were free to take alternative paths. The study concluded with a debriefing questionnaire to assess the user's opinions about the authoring process.

Pre-Processing Content

For the playthrough session, we discarded four tutorials for having missteps (i.e. an incorrect step followed by a "back" action), one for having stereotypical references to difficulties felt by blind people, and three for poor audio quality. When recording tutorials, users had to demonstrate the task while giving instructions which resulted in audio files with long periods of silence. To address this issue, we removed the silences of all audio recordings. We intentionally did not control tutorial delivery, precision of vocabulary or required level of skill. In this study, our goal was to assess how to go from human generated tutorials, with all their idiosyncrasies, to accessible tutorials.

Playthrough Session

First, *Consumers* were informed they would be asked to complete a set of tasks. During the tasks they would be following instructions that had been previously recorded by other people, both sighted and blind. At any point during the task, when participants wanted to control the playthrough of the instructions (e.g. stop, play, repeat) or when they required additional information or assistance, they could prompt the researcher. When a clear question was asked, the researcher would answer it (e.g. *"Where is it?"*). When participants asked for assistance but could not verbalize what they needed (e.g. *"I cannot find it anywhere. What should I do?"*) the researcher would help them based on what he observed caused the issue (e.g. "*You already went through the target, but it is not 'create' it is 'new conversation*'").

Instructions were given step by step or whenever the participant asked. To avoid audio conflicts, instructions only started when the screen reader had nothing else to announce. Participants were asked to complete the task by following the instructions and encouraged to think aloud whenever they stopped to require assistance. The only limitations imposed were: (1) the researcher could not physically assist in any way, and (2) the researcher could not take the initiative to provide further instructions unless the participant was stuck in a step for more than three minutes. We audio recorded the entire session and observed user interactions with the smartphone. A second researcher annotated all requests and assistance provided by the intervening researcher.

For *Consumers* to get accustomed to the device and the study procedure, they completed TT1 created by the research team. Once they completed the task, and felt comfortable navigating the device, we asked *Consumers* to complete the four tasks. Prior to starting each, they were informed what they would be attempting to do by following the instructions (e.g., ‘Creating a group chat in WhatsApp’). Each participant followed tutorials created by both groups. Order of the tasks was counterbalanced, and every validated tutorial was followed at least once. In the debriefing session, participants discussed the challenges faced and provided insights on possible features. Finally, we asked *Consumers* what information they believed to be essential to an instruction.

Table 7 - Code frequency in the tutorials instructions.

Codes (%)	Type					Screen		
	Visual	Text	Incorrect Text	Function	Type	Function Descr.	Layout Descr.	
Blind	0.0	77.8	0.0	9.4	7.7	8.5	2.6	
Sighted	8.6	74.3	19.2	21.0	8.6	2.9	1.9	
Codes (%)	Action			Location			Feedback	
	Selection	Navigation	Gesture	Absolute	Relative	Hierarchical	Audio	State
Blind	59.0	33.3	52.1	7.7	2.6	1.7	7.7	22.2
Sighted	61.9	18.1	1.0	19.0	13.3	11.4	1.0	21.0

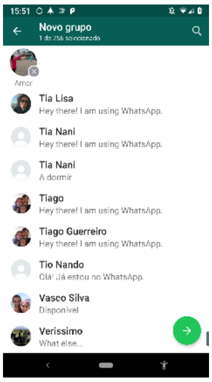
5.2.8 Data & Analysis

We conducted a thematic analysis, leveraging the flexibility of the method to reflect over the data collected (Braun and Clarke 2006). We transcribed the instructions provided by *Authors* while creating the tutorials. For the authoring session, we sought to understand the characteristics of each instruction. Therefore, two researchers inductively created two codebooks from ten tutorials, one from each participant and at least one per task. Codebooks were iterated, discussed among the research team, and merged. Another set of ten tutorials were coded independently and researchers reached a Cohen's Kappa agreement of k=0.82. The final codebook is shown on Table 7, aggregated by theme.

During the playthrough session, a researcher was observing interactions, behaviours, annotating requests, their motivation, and the additional assistance provided. Thus, given the different focus of the second session, we created a second codebook from all the information collected. The codebook was iterated and refined by the two researchers. We aggregated the observations, requests and motivations in four major categories: Instructions' Content; Gesture & Navigation; Location & Layout; and Feedback.

5.2.9 Findings

In the authoring session, we collected 40 tutorials, 20 from each group, with a total of 128 individual instructions recorded by sighted *Authors* and 128 by blind *Authors* (summing a total of 256). Three tutorials created by blind people included extra steps during the recording (e.g. enabling Bluetooth). One blind participant, in one task, only demonstrated



- *Now we do the up down gesture to go to the last element of the page. Now we are on Next we double tap. (BA)*
- *We click forward. (SA)*

- *Now it is showing a list of radio stations, we are going to swipe from left to right with one finger until Radio Comercial. It will say radio logo, radio and name. (…) (BA)*
- *Multiple stations appear, we choose the one we want, Radio Comercial. (SA)*

Figure 27- Two examples of two steps described by four different authors. Two blind authors (BA) and two sighted (SA).

the steps without giving any instructions. The remaining tutorials were created successfully.

In the playthrough session, the twelve participants explored all tasks successfully by following the tutorials and relying on the assistance of the researcher. In total, participants followed 240 steps and requested additional assistance in all tutorials. Specifically, in 83 (34.6%) of the steps, participants requested information that was not present in the instruction given.

Below, we detail our findings, organized into the four major themes that emerged from the playthrough session: Instructions' Content, Gesture & Navigation, Location & Layout and Tutorial Feedback. The discussion on each topic is also supported by the analysis conducted on the tutorial instructions (Table 7). Frequencies are used to illustrate the findings; however, they should not be taken as quantitative measures of the relevance of each problem. Finally, we report on the participants' feedback about the tutorial authoring process and on the value of the tips provided.

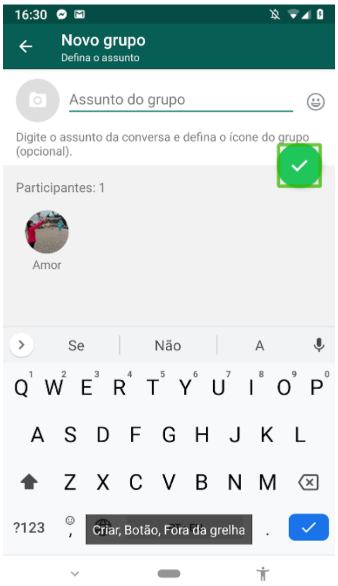


Figure 28 - WhatsApp create group screen.

Instructions' Content

In 51 (21.3%) of the instructional steps provided, *Consumers* could not understand or identify the content being described in relation to the current screen.

Textual descriptions were not always provided. In most steps (above 75%), the instructions had the target textual description.

"In the main menu click the app RadiosNet", S1

For some of the remaining instructions, *Authors* gave less detailed information focusing on the tutorial goal:

"We want a group conversation with one of the contacts, after you select a contact (...)." S1

"(...)*until we find the intended message*", *B4*

At times, sighted *Authors* were unable to provide a target description, leading to long and possibly confusing instructions. One example from T2 (WhatsApp): sighted *Authors* did not know what to call the create button (Figure 28), a green check mark, and gave a long confusing instruction:

"After you select the subject, it will appear on the bottom of the screen and then click. Click not on the upper right corner but a little bit below, but still in the far-right side of the screen and click" S1

Blind *Authors* were more verbose. Sighted *Authors* provided shorter instructions only indicating what to do in each step (Figure 27). Blind *Authors* provided additional information about the current state of the tutorial and its overall goal (33% and 9% of instructions, respectively) (Figure 27). Despite being more verbose, only 8% of blind people's' instructions referenced any kind of audio feedback, and none described any type of audio cue.

"It will say in all of them the 'radio logo', 'radio' followed by the name." B1

Target Description. Although most instructions had text descriptions, 19% of the ones provided by sighted people did not match the item label. This was expected, given the known issues with the variability of vocabulary used by people when interacting with systems (Furnas et al. 1987). Not surprisingly, on the debriefing questionnaire, all sighted participants mention how hard it was to translate a visual icon into a textual description. Therefore, at times (21% of the instructions) they relied on describing the target function rather than its name.

"In the bottom right corner, look for the icon that starts a conversation." S5

For Blind *Authors*, target function (in 9% of the instructions) appears to be described to alert users about the outcome of their interactions.

"Now we get to Radio Channel and we are going to make it play" B3

When following instructions, if the target was anything but verbatim (e.g. "New Conversation" vs "Create Conversation") *Consumers* assumed there would be another option that they had yet to find that corresponded exactly. This is particularly relevant in the first utterances of the word, which are relied on to quickly skim through content.

Gesture & Navigation

In 32 (13.3%) of steps, *Consumers* required additional assistance due to issues with gestures and navigation. This includes issues that resulted from a combination of the navigational content of the instructions and the participants navigational behaviours.

Blind *Authors* instructed more often to navigate. For sighted *Authors*, mentioning navigation is only relevant when the target element is not visually available on the screen. However, for blind people that rely on swipe gestures to navigate, every target needs to be navigated to. Sighted *Authors* instructed users to navigate in 18% of their instructions, while blind *Authors* did it in 33% of the instructions.

"We are going to look for the message by swiping with one finger until we find the message that in this case says 'Hello'", B2

Blind *Authors* were more aware of gesture subtleties. Only one instruction by a sighted Author contained a brief explanation on how to perform a particular action. On the other hand, 52% of blind authored instructions contained the additional information on how to perform a gesture. However, how to instruct the user raised some questions.

"I am going to keep it pressed", S2

"Now, we locate, or we swipe from left to right or by exploring the screen until we find the message we want to forward. Then we double tap and hold on the second one", B1

Conflicts with user expertise and interaction preferences. Instructions that guide the user by saying 'go to the right corner and select X', or 'swipe until you find X' can be disruptive for users who are only familiar with one interaction method (i.e. Explore by Touch versus Sequential Navigation). In multiple instances, *Consumers* tried using an unfamiliar method with no success. They were convinced that since the tutorial instructed them to do so, it was the only way to reach the target. Thus, participants who relied on either method exclusively, had to request help to understand how to proceed.

The same problem happens when sighted *Authors* instruct participants to perform gestures (e.g. to forward a message with a screen reader, users have to double tap with a long press on the second tap; instead of just a long press). Since neither sighted Author nor *Consumer* were aware of the dissonance between the interactions, the latter required further assistance.

Navigational deadlocks. Although we observed that instructions that include how to reach a target can be problematic, the exact opposite can also be true. For novice users that still struggle understanding some navigational behaviours, such as lists, information on how to reach a target can be fundamental. To reach the option "Applications" in the device "Settings", users have to either perform a scroll, or navigate by swiping from left to right until they reach the end of the list displayed on the screen. However, if the user is relying on navigating from right to left, the list will not scroll down, it will just cycle through the elements on the screen repeatedly. Thus, the user will never find the intended target, leading to a navigational deadlock. When multiple lists are present on the screen, the user can get "locked" navigating one until it reaches the end of its content, which in auto updated views can be never.

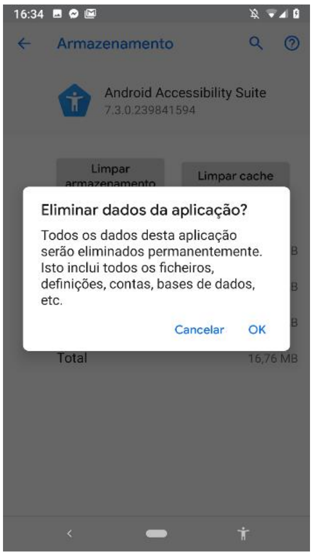


Figure 29 - Clear app data confirmation pop-up

Location & Layout

In 38 (15.8%) of steps, *Consumers* required assistance related with the location of the target element or further details about the overall layout of the screen.

Sighted *Authors* gave more, and often useless, location instructions. Blind *Authors* gave location instructions (absolute, relative or hierarchical) 15 times while sighted *Authors* did it 46 times. Although 42% (19% absolute, 13% relative, and 11% hierarchical) of instructions by sighted *Authors* came with location, many of them were inadequate and even misleading. In the following example, while location was provided, the *Consumer* was unaware of the location and size of the pop-up menu (Figure 29).

"(...) click in the OK that is on the bottom right corner of the pop-up", S4

Target location was complementary. *Consumers* wanted to be notified about the absolute locations of the target they needed to reach. Since some relied on ‘Explore by Touch’, absolute location could be crucial to find the target effectively. Others asked for location instructions when they got stuck in navigational deadlocks. A few that relied on mixed interaction methods wanted to optimize their navigation behaviours. To do so, they needed to know the target whereabouts to be able to start their navigation closer to the intended target, prior to linear scanning. Location seems to be complementary, and when given, one must be aware of its potential consequences. It all depends on the user expertise and interaction behaviours.

Describe the screen overall layout. For multiple *Consumers* it was important to create a mental model of the screen before starting to navigate. However, less than 10% of the instructions contained additional information about other functionalities available in the screen. Sighted *Authors* made no attempts to describe layout, despite being able to quickly grasp a screen structure. From the 128 instructions given by sighted *Authors* only two attempted to describe screen layout. In contrast, for blind *Authors* to describe a layout, they must first explore all the interface. Even so, only three of the 128 instructions by blind *Authors* contained layout descriptions.

Consumers asked how the content was organized as they tried to figure out how the elements were disposed *("Is it a grid?")*. In some cases where the screen was composed by two or three major structures (e.g. title bar and list), the answer was simple. However, there are complex layouts that can be time consuming to describe, and at times even confusing (e.g. multiple list views, some horizontal others vertical, with other unstructured content). In these instances, the researcher providing additional assistance struggled to provide a clear and concise description of the layout.

Tutorial Feedback

In this section, we aggregated instances where feedback should have been provided to facilitate user interaction at a key point during navigation, or when feedback was provided inappropriately, causing users to request assistance.

Confirm target. Similarly, to what has been previously reported in Vigo et al. (Vigo e Harper 2013) on the coping tactics employed by visually impaired people on the web, *Consumers* asked for reassurance and confirmation: checking what to do prior to engaging in a navigation and confirming once they reached a target. For a successful interactive tutorial, one may need to ensure users are given appropriate feedback to enable them to seamless detect they have reached the target and reassure them that they are following the intended steps.

***Consumers* did not understand why they could not find their target.** When *Consumers* spent a significant amount of time exploring the screen and detected repetitions without finding the target element, they prompted the researcher for assistance. This could be because the element was not displayed on the screen; or *Consumers* were stuck in a navigational deadlock; or because during exploration they missed the target element. Although the consequences are the same, the required actions to address them are distinct. Thus, it can be crucial to detect each scenario.

***Consumers* were unaware of incorrect steps.** In 21 (8.8%) of the steps *Consumers* deviated from the intended path. In all, *Consumers* were unaware they did so. In two instances the step was a shortcut that jumped the tutorial two steps forward. In these instances, the researcher controlling the audio tutorial compensated and skipped the middle step. In the remaining steps, participants were notified they had deviated from the intended path after they asked the researcher for further instructions; it is noteworthy that all requested assistance to resume their previous state.

Authoring and Tips

Sighted *Authors* at times did not follow the provided tips, and struggled to provide descriptions to visual elements, sometimes even to provide location.

> *"I will press again (hadn't mentioned or previously pressed that button) in the button on the right line below, in the bottom right corner.", S2.*

5.2.10 Discussion

We explored the ways in which people create tutorials for mobile interactions, and the challenges faced by blind people when following those tutorials. We discuss the lessons learned that inform our future solutions, and which should be of interest to researchers and practitioners working on nonvisual mobile accessibility.

Required information & feedback

Different people will require different instructions and control depending on a variety of factors. The only common requirement for all instructions and across participants was accurate target description. All other information can be beneficial or detrimental to the users. The types of information/feedback required were the following: target location;

target function; screen state; layout description; target focus confirmation; alert on path deviation; gesture guidance; and task/feature clarifications.

In past work, Lieberman et al. (Lieberman, Rosenzweig, and Fry 2014) have explored three levels of control over each step. However, with an understanding of the information and interaction needs, we can go further and, not only provide different levels of guidance, but also adapt contents within each instruction. It is important to collect different types of information for all tutorial steps during authoring. Only with a rich information set can we develop flexible playthrough tools. This may come with the cost of overburdening authors; thus, we must work towards solutions that support the authoring process.

Authoring support

Blind and sighted people created instructions with different content. Blind people were more verbose and often provided guidance on how to navigate, which again can be beneficial or detrimental.

On the other hand, despite the tips, sighted participants still provided inaccurate instructions, suggesting we may need to find alternatives. Discarding sighted people from the pool of authors is not one we should willingly follow due to its impact on availability and coverability of tutorials, particularly when we consider that some of the information required during playthrough is easier to be provided by sighted people. Sighted people, in previous work, have been successfully leveraged to answer visual questions (Bigham et al. 2010). Future solutions should be able to leverage the differences in content, created by both author groups, to provide accessible tutorials.

One possibility is to increase the authoring burden by increasing training and provide more context, as we suggested in the previous study. However, given the issues found despite the tips provided, a more scalable approach may be collecting additional data during the authoring process (e.g. layouts, steps, workflows, labels) to reduce the dependency on the accessibility knowledge of tutorial authors.

To collect all the different information required, we can explore how to break down the authoring process, and prompt people (authors) to provide different types of information in small tasks (e.g. Item location, Layout Structure), without any training or understanding of the underlying requirements. This can be compared to what has been previously proposed and achieved by Gleason et. al (Gleason et al. 2018) in enabling non-experts to participate in the installation and maintenance of indoor localization infrastructures.

This approach would enable us to both guide contributors through the authoring process and, if need be, rely on different contributors for different types of information. Furthermore, we can explore how to make the most of contributions by maximizing the information collected and/or derived automatically. For instance, since we can ask contributors to demonstrate tasks, we use the opportunity to collect target descriptions, thus avoiding inaccurate descriptions. By breaking down the authoring process, we can combine contributions of multiple people to create a complete representation of the task and all its peculiarities. In addition to relying on authors as the sole contributors, in line with previous work (Lafreniere, Grossman, and Fitzmaurice 2013), *Consumers*' interactions with content can be leveraged to enhance assistance (e.g. providing multiple demonstrations of the sequence of navigational steps taken).

Only by supporting the authoring process and leveraging multiple sources of information will we be able to design adaptable playthrough solutions.

Account for user expertise, behaviours and preferences

The same instructions can be interpreted differently by users, and what prevents some from completing the task can be what enables others to do so. For example, for users who primarily rely on Explore by Touch, it is of the utmost importance to understand the interface underneath. If users understand they are interacting with a grid, they will scan very differently than if they believe they are facing a list. We must also be aware that user requirements might change per step, or even in the same step, when certain navigation patterns occur (e.g. detecting users are stuck navigating an auto-updating list and their target is not on that list but in another element of the layout).

We believe part of the solution can be to continuously model and monitor user interaction behaviours during playthrough. Previous research in user modelling (Montague 2012) has already explored continuously updating models based on current behaviours, to provide optimal settings for each interface. We can imagine a similar approach to tutorial playthrough, where one can adapt based on: user profile; interaction behaviours (i.e. past and current); navigation pattern detection; content instructions and personalization.

Flexible instructions & App Modelling

If we can collect different types of information and develop solutions that take into consideration user interaction behaviours and immediate needs, we will be able to provide flexible instructions. As recognized by Lieberman et al. (Lieberman, Rosenzweig, and Fry 2014), users may require different levels of guidance at different steps. By default, users should be able to access all types of instructions during playthrough by request or based on triggering interactions.

We may start to adapt instructions verbosity and gestures guidance. Expert users felt instructions were too long with unnecessary content. However, for less experienced users, detailed instructions may be crucial since they are not as aware of the navigation nuances of different interface elements. One example is providing users with additional information on navigational locks, or if the target element is, or ever was, on screen.

When possible, instructions should be generated or adapted to current context and past actions. Interactive tutorials should detect the variety of available paths to complete a task, alerting users on deviations and providing mechanisms to recover. Building such systems will require a deeper understanding of app structure and navigation workflows, currently out of reach to third party assistive technologies. However, if we can model application structures and workflows, we will be able to create smarter assistive tools.

Enable dialog, a fallback mechanism.

The previous considerations stem from the unpredictability of the user’s individual requirements when trying to learn or accomplish a task. We discuss how we can broaden the adaptability of instructions and assistance by considering the variety of points of failure and doubts, and pre-emptively prepare for them. As our findings suggest, invariable instructions were not enough. With adapted solutions, we might get closer to a fully automatic assistance. However, we believe the only answer to byzantine problems is to rely on others once more. We can create solutions that leverage others beyond a single contribution and enable a dialog mediated by the technology. The outcomes of this channel will further fuel the accessibility and adaptability of the content, thus creating living artefacts capable of perpetual improvement.

Limitations

We conducted a study with five blind and five sighted authors that created tutorials for four tasks, that we exposed to 12 blind participants. Although this is a small number of participants and tasks per user group, it allowed us to identify a variety of novel information needs triggered by nonvisual tutorials. Nonetheless, further research with a larger user pool, with different expertise levels, and set of tasks (e.g. navigate a video, play a game) may uncover additional needs.

5.3 Summary

In this chapter, we first explored the acceptance of human-powered assistance for smartphone support. When the support is provided by a domain expert with full context it can be effective, and users perceived it as not relying on anyone else, enabling self-organized learning and promoting self-efficacy.

Next, we conducted a user study exploring the authoring and playthrough of nonvisual smartphone tutorials. We identified the different information required by users during playthrough when following instructions by others. It was clear that the instructions created were not enough. All participants required additional assistance that was not contemplated in the instructions. At times, they could even be the cause of the issues.

When following a tutorial, the differences in users' expertise, interaction behaviours and preferences dictate the type of instruction adequate for each user. There is a need for novel solutions in interactive nonvisual tutorials, both in authoring and playthrough.

There is an opportunity to explore how to support users during authoring to create useful information, taking advantage of each author specific knowledge. Moreover, we can start leveraging the data collected during the authoring process to enrich, or even create, new instructions. On the other end of the spectrum, we need to compensate for the unavoidable flaws that come from: 1) the authoring process by non-experts; and 2) the limitations of rigid instructions by looking into novel playthrough mechanisms for nonvisual assistance.

Chapter 6

Designing a Rich Smartphone Assistant

With an understanding of the challenges blind people face, and with the knowledge of what and how assistance should be provided, we found ourselves in the position to posit how an effective human-powered solution could be designed.

We believe assistance should be more than a set of static instructions (i.e. the typical manual or help menu) and should be able to support the user in the tasks they are trying to accomplish. It should not just be a fallback mechanism, that relies on users being able to express their doubts. It should also avoid being the exact opposite: a proactive, intrusive assistance. The popular example is 'Clippy' by Microsoft, which was largely unsuccessful due to a low accuracy, and because it provided no provenance to its suggestions. A successful assistant must be both competent and trusted (Maes 1995).

We propose the creation of a Rich Interactive Smartphone Assistant, a support mechanism that can adapt accordingly to user interactions and context, is available everywhere, and is prepared to generate assistive content based on users' requests. To deal with the idiosyncrasies of human behaviour and requests, such system is therefore required to be highly flexible, and capable of an understanding, beyond current automatic approaches. Thus, at this point, a rich interactive assistant must, in some capacity, rely on human contributors to generate its assistive content.

In this chapter, we posit how a Rich Interaction Assistant should be designed based on our previous studies, and describe our implementation of RISA (**R**ich **I**nteractive **S**martphone **A**ssistance). We describe the design decisions behind RISA features, and detail the technical challenges we had to overcome to provide in-context, system-wide support. We contribute with **C6) a human-powered nonvisual task assistance that relies on non-expert content authors**.

6.1 Rich Smartphone Assistance

In the proposed solution there are two distinct but equally important stakeholders: people that contribute with content (*Authors*), and end-users who will leverage the assistive content created (*Consumers*). We envision a system that requires minimal effort from *Authors* to contribute, and relies as much as possible on automatic mechanisms, to enrich and adapt assistance. For *Consumers*, assistance should be pervasive (i.e. available across the whole device in any app), congruent (i.e. always provide the same level of support and hide the multiplicity of possible sources of assistance) and capable of supporting any task *Consumers* wish to learn.

To accomplish the vision and based on our previous findings, we abided by the following principles. **1) not require any training or professional assistance**. *Authors* can be hard to come by, and it quickly becomes an uphill battle if we start requiring training. A service provided by professionals would entail significant costs to users, or entities that provided the service. To rely solely on amateurs, we need to **2) streamline the authoring process.** In the previous chapters, we identified what information was required to build an assistant that would be able to guide through complex tasks, and found that content by non-experts would not suffice. Thus, a rich smartphone assistant should rely as little as possible on *Authors*, relying whenever possible on automatic data collection to reduce the burden on the authoring process; and facilitate the **3) consistency of delivery** and **4) flexibility of assistance**. To support flexibility, we need to design solutions that monitor user interactions, and have basic knowledge of the app structure to be able to provide **5) path recovery support.** As we have previously discussed (*Chapter 3 - Smartphone Adoption*), we need to ensure there is always a human **6) fallback mechanism**. Finally, we must **7) aggregate knowledge**. When knowledge is created to assist, we should not waste it in a single instance for a single Consumer. Contributions are a precious commodity, and we must take full advantage of their potential to avoid overburdening *Authors* and guarantee, with each subsequent contribution, a larger coverage of the assistant required.

6.2 RISA

Following the aforementioned principles, we set out to develop RISA, a nonvisual smartphone task assistant. RISA was designed to rely on *Authors* to demonstrate the task. Then, from the data collected it generates a task assistant.

> *Mary receives a notification that there is an open request asking how to share a previously seen YouTube video on Facebook. She is used to do it, so she takes on the request. RISA asks Mary to demonstrate the task, adding additional information at certain points, such as interface description and other information about what she is selecting. Mary concludes the demonstration giving it a title and description. RISA, with the information collected implicitly, and explicitly, during the demonstration, creates a task assistant, that becomes available to anyone in need.*

Consumers can then rely on RISA to learn how to perform the task. When learning a task, RISA guides C*onsumers* while monitoring their interactions and the app interface.

> *John wants to learn how to share a YouTube video he has seen recently on Facebook. He goes to RISA and checks if there is a task assistant that will guide him through the task. If there isn't, he can always create a new request and describe his task request. Since someone else had asked for the same thing before, John selects the task assistance. Step by step, RISA will announce where John should go, giving him auditory cues and, whenever requested, additional hints about the interface or target option. RISA adapts its guidance and hints accordingly to the smartphone context, and John's past interactions. John can freely explore and follow the steps at his own pace. Using RISA, John is able to learn and share the video with his friends. The day after, John rushed to teach his mother how to perform the task, knowing that she also struggles with it.*

During the development of RISA, we also conducted two informal sessions with each stakeholder (i.e. *author* and *consumer*) where they freely explored RISA and provided some additional feedback.

Below, we discuss in detail the implementation of RISA. First, we provide an overview of the system. Next, we describe in detail the authoring procedure, the technical challenges RISA overcomes, the data collected, and other features relevant to *Authors.* Next, we cover how RISA was developed to support *Consumers* when providing task assistance. We describe the Task Assistance procedure, all its features and behaviours, and how technical challenges were addressed.

6.2.1 System Overview

RISA is an Android application and an accessibility service. It uses Firebase mobile development platform for its authentication, real-time database, cloud-storage, crash-analytics, cloud-functions and cloud-messaging. It was developed using Android Studio and targets Android SDK API 28 (9.0) with a minimum SDK 26 (i.e. 8.0). It uses an adapted version of the external library ChatKit[20] for its chat interface. Below, we briefly describe each of the components.

Application

When first installing RISA, the user will have the choice of becoming a contributing *Author*, or solely use RISA as a task assistant (i.e. *Consumer*). Users are required to register using a google account. Depending on the decision, users will be presented with a different set of app features. The application is responsible for requesting all the required permissions, allow to browse the different available assistive content, request new content, and notify users on changes. *Authors* have access to all RISA's features, while *Consumers* have a simplified interface. However, only users who have Talkback enabled can fully leverage RISA's task assistant capabilities, due to some being highly coupled with screen reader behaviours.

Accessibility Service

An accessibility service in Android is a background service that receives callbacks from the system when *AccessibilityEvents* are fired. These events denote transitions in the user interface (e.g. focus changes, clicked element). Through such service, it is also possible to query the content of the current active window, even outside one's own app. The RISA

[20] ChatKit, (https://github.com/stfalcon-studio/ChatKit)

accessibility service is responsible for the authoring process and for the task assistant. During the authoring process, we use the service capabilities to collect:

1. Current application package name;
2. Windows' list of interactive elements (e.g. can be clicked, long clicked);
3. All scrolls performed and in which elements;
4. All text written and in which elements;
5. Interaction events (i.e. click, long) and elements that triggered a transition;

When the service is active in task assistant mode, we use it to control the flow of the task assistant, and adapt guidance depending on active window and users' behaviours. The service was designed to:

1. Monitor active window elements;
2. Detect intended target based on the task at hand;
3. Monitor user interactions;

As an Accessibility Service, RISA faces the same implementation restrictions as any other assistive technology developed for Android devices (e.g. Talkback). For example, most games are designed in a way that don't trigger Accessibility Events, and their window content is not *queryable*. Therefore, they are not accessible to screen readers or any other accessibility services like RISA.

Firebase Mobile Development Platform

RISA relies on *Firebase Authentication* to uniquely identify its users, and adapt the content presented based on the type of user, and content ownership. The *Real-Time* database is used to store all the information collected during the authoring process, chats about the assistive content, requests, and user profiles (i.e. installed applications, type of user). During the authoring procedure, any audio recordings are stored using the *Cloud-Storage*. The *Cloud-Functions & Messaging* features are used to notify *Authors* and *Consumers* of new requests and answered ones. *Crash-Analytics* was used to facilitate the debugging process across multiple devices.

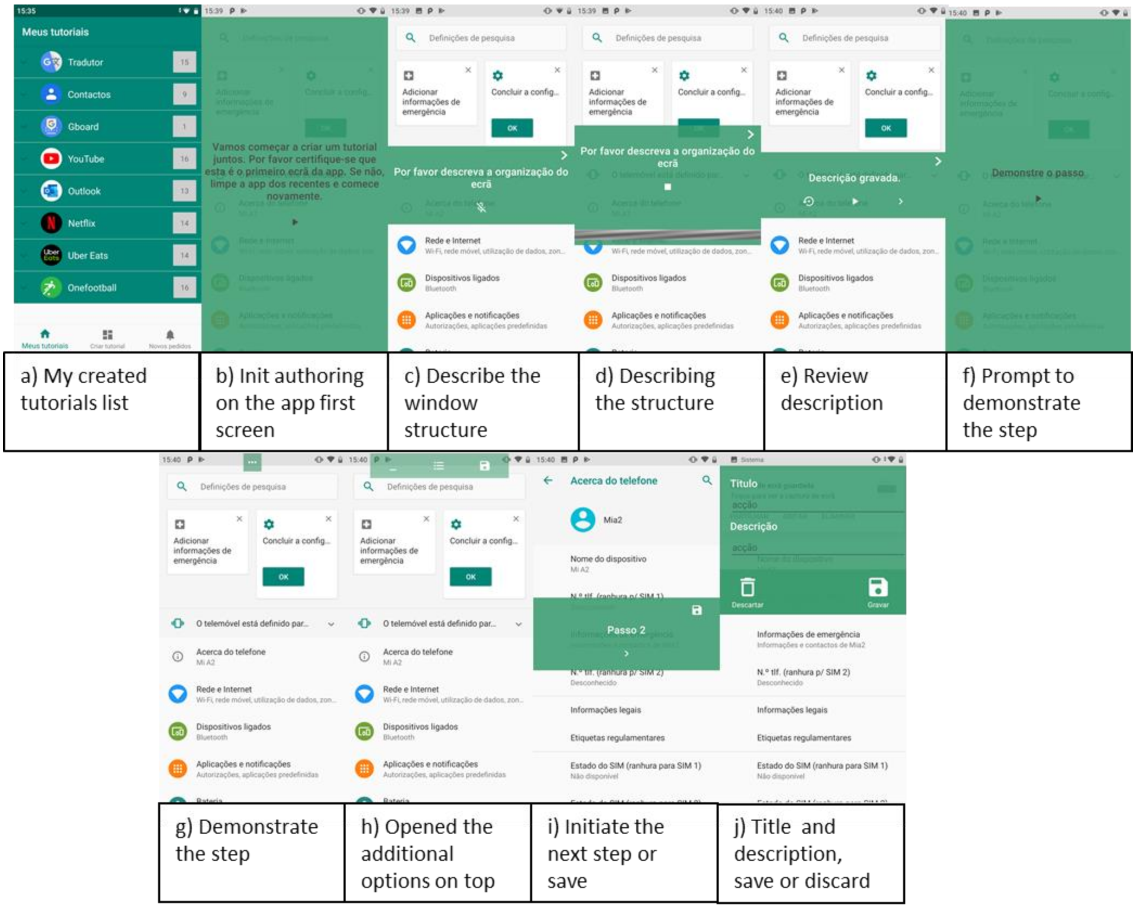


Figure 30 - Example of authoring a task with RISA.

6.2.2 Authoring

In RISA, *Authors* are not required to have any a priori knowledge about the tool, and are guided, step by step, in the creation of assistive content. *Authors* simply must know how to perform the task. All other information is prompted when necessary. The authoring procedure of RISA is depicted in Figure 30. *Authors* start by either selecting an app, or a request previously made by *Consumers*. Then, in each screen, *Authors* are asked to describe the structure of the active window (Figure 30 – c/d/e) via voice input; *Authors* can skip this step if the last window described follows the same structure (i.e. when *Authors* perform interactions that do not change the structure of the active screen - 'like' button). After reviewing the recording, *Authors* are asked to demonstrate the step (i.e. perform the interaction). For most steps, this covers the whole procedure.

In steps where text is written, or possibly sensitive, or dynamic information is detected, users are presented with two prompts: one to ask if the information is sensitive/private/dynamic, and one about the type of content that was inserted/selected. Both impact how assistance is provided. We discuss, in the following sections, how sensitive or dynamic information is detected.

Authors will repeat this sequence until they have fully demonstrated the task. At any point, they can consult more options to check the previous steps or save the task. To finish, *Authors* are required to add a title and a description to what they just demonstrated.

Assuming the worst-case scenario the following will occur: *Authors* will be prompted in every step, will have to write a word or two, and provide an audio description. If we assume the average user will write on average 19 words per minute (MacKenzie and Zhang 2001) on smartphones, then demonstrating the task for RISA will take: twice as long (i.e. an extra prompt for each interaction), plus a minute for each 19 words written, plus the time it takes to quickly describe the layout of an interface, which is highly user

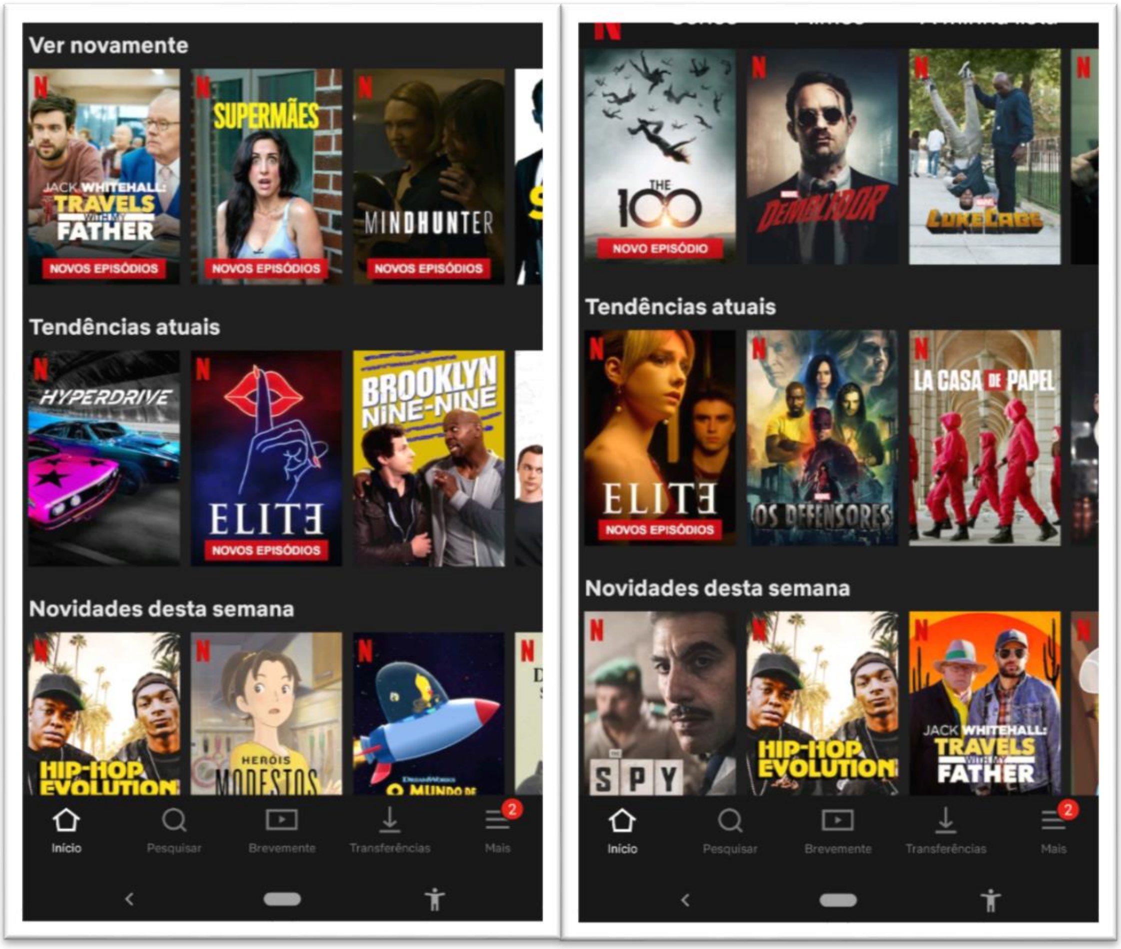


Figure 31 - Neflix with a different feed at a different point in time in the same week.

dependent. We can conservatively estimate that authoring a task for RISA will take, at least, twice as long. The time taken will be highly variable, and highly coupled with tasks and *Author* profile.

Dynamic Content

Many, if not most, smartphones apps have some sort of content updates, either due to internet connectivity or because of users' interactions. App interactive elements are highly volatile, particularly their text content. They change due to user profiles, time of day, events, etc (Figure 31). Currently, we do not possess the ability to uniquely identify content, nor the ability to programmatically know when an element is static (i.e. always has the same characteristics including text) or dynamic (i.e. changes at least text content). To be able to create a task assistance that can handle not being presented with the exact same content, we also require *Authors* to tag selected elements as dynamic. When an interactive element is selected, that is not part of a whitelist created of common terms for interactive elements, *Authors* are asked if the element is dynamic; if it is, they are prompted to write a text description of the content. This is a rather naive approach that could be improved in the future, by either complementing the whitelist with validated data or other approaches to detect dynamic data in the app structure.

Sensitive Content

Interactive elements can have sensitive information (e.g. a message that you have sent can be part of a demonstration on how to forward a message; an account balance can be pressed for details). To streamline the authoring procedure, RISA only asks if it is sensitive information if: the *Author* wrote; it is not one of the whitelist labels and the text detected in the interactive element is composed by more than one word; or if it is a number. The process ensures *Authors* can create assistive content for any task without having to compromise their privacy or security.

Table 8 - Data collected by RISA during the authoring process.

Task Demonstration	Screen	MyView
author: <String>	title: <String>	activity: <String>
title: <String>	packageName: <String>	classname: <String>
description: <String>	Timestamp: <Long>	boundsInScreen: <Rect>
packageName: <String>	interactionType: {CLICK, LONG_CLICK}	clickable: <bool>
Nsteps: : <Int>	audioDescription: : <String>	scrollable: <bool>
Screens: [Screen]	clicked: View	closestText: <String>
	interactive: [MyView]	contentDescription: <String>
	scrolled: [MyView]	text: <String>
	textEdit: [MyView]	eventText: <String>
		authorText: <String>
		(…)

Data Collected

We rely on the authoring process to collect a mix of metadata and interaction data. We record every detail we can from each screen, in each step, as well as all actions *Authors* perform. The combination of the two is what allows us to create a responsive task assistance that is aware of not only the sequence of options, but what actions were needed to reach them.

The information collected is divided into steps. Each step is composed of a screen, and the interactions that were performed on it (Table 8). For each, we collect the title of the active window, package name, timestamp of the interaction that caused the active window to change, the type of interaction that caused the change, the file path of the audio description of the structure, the view that was interacted with (i.e. clicked), the list of views that were interactable, scrolled, and where text was inserted/removed. Each view has the information stated in Table 8 and all additional fields of an *AccessibilityNodeInfo*[21]. Every view queried through an accessibility service has the fields of content description and text; however, they are often empty.

RISA relies on textual descriptions of elements to identify targets and to guide *Consumers*, thus descriptions are essential. Often, the element that is clicked, or interacted with, is not the one that has a description. The descriptions can be in a hierarchical child or parent of the interactive view. For example, an interactive LinearLayout with a

[21]AccessibilityNodeInfo (https://developer.android.com/reference/android/view/accessibility/AccessibilityNodeInfo)

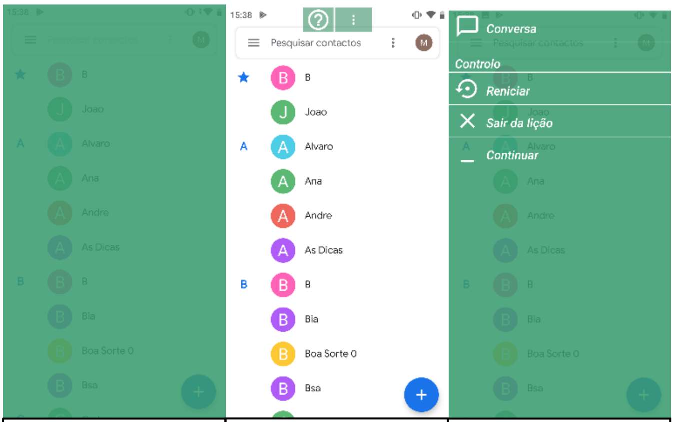
TextView child. The TextView has the text that is read by a screen reader, which is the one we want in RISA, to be able to guide *Consumers.* Thus, for every view, we locate the closest text or content description, first through its children, then its parents. Event text is collected when it exists, as this can also be representative of the view that was interacted. In instances where the text is flagged as sensitive, or dynamic, authorText contains the *Author* written description of the element.

Support Created Content

When a demonstration is created, the *Author* maintains ownership. It is always possible to check its details after it is created, and to delete it permanently. Every task has associated its own discussion thread, where *Consumers* and *Authors* can discuss anything, using the built-in chat features. The interface for the discussion thread is an adapted version of ChatKit. When a new task is created, the *Author* becomes automatically subscribed to its discussion thread. Whenever something is posted, the *Author* receives a notification. The *Author* can then answer and tag the information as relevant to a particular step. Alternatively, *Authors* can also create new tags and mark the thread posts as one from a particular tag. This may be relevant in threads where different discussions take place. Creating these tags enables *Consumers* to not only check information based on the step they are in, but also based on a tag the *Author* deemed relevant. Whenever an *Author* posts, all users who are still subscribed receive a notification.

Answering Requests

When users register as *Authors,* they are automatically subscribed to notifications of requests to all their installed apps. They can always unsubscribe from receiving notifications from an app or set of apps. When a request is made, any *Author* can create a task to answer it. As soon as a demonstration is created, everyone has access to it, and the request is removed from the list of pending requests. Thus, RISA’s available knowledge grows with every contribution.



a) Listening to description or *hint*	b) Free to explore, top bar with *hint* and *more options*	c) Chat and other control features in *more options*

Figure 32 - Screen examples of RISA playthrough.

6.2.3 Task Assistant

When users register as *Consumers* in RISA, they can browse through each of their installed apps and check which have supported tasks, checking their title, description, and steps. If a task is not available, *Consumers* can create a new request describing what they are looking for through text, or by sending a voice message associated with a particular app. When *Consumers* select a task, RISA initializes the app in its first screen, and begins guiding step by step. During the task, RISA behaves somewhere in between an assistant than only intervenes when requested, and an interactive tutorial. It relies on the data collected during the authoring process to generate each of its instructions, and hints, through the device Text to Speech Engine (TTS), thus ensuring a consistency of the delivery of the instructions between different *Authors* and different applications.

Guiding Step by Step

RISA always follows the same procedure. First it blocks all screen interactions with an overlay (except for the home and back button) while it announces the description of the next target element (Figure 32- a).

Then, *Consumers* enter a free exploration mode, where they interact with the app as they want, without interaction interference from RISA (i.e. RISA does not prevent or makes any interaction). During free exploration, RISA creates an overlay at the top centre of the screen, with a *Hint* and a *More Options* element, and monitors *Consumers* interactions. The top centre of the screen was chosen as the default position because it does not interfere with any app layout that we are familiar with. Typically, there are no interactive elements in the top centre, and when there are, they usually occupy a larger area than RISA is covering. Moreover, since it is directly underneath the typical place for smartphone speakers, it has a clear tactile cue for quick access.

When the element *Hint* is focused, RISA blocks again the interaction as in (Figure 32- a), and provides the next hint depending on the target and the interactions monitored (described in detail in a following section). Through *More Options, Consumers* can access the chat for the task and control the general task flow (Figure 32 - c).

During free exploration, if *Consumers* focus the target element, RISA provides an audio cue (i.e. a short 'beep') indicating they are on the correct target. If *Consumers* select (i.e. click, long press) anything but the correct target element, RISA gives another audio cue (i.e. a short 'boop').

To support path recovery, when the *Consumer* has deviated from the path, RISA checks the current screen for all previous targets. If it finds one, it will initiate the default procedure. When a correct target is selected, RISA blocks the screen again and provides the next target description. RISA repeats this behaviour until the last step is performed, then it provides an audio message stating the task was successfully completed, and a short audio clip symbolizing success.

Text to Speech

RISA uses the Text to Speech (TTS) engine available on the device, and selects a different voice from the one being used by the screen reader to ensure the intelligibility of speech (Brungart and Simpson 2005), facilitating *Consumers* ability to distinguish between RISA and the screen reader.

Since Android 8.0, the accessibility volume can be controlled independently of the multimedia volume. To not interfere with the default behaviours, and feedback of the device screen reader, RISA always uses the multimedia API to play instructions. Thus, it is possible, in extraordinary conditions, that the screen reader and RISA speak simultaneously. When it happens, *Consumers* can simply tap the screen to silence the screen reader or adjust the multimedia volume to only silence RISA.

Target Announcement

Statements

Previous:	*" "*	*OR*	*"If you selected the intended option, proceed and"*
Interaction:	*"Select"*	*OR*	*"Do a long press at"*
Target:	`View.ClosestText`	*OR*	*"the element of the type"* + `View.AuthText`
Previous Edit:	*"In this step you will have to write in"* + `nº_of_EditText` + *"edit text boxes. Then"*		
Edit:	`Nº` + *"write"* + `View[Edit].AuthText` + *" then "*		

Default behavior

Announcement:	Previous	+	Interaction	+	Target

Example: *"Select Create Contact"*

Example sensitive/dynamic: *"Select the element of the type message"*

When text was written

Announcement:	Previous Edit	+	Interaction	+	Target	+	Edit [1..N]

Example: *"In this step you will have to write in two edit text boxes. Then select confirm. First write name. Second write number."*

Figure 33 - Available options for target announcement.

Target Announcement

RISA default target announcement behaviour is to state the required interaction, followed by the target closest text (Figure 33- Target) (e.g. "*Select Create Contact*"). If the target was marked as sensitive, or dynamic, the target description is the *Author's* written description of the element (Figure 33- Target). In such cases, it is not possible to determine when the user is on the correct target, nor if the correct one is selected. Thus, following a selection of an unknown target, the next target announcement will include the disclaimer *"if you selected the intended option, proceed and …"* (Figure 33- Previous).

If during authoring, text was written in any edit text boxes, then the target announcement changes as depicted in Figure 33– When text was written. First, RISA announces how many edit boxes will have to be edited, then what element to select afterwards, followed by an enumeration of what to write in each of the edit text boxes.

Table 9 - Hint types.

Hint Types	Composition
A) EDITBOX DETAIL	<Nº> + "*write*" + EditView.AuthText
B) LONG CLICK	*"To do a long click, double tap on the screen and keep the finger against the screen on the second tap"*
C) TARGET POSITION	Target position is divided in nine locations. The four corners, the four edges and the centre.
	"Top right corner"; "Left edge"; "Centre";
D) SCROLLABLE	*"Option is not on screen, try to scroll in the list to find it."*
	"Try to navigate in the list that has" + ViewScrolled.ClosestText + *"until you find the option"*
E) SWIPE SCROLL	*"Find"* + ViewScrolled.ClosestText + *" and navigate the list by swiping from left to right until you find the target"*
	"Try to swipe from left to right after first focusing the element on the top left of the screen"
	"To scroll a list you can swipe from left to right and to the left again in a single motion."
F) TARGET NOT FOUND	*"Example"* + View.ClosestText
	"Target not on screen, if you are unable to recover, try to restart the task"
	"Target not on screen, try to back to the previous screen"
G) LAYOUT DESCRIPTION	Recorded audio file from the task Author is played
H) REPEAT	Repeats target announcement
I) RECOVER	*"You can resume the task from any step"*
J) CHAT	*"If you are unable to proceed, you can check the task chat through more options."*
K) CONTROL	*"If you can resume the task, select more options and restart"*
	"Select more options for other task control options"

Hints on Request

When the *Hint* element on the top centre is focused, RISA plays a single hint from the next hint type available (Table 9). RISA's default behaviour is to play the next hint, from A to K as depicted in (Table 9). If a hint is not available, or does not make sense given the target or previous monitored interaction, then it skips to the next one. When it reaches

K) Control, it circles back to A) EditBox Detail. The flexibility of the provided hints is fundamental to be able to correctly assist the *Consumer,* depending on current context and target. Hints are adapted depending on the *Consumer's* path, if and how far it deviated, and how much information RISA gathered during the authoring procedure. Below we detail how each of the different hint types behave.

A) EditBox Detail, the first hint type, is only available if there were EditBoxes edited during this step in the authoring process. It is also the only hint type where RISA will not skip to the next type until it has provided all hints of its type.

B) Long Click is only available if the element was long pressed.

C) Target Position is available if RISA can find the target on screen; thus, for targets tagged as sensitive or dynamic, it is usually not available. Target location was divided in nine sectors: corners, edges and centre, since corners and edges have been shown to be preferred by visually impaired people (Kane, Wobbrock, and Ladner 2011). Since larger targets can be in multiple locations, the labelling priority is always given from corners to edges, to centre, from left to right and from top to bottom (e.g. an element that occupies the full top edge will provide a hint stating "*top left corner"*). The home/back button bar is ignored for the calculation of the location.

D) Scrollable is available if the *Consumer* hasn't deviated from the path, if the target is not on screen, and if the target is known to be in a scrollable element. During the authoring process, RISA collects the closest text to the scrollable view where the target element is. RISA then uses this text to try and find a representative element of the scrollable list; if it does then it announces the hint (i.e. navigate in the list that has X).

E) Swipe Control is available under the same restrictions as D) Scrollable. It provides a detailed explanation of how to interact with scrollable elements.

There are three methods to scroll lists on Android: page down gesture, sequential navigation, and two fingers scroll. For the page down gesture, and for the sequential navigation, the current element in focus must be one within the desired list. When the gesture used is a two-finger scroll, then the initial positions of the two fingers determine the list which is going to be scrolled. If the list is horizontal, then the two-finger gesture must be done from right to left instead of top bottom.

Due to the known struggles with list navigation, when the scrollable element is a vertical list and occupies a large part of the screen, the hint provided is the page down gesture (i.e. *swipe right left right in a single motion*). Otherwise, to prevent scrolling on unwanted, lists RISA only advises to use linear navigation.

F) Target Not Found is available if the target was marked as sensitive and the *Consumer* hasn't deviated from the path; or is dynamic and not on screen; or if the target is not on screen, is not inside a scrollable element, and the *Consumer* has deviated from the path.

If the target is dynamic or sensitive, then the hint is an example of what it was (i.e. "*Example Ana"*) or what was written by the *Author (i.e. "Example contact name")*.

If the *Consumer* has deviated from the path, then depending on how many interactions off the path the *Consumer* is, the hints differ to restart, or try to go back.

G) Layout Description is available if the *Consumer* hasn't deviated from the path. It is the recording of the layout description given by the task *Author* for the current screen. Since RISA does not impose any limit to the length of the layout description, for this hint only, if *Consumers* double tap on the blocking overlay (Figure 32- a) it stops the hint and resumes free exploration.

H) Repeat, is always available and it repeats the target announcement. In instances where the *Consumer* has deviated from the path, it checks to see if any previous target is on screen. If it is, it makes the target announcement for it as per standard behaviour.

I) Recover, is only available if the *Consumer* has deviated from the path and it ensures *Consumers* are aware, they can resume the task from any step.

J) Chat is always available. When nothing else is enough, this hint mentions *Consumers* can rely on the task chat to ask questions, or to browse previously shared information. The Chat is RISA's last fallback mechanism.

K) Control is always available and points *Consumers* to the task control options in case they need to stop or restart the task.

Matching Views

Unfortunately, there is no mechanism available on Android that allows us to uniquely identify views or screens as an external service. View *ids* and other possible identifiable information are generated during runtime, and change at every rendering; thus, it is not possible to programmatically identify matches between views, or screens, without compromises. For RISA to work, we needed to be able to match targets to current screen elements, and be able to verify whether *Consumers* have focused/selected the intended target. We rely on a combination of the views' characteristics, and the information provided by *Authors* (i.e. view is dynamic/sensitive) to match target views with current screen views. Target location is never used for matching purposes as it may differ from device to device, screen orientation, app version, time of day and many other factors. Since RISA relies on textual descriptions of elements to guide users, view matches rely on the similarities of the text content.

When locating a target on the screen, RISA checks all the interactive elements (i.e. clickable, long clickable) that are visible on screen. If it finds a target that shares the same package name, class name and closest text, it assumes it has found a match. If multiple matches are found, RISA compares element sizes, then hierarchical level (i.e. depth of each element in its DOM tree); if matches are still equal, then it selects the first element to appear on the DOM tree. If no match is found, it searches for interactive elements that share the same class name, package and that contain the closest text of the target, if its text length is larger than 2 characters. This addresses issues where an interactive element contains multiple interchangeable text view elements.

If the view was tagged as sensitive RISA does not look for any matches and assumes that the next interaction taken by the *Consumer* is the correct one. In these instances, as previously mentioned, RISA alerts the *Consumer* in the next target announcement that he/she is responsible for ensuring the step was followed correctly.

When the target is tagged as dynamic, RISA follows the same matching procedure as if it wasn't tagged at all. If it doesn't find any viable matches, it follows the procedure described when a target is marked as sensitive.

Whenever an element is focused, a focus *AccessibilityEvent* is triggered. RISA assesses the content of the event to find if the element focused is the target one. If it is, the correct audio cue is played. Focused events do not always have the same source element as the Click event on the same apparent element. Additionally, events often have null source views. RISA checks if the event has any type of text (i.e. content description or text). If it does, we use it to find a match instead of the source view. Since an event often aggregates information from multiple views, RISA checks if the event contains the closest text of the target element, if it does, a match is found. If the *AccessibilityEvent* has all text fields empty, RISA checks if the source view is available. If it is, RISA matches the view with the same criteria as when locating a target.

6.3 Limitations

The limitations caused by dynamic content are not solved, and they impact the overall implementation of RISA. Although RISA can work around this limitation, it does so by further burdening the authoring process, and causing uncertainty in some task steps.

6.4 Summary

In this chapter we describe what we posit as a Rich Smartphone Assistant, and describe our implementation of RISA, a nonvisual assistant. RISA is the first system-wide in-context assistant for smartphone tasks. We developed RISA to be app independent, and capable of supporting users in a wide variety of tasks, interfaces, and behaviours. We **streamlined the authoring process,** ensuring *Authors* only need to know how to perform the task. We ensured the **consistency of the delivery** by collecting interaction and interface data that enables us to generate homogeneous instructions, despite the possibility of relying on different *Authors*. RISA monitors users' interactions to **adapt its instructions and hints to the current context**. Monitoring also enabled us to provide path recovery support. The discussion thread and task request features resemble the Hint Me solution described in *Chapter 5 - Human-Powered Support*, and ensure users always have a **fallback mechanism** which they can rely on. Finally, RISA knowledge is shared among all users, thus **with every contribution, the support available grows**.

Chapter 7

Evaluating Efficacy and Perceived Self-Efficacy

Throughout this work, we have highlighted how smartphones, despite being conceptually accessible, still pose challenges. In our first study there was a quote that has guided the direction of this research.

> *"Pretend I don't have a phone"*.

The participant that said the above was able to answer calls, send messages, find apps and check the time since the very first week in every laboratory assessment; from the second week onward, he was able to call numbers and contacts. However, he did not feel confident in his own ability to control the device. The lack of self-efficacy defined how he used his device, and how he portrayed its use to others. Family was not supposed to call or worry. The device was accessible and useless. We understood that the issues go beyond what users can or cannot do. Users need to feel confident in their ability to control the device and perform tasks. The availability of a knowledgeable support network appeared to be what many needed in order to feel comfortable exploring and using the device, knowing they always had a fallback mechanism in case they needed.

"*Unless people believe they can produce desired effects by their actions they have little incentive to act. Efficacy believe is, therefore, the foundation of action."* (Bandura 1998)

We believe for a Rich Smartphone Assistant to be truly effective, it must not only support users in completing the task, but also affect users' perceived competence with it. If it does, and if assistance is pervasive, there is a chance it will impact how app and device accessibility is perceived. Consequently, accessibility perceptions can further impact users' sense of self-efficacy.

In the previous chapter, we described in detail all the available features of RISA. In this chapter, we assess if a task assistant can effectively support people in their mobile tasks and affect their perceptions of self-efficacy. We do this through a comparative user study where we recruited 13 sighted participants to create assistive content for six tasks; then 16 blind participants followed them in two playthrough sessions. Our task assistant was able to support participants and affect their perceptions of app/device accessibility and self-efficacy. We use the term self-efficacy in relation to smartphone use, to describe the participants' beliefs about their capabilities to use the device and its applications.

In this chapter, we contribute with **C7) an assessment of the effectiveness of a human-powered nonvisual task assistance**, and **C8) an assessment of the impact of effective task assistance in perceived self-efficacy**.

7.1 User Study: RISA a Human-Powered Task Assistant

To assess the effectiveness and the impact of a rich interactive assistant, we conducted a comparative user study where participants performed tasks, with and without the assistance of RISA. First, to create the content to be used in the study, we conducted an authoring session with sighted users with no accessibility knowledge, where participants were asked to demonstrate a set of tasks. Next, to assess the effectiveness, and the impact of the assistant, we recruited 16 blind people to participate in two playthrough sessions. In one, participants performed a set of tasks with no assistance, as they would do on their own device. In the other, participants performed a set of tasks with the help of RISA.

7.1.1 Research Questions

We set out to answer the following research questions:

1. Can a rich interactive assistant enable non-specialists to create effective smartphone assistance for blind users?
2. Does effective smartphone assistance affect users’ perception of self-efficacy related to their smartphone competence?

Table 10 - Task condition and description.

ID	Condition	*App*	*Task*
TT1	-	**Contacts**	Add new contact
TT2	-	**Contacts**	Add <X> to favourites
T1	A	**Youtube**	Find a video from the History and share it to Facebook
T2	A	**Nextflix**	Check the in app notifications and find a new series. Add it to your watch list and download the first episode.
T3	B	**UberEATS**	Find McDonalds Saldanha and order a Big Mac Menu, medium size with Coca-Cola. (No need to worry about spending money, there are no card details on the smartphone) Check-out.
T4	A	**Google Translate**	Switch translation from Portuguese-French to Portuguese-English. Listen to a phrase that has been previously stored "Quero comprar cebolas". Play it in English.
T5	B	**OneFootball**	Check the competitions we follow. Check the teams in Liga Nós. Check Belenenses squad.
T6	B	**Microsoft Outlook**	Turn on the 'Do not disturb" mode.

We chose a top app of selected categories (Google Play Store - Portugal ranking) that had enough complexity to create a task with more than three steps and basic accessibility compliance (i.e. not having every element unlabelled). For the category *Music & Audio*, although *Youtube Music* was the top app, we chose its more pervasive sister app *Youtube*, that does not appear on top categories due to being pre-installed in most devices.

Tasks varied in number of required steps and workflow structure. The tasks T2 and T4 had a single sequence of steps that had to be performed with no alternative paths available and never passing through the same screen. The remaining tasks had alternative paths available in some of its required steps (e.g. multiple ways to reach the restaurant). In T3 and T4, participants had to go through the same screen at least twice to complete it.

In the authoring session, tasks were counterbalanced. For the playthrough session, the tasks were randomly assigned to group A and B (Table 10). Within a group of tasks participants performed them always in the same order. For group A the order was randomly generated as T1, T4, T2; and for B T6, T5 and T3. The group of tasks was counterbalanced between the two conditions (with RISA and without). The first assigned condition was also counterbalanced between participants.

7.1.2 Apparatus

We used a Xiaomi Mi A2 device running Android 8.1. For every app with a task, a shortcut was made available on the device home screen. RISA was pre-installed on the device. In the playthrough sessions we use Talkback, the default Android screen reader. Participants were requested to use headphones for the playthrough sessions.

7.1.3 Tooling

We relied on RISA authoring capabilities for the authoring session, and its task assistance features for the playthrough sessions. Since we provided the device for the study, no information on the device was deemed as sensitive; thus, RISA was modified to never ask if the information was sensitive.

As we have previously reported in *Chapter 5 –Human-Powered Support*, a Q&A or other last resort fallback mechanism is needed. However, for this study we were specifically concerned with the assistance provided through the demonstration process and not its fallback mechanisms. Therefore, the chat feature for RISA described in the previous chapter (section Support Created Content) was unavailable during the study.

Four of the applications had dynamic content (i.e. *Youtube*, *Nextflix*, *UberEATS*, *OneFootbal*). RISA was developed to deal with both dynamic content, and dynamic target elements. In this study, we focused on assessing the effectiveness and effects of pervasive assistance, regardless of RISA's ability to adapt to this type of content. To limit its impact, users were never guided through a step where RISA was unaware of the intended target, thus no instruction was ever complemented with the disclaimer *"if you selected the intended option…"* (i.e. *Chapter 6 – Target Announcement*).

7.1.4 Authoring Session

For the authoring phase, we recruited 13 sighted participants with no or limited accessibility knowledge; only one had previously tried a desktop screen reader and none knew how blind users interacted with smartphones. Ages ranged between 19 and 52 (M=26.77, SD=8.80), two iOS users and 11 Android. In the following sections we will refer to authoring participants as *Authors*.

We informed participants they would be demonstrating how to do a set of tasks, while using the RISA authoring feature, and that their demonstration would be used in the following weeks to guide blind people. First, participants were given a brief overview of the RISA authoring features, then they were guided through the authoring process for TT1. Next, participants demonstrated TT2 without any guidance, but were encouraged to ask any questions they had. After the training tasks, participants were asked to demonstrate 6 tasks (Table 10). Before recording each demonstration, participants were guided through the task, ensuring they were familiar with it. They were encouraged to explore and ask any questions. When they felt comfortable with it, they started the RISA authoring process. Participants were rewarded for their time with a gift card.

Authors were able to successfully demonstrate a total of 78 tasks, 13 of each app. From those, we discarded 11, due to technical issues with the recordings of the interface descriptions. Every *Author* had at least three of their demonstrations followed.

Table 11- Playthrough session participants. Competence and Expertise were self-assessed in a scale from 1 to 5, with higher being better. Device model was self-reported.

ID	*Birth*	*Age Onset*	*Gender*	*Education*	*Smartphone*	*Years of use*	*Smartphone Competence*	*Screen Reader Expertise*
p1	1967	19	M	Masters	iPhone 7 Plus	4	4	4
p2	1974	0	M	Licentiate	Samsung S8	6	4	4
p3	1971	9	F	9º Grade	Samsung J5	1.5	5	5
p4	1964	0	F	9º Grade	iPhone	1	3	3
p5	1969	20	F	12º Grade	iPhone	5	4	4
p6	1978	30	M	12º Grade	Android/iPhone	6	4	5
p7	1961	0	F	12º Grade	iPhone SE	0.6	2	4
p9	1947	53	M	12º Grade	iPhone 6	4	3	4
p10	1949	40	M	12º Grade	Nokia - Android	5	3	4
p11	1969	4	F	12º Grade	iPhone	0.4	3	4
p14	1987	19	M	12º Grade	Samsung	4	4	4
p13	1968	0	M	12º Grade	iPhone	2	3	4
p16	1987	0	M	Licentiate	Samsung j3	2.5	4	4
p15	1979	0	F	Licentiate	iPhone	7	3	5
p17	1992	14	F	12º Grade	iPhone	6	3	4
p18	1980	18	M	Licentiate	Sony - Android	6	5	5

7.1.5 Participants

For the playthrough sessions we recruited 16 legally blind participants, ages ranging between 26 and 71 (M=46.63, SD=12.64), 9 iOS users and 7 Android, with a variety of self-reported smartphone expertise (Table 11); all had previous experiences with Android devices. Participants were required to be able to at least make a phone call and send a text message with their smartphone.

7.1.6 Procedure

Participants were informed that the purpose of the study was to understand the impact a smartphone task assistant could have. After a demographics and expertise questionnaire; participants were asked about their familiarity with the applications and tasks chosen for this study. Only three tasks had been performed before: two participants had previously shared a *Youtube* video on *Facebook*, but not from History; and one participant had ordered food from *UberEATS* with her iPhone. Participants were informed they would be performing a set of tasks in two conditions: 1) with the assistance of RISA; and 2) as if they were trying to perform the task by themselves. Each condition was counterbalanced and performed in different sessions, in different days, each lasting about 1 hour.

In both, participants were first asked to find the contact Andre to get used to the device. Then, participants performed two training tasks (i.e. TT1 and TT2). In 1) with RISA, TT1 was used to present to participants the playthrough features of the assistant. Participants were guided through the task, encouraged to consult hints, and explained the type of hints and audio cues provided, the assistant guidance behaviour, and the recovery mechanisms.

For TT2 with RISA, participants were asked to follow the assistant alone, and encouraged to ask any question they had. After participants finished TT2 they proceeded to do set A or B of 3 tasks (Table 10) in one of the conditions. Tasks within a set were always presented in the same order. Participants performed a different set in each condition. We ensured each set was performed the same amount of times in each condition.

In the condition with RISA, we randomly selected one of the demonstrations created for that task, ensuring that each participant followed demonstrations by three different *Authors.* No pair task-author was ever repeated.

Participants were asked to complete the task at hand to the best of their efforts. Each task was introduced as described in Table 10, "In application X do Y", participants started from the first screen of the application. In 2) (without RISA), participants were told to notify the researcher when they had completed the task, while the researcher also observed and took notes; in 1), once they had performed the final step, RISA announced the end of the task. If participants did not make any progress towards ending the task after 5 minutes, the task was interrupted, and considered to be unsuccessful. The task was also stopped if participants alerted the researcher that they were not able to complete the task and showed visible signs of frustration, or had given up. Participants could always ask to repeat the task description, but no additional help was provided.

Every interaction with RISA and the device was recorded. At the end of each condition, each participant filled a brief questionnaire about smartphone self-efficacy, based on Bandura's (Bandura 2006) work, with a 100-point confidence scale, 0 ("Cannot do at all") to 100 ("Highly certain can do"). The 11 questions were designed based on our findings of the challenges users face when interacting with smartphones (e.g. learn a new App, understand when something does not work).

Rate your degree of confidence as of now by stating a number from 0 ("Cannot do at all") to 100 ("Highly certain can do").

1. Learn to use a new app.
2. Perform an update.
3. Learn how to use new app features.
4. Find what you are looking for.
5. Understand an app.
6. Understand a screen.
7. Know which gestures to perform.
8. Recalling what I can do in each app.
9. Recalling how to do what I want to in each app.
10. Understand when something does not work or has an unexpected behaviour.
11. Be aware of all the options available.

In the debriefing of the second session, we conducted a semi-structured interview to understand user confidence, with, and without RISA, perceived accessibility and self-efficacy. Participants were encouraged to share their thoughts on learning, exploring, assistance, and about RISA. Participants were rewarded for their time with a gift card.

Table 12 - Codebook for the analysis of the interviews conducted in the playthrough session.

	Code	Description
Assistance Provided	App Assistance	In app tutorial, help menu, manual
	RISA	Reference to RISA
	Tips	Reference to tips in RISA
About what?	Task	Reference to a particular task
	App	Reference to an App or apps in general
	Device	Reference to a device or operating system
	Interface	Reference to an interface element
	Location	Reference to location (layout element location)
	Structure	Reference to layout structure
Recovery	Recovery	Reference to deviations from the path and recovery problems/techniques
Sentiments	Attitude	Attitude (positive or negative) e.g. “important, interesting”
	Desire	Desire to learn/acquire/do something
	Curiosity	Curiosity as a motivation
	Necessity	Necessity as a motivation
	Security	Privacy and security concerns and fears
	Confusion	Reporting confusion when interacting with a smartphone, or misunderstanding of any device/app feature
	Self-Efficacy	Reference to user ability/confidence to control his smartphone both positive and negative
Characteristics	Accessibility	App, device, interaction, interface or other accessibility comment
	Efficiency	Reference to efficiency
	Efficacy	Reference to efficacy
	Utility	Reference to usefulness, helpfulness
	Availability	Related with availability of assistance
Learning	Learnability	Related to the learning process. Difficulty, ease of use.
	Exploration	Exploration has the means to learn.
Who is involved?	Social Circle	Friends, family colleagues
	Internet	Reference to blogs, videos and other internet content
	Groups	Related with groups or forums online (e.g. WhatsApp, google groups)
Expertise	Novice	Reference to novice expertise
	Experts	Reference to experts/expertise
Interaction & Feedback	Voice Interaction	Voice to Text, Siri, Google Assistant
	Feedback	Audio, haptic, tactile feedback
	Screen Reader	VoiceOver, Talkback or other
	Gesture	Gesture related
Problems	Barrier	Reports of a barrier to use, learn or acquire
	Cumbersome	Despite something being “accessible” is too cumbersome/slow to use. (e.g. writing is cumbersome so people use voice)

7.1.7 Design & Analysis

To understand the impact of effective assistance, we must first ensure the assistance provided was successful. Thus, in the following *Quantitative* section, we report the impact of the condition on task and participant success rate. A task was considered to be successful if the participant performed all the required steps, independently of the order. For example, if for the *UberEATS* task, the participant did all the steps but ordered from a different McDonalds than the one requested, the task would be considered unsuccessful.

We used task success rate as the dependent variable and relied on a mixed effects model analysis following the procedure in (Seltman 2012). We modelled Condition and Task as fixed effects to ensure we accounted for the possible effects of the differences in task complexity. Task and participant were also added as random effects to accommodate repeated measures. Task ($p=0.879$) and Condition*Task ($p=0.815$) were not significant fixed effects; thus, to simplify the model we only accounted for the fixed effect of Condition ($p<0.001$).

We transcribed all interviews and conducted a qualitative analysis using, primarily, a deductive coding approach. Two researchers independently coded one interview to revise the initial codebook, adding two new codes and changing other two. The final codebook is reported in Table 12. Then the two researchers coded four additional interviews independently. We calculated a Cohen’s kappa agreement of $k=0.66$ ($SD=0.32$), which represents a fair to good agreement. One researcher completed and revised the remaining interviews.

We were interested in understanding self-efficacy, perspectives on accessibility, confidence when using smartphones and whether they were affected by having effective assistance. We further support our qualitative analysis with the results of the questionnaires on self-efficacy. Considering a within subject design, and since self-efficacy is normally distributed, we applied a paired samples t-test.

7.1.8 Findings

In this section we focus on the analysis of the results of the playthrough session. Thus, when we refer to participants, we are specifically discussing playthrough participants.

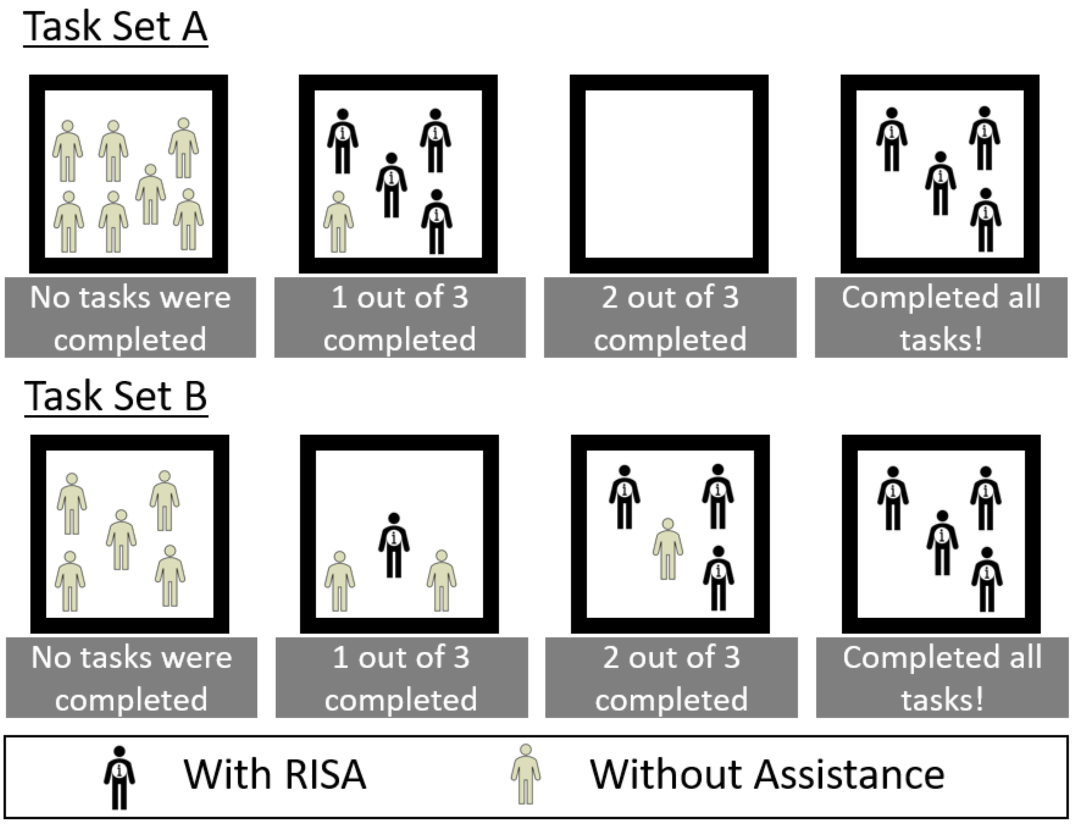


Figure 34 - Number of tasks successfully completed for task set A and B for condition with RISA and without.

Quantitative

Success Rate. When using RISA, participants were more likely to be successful. The condition with RISA had an average success rate of 72.9% (SD=0.45) against 10.4% (SD=0.31) without RISA. The Estimates of the Fixed Effects calculated using the mixed model show that Condition (F=64.849, p<0.001) had a significant effect on a participant's success rate. The condition without RISA, had a negative effect on the task success rate between 76.9% to 46.5%.

In Figure 34 we can observe the success rate of the two different task sets per condition. Although RISA was unable to support all users in completing their tasks, half of the participants completed all tasks, and all completed at least one of the three tasks successfully with RISA. On the other hand, without RISA, the majority of participants (12) were unable to successfully complete a single task.

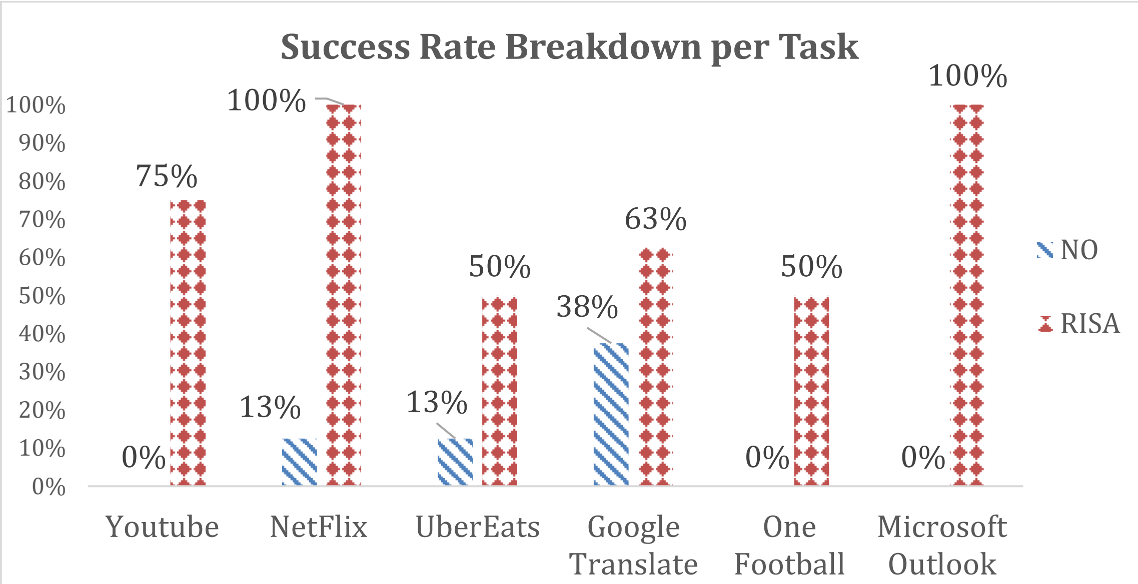


Figure 35 - Success rate breakdown per task between the condition without RISA (NO) and with RISA.

When we analyse the success rate per task Figure 35, the data suggests that for some tasks RISA was highly effective (i.e. *Nextflix* and *Outlook*), while for others we only observed marginal improvements (i.e. *Google Translate*). For *Google Translate,* participants had at times multiple elements with the same name, and only one was correct. Although RISA provided an audio cue when the correct element was focused, the wrong one was first on sequential navigation and close to an edge, and was therefore easier to reach.

Hints Consulted. With RISA, participants were able to consult hints (*Chapter 6 - Hints on Request*) at any time during exploration. A total of 300 hints with an average of M=6.00 (SD=5.77) per task, and M=6.14, SD=4.72 per participant were triggered during playthrough sessions. In 20.8% (10) of the tasks, participants did not rely on any hints, and only followed the step by step instructions. Of those, in 8.3% (4) of the tasks, participants were unsuccessful. As expected, participants had a wide variety of hint consulting behaviours, ranging from individuals who consulted a single hint, during the three tasks, and only completed two tasks successfully; to one that consulted 60, and completed all tasks successfully with RISA. Hints require users to actively seek out assistance. In this study only some participants leveraged them to their benefit.

Qualitative

All participants reported being interested in learning new applications. Some were driven by curiosity and the possibility to improve their daily lives, while others reported to do it out of necessity.

"Everything is always changing... The systems are always updating, it is always better to learn new stuff." P3

Structure based learning. When learning a new app, particularly when out of curiosity, participants reported how it is a slow and methodical procedure of trying to make sense of the underlying interface. All participants described how first it's all about exploring the screen, to try and create a mental model that allows then to interact successfully. The amount of effort it takes is highly affected by user expertise and app accessibility.

"*Some apps are complex. It takes a while before I figure out what I can do, how to navigate, it's complicated." P6*

"One of the last ones I installed was a guitar tuner. I had to explore by touch to find the buttons that would make each string sound. It is not even remotely accessible by touch." P1

RISA was effective in providing assistance but was not perceived to support learning through exploration, but rather as an alternative. For expert users who are comfortable learning through exploration, RISA's current support might not be the adequate solution.

"(RISA) is important when we want to complete a specific task, but for exploration I cannot tell if it is functional, if it helps. RISA gives accurate references, what happens is I focus on them and completely ignore the rest." P15

"(RISA) For more expert users I believe, although it can help, it might not be necessary, or at least as useful." *P16*

Task based learning. The alternative described by participants to exploration was 'task oriented learning'. When learning something out of necessity, when time was a factor, participants reported how they are simply trying to figure out how to do a task ignoring all rest. Additionally, participants also start to describe their learning process as task oriented once they have a basic understanding of the app. RISA was associated with this kind of learning by all participants.

"I always go through the options it has, then I try to accomplish my goal (...), if I can't I ask for help (...), then with repetition I am able to do it..." P11

Participants believed RISA would enable them and others to be more efficient when trying to complete a new task.

"*(RISA) helps getting where we want to go faster. With efficacy and without making mistakes." P6*

However, despite the benefits in the success rate, participants believed RISA contribution to be efficiency. In this study, users had a time limit to perform the tasks. Thus, even when unsuccessful, participants did not report they were unable to do the task. Rather it would take more time and patience. All but one participant reported RISA would have a positive impact on a novice learning experience.

"*I believe it's quite a useful tool, particularly for those who are starting to explore smartphones. It ends up guiding the user faster, more efficiently to the right places where one should press." P16*

"(Who could benefit from having RISA?) More experienced people (...) I think it's too complicated." P3

Additionally, 13 participants reported RISA would be helpful for everyone when trying to learn a particular task, or when first interacting with a new app.

"Almost everyone, even for more experienced people. Even when they learn a new application by instinct, they would get there faster. I would recommend even for experienced people." P11

RISA Improvements. A common request across participants was for RISA to be integrated in current voice assistants (i.e. Siri) to be called upon whenever needed.

"Should be able to activate by voice: 'Hi assistant, how do I do this?'" P18

When targets are in the middle of the screen close to the right or left edge, RISA hint announces left edge. Two participants wanted location tips to be more specific and reiterate right edge **centre** (i.e. currently if the location was edge but not centre RISA would state corner). Two other participants mentioned how they should be able to control the verbosity of the hint and guidance to ensure they do not have to wait for RISA in longer descriptions. Since participants only engaged with tasks that were available in

RISA, we had no reference to its availability. However, two participants expressed the need to have a wide coverage of assistance. They wanted to be able to choose the tasks they learn and not a pre-set of tasks, as is traditional with tutorials.

"*I would add (tasks) to the assistance so I would be the one choosing the task that I want to learn." P1*

Applications perceived accessibility. Participants reported their self-efficacy to be influenced by how they perceive an app's accessibility. When first exploring an app, if basic accessibility guidelines are not followed (e.g. buttons have labels) participants can quickly dismiss the app as inaccessible and give up.

"*It is always a challenge to understand if an app is accessible or not. If not, it is enough to make me give up on it." P15*

"It all depends on how an app is built. My confidence is proportional to the degree of intuitiveness of the app's features." P16

Two of the tasks had inaccessible elements: *Nextflix* had elements with no labels, and *UberEATS* had labels in a foreign language, unknown to most of the participants, especially when we consider a Portuguese screen reader trying to read in English. One of the major struggles we observed, during the trials for these tasks, was precisely dealing with this content when navigating. When using RISA, participants quickly dismiss inaccessible content, or not based on RISA guidance. However, when without RISA, participants appear to be stuck more often exploring this content, wondering where to go, or what was happening. Participants associated RISA with being helpful/useful and the cause for some tasks being easier to perform than others. Effective assistance appears to have an effect on the perceived complexity of the task.

"Without any room for doubt, today everything was simpler. The options were not always obvious, and with the assistant I knew exactly what to look for." P17

"(Which tasks were easier?) From yesterday. Because of the assistant, it allowed me to quickly know what I needed." P5

Confidence and effective assistance. Participants expressed they are not confident when trying new apps. The fears are like the ones found in other computer like devices: fear of breaking the device, deleting items, unintentionally making payments. People are often left with no choice other than trying, even if they are not confident.

"A bit afraid of messing up, that I block it. Deleting stuff that I shouldn't or installing something that I have to pay and not understanding it." P3

"*Not confident, not really. However, I try to understand what is going to happen."P2*

RISA was thought of as a tool that would: 1) facilitate the learning process; 2) enable to do tasks faster; 3) provide clear paths avoiding doubts and fears; and 4) facilitate the understanding of interfaces. After using RISA participants reported higher rates of confidence, particularly in their ability to understand their current screen and the overall flow of the application.

"*For me it is a bit difficult to explore new applications (...) (With RISA) maybe I would be more confident." P4*

We found that RISA affected the participants' ratings of self-efficacy. The average score of the sum of the scales for self-assessed self-efficacy M=91.31 (SD=14.23) with RISA, and M=78.75 (SD=19.91) without (i.e. higher scores represent higher efficacy). The paired-samples t-test conducted revealed a statistically significant difference between with RISA and without, t(15)=3.11, p=0.007 (two-tailed). The mean decrease without RISA was 12.56 with a 95% confidence interval ranging from 3.96 to 21.16. The eta squared statistic (0.39) indicates a large effect size (Cohen 2013).

RISA as an information tool. One of the raised issues with learning new apps are the descriptions of the apps on the App Stores not being helpful or factual. Participants saw other potential uses for RISA. Instead of relying on RISA to learn tasks, they saw it as a quick way to check what was possible to do in a given app.

"I would look through the relevant tasks, instead of looking around, RISA would give me all the information I wanted." P14

7.1.9 Discussion

We conducted a comparative user study to assess if we could provide effective smartphone assistance, by relying on untrained individuals, with limited accessibility knowledge. Furthermore, we assessed how it impacted users' perceptions of accessibility and self-efficacy. Herein, we discuss our findings, that should be of interest to researchers and practitioners working on nonvisual mobile accessibility.

Shift in learning practices

Throughout our work, people have described how they first engage with new smartphone applications. First, people explore the whole screen, to create a mental model of the available options. This procedure is like how sighted people engage with new apps, quickly grasping the interface through visual cues. However, for blind people the procedure is long, methodical and has a substantial workload for complex, or out of the ordinary, applications.

Learning how to interact with technology, seems to resemble how one learns a second language. Learning how to first perform gestures, and then learning through exploration by first creating mental models of the structure of an application, appears to be the equivalent of the focus on language structure. Since the 1970’s, the assumption has been that focusing only on language structure is not enough, but rather there needs to be an association with an ability to express meaning (Skehan 2003; Prabhu 1987; Widdowson 1978); the smartphone equivalent would be the ability to perform tasks. Since then, task-based teaching (Prabhu 1987) has been the prevalent construct by which teachers have created their syllabus (Seedhouse 1999) and the ‘dominant paradigm in the teacher education literature’ (Lynch and Maclean 2000). One of the arguments against task-based learning in second language learning is the difficulty in creating tasks that are representative of the real world (Seedhouse 1999; Lynch and Maclean 2000). We argue that technology is still far behind current teaching practices, and for blind people, the status quo is a decades old approach that forces form before function. The limitations of transposing real-world, useful tasks to the classroom does not apply to technology learning, or assistance, as tasks can easily be performed in a real context.

RISA has allowed to shift the focus from form to function by allowing users to learn in a task-based environment. RISA was not thought of as an assistant that was able to support a methodical exploration of structure, but rather as an alternative. Particularly, for novice users, it was described as an easier way to get familiar with applications by performing tasks, while ignoring the rest that would create additional workload. With such assistance, users reported they would be able to focus on tasks, learn, and perform them quicker.

After being familiar with a particular set of tasks, users believe it would be easier to understand the application, given they would already be aware of some of the workflows, and interactions behaviours. Similarly to learning a second language, once learners are proficient, a task base scenario might only help for unowned contexts.

Non-specialists were able to provide effective assistance

In *Chapter 5 - Human-Powered Support* we found a mismatch between the assistant provided by non-experts and the required by end-users. With RISA we were able to fill this gap by structuring the authoring process, implicitly collecting app and interaction data during authoring; and by, monitoring user interactions and assessing app structure during assistance. Untrained sighted individuals created assistive content that had a positive impact on participants task success rate when compared against no assistance.

In the previous chapters, we have described in detail the challenges that users face when interacting with smartphones. Many are not able to independently tackle these issues and have to rely on other people. Unfortunately, some do not have anyone to rely on or would rather not bother their peers, friends and family. We intentionally compared our system against the other current available independent, alternate way, that most report to rely on, exploring by themselves. Our baseline comparison does not portray the full spectrum of possibilities. Many people will only try to learn something new if they have help, or if someone told them about it. Nevertheless, our goal was also to enable these users to have the choice to not have to rely on others and be in control of their learning process. We have previously described how unknowledgeable individuals may be unable to assist, even when co-located. Solutions like RISA may be developed to support co-located assistance, where demonstrations on one's own device is translated into a format that end-users are able to engage with.

We considered RISA to be effective since it had a positive impact on success rate. However, RISA was not able to support everyone in every task. Still, with RISA, the success rate rose from ~10% to ~70%. Playthrough participants completed six different tasks, in six different applications, from six different categories. It appears RISA is flexible enough to provide support in a variety of application, contexts, and interfaces.

Self-efficacy was affected by effective assistance

Self-efficacy has been reported to have a positive effect on the decision to use web technology and influencing actual use (Yi and Hwang 2003). Assessing if a solution is effective or not, is not enough to assess its potential impact. In this study, we go beyond the traditional metrics and make the first attempt to assess blind people self-reported smartphone self-efficacy. We designed a self-efficacy questionnaire to be sensitive to differences in self-assessed confidence to complete a variety of tasks associated with smartphones. The questionnaire revealed a significant effect of RISA effective assistance in its measure. With RISA, participants reported to be more confident in their ability to understand interfaces and apps. RISA use of clear targets enabled participants to ignore all possible confounding and inaccessible elements. We found participants attributed the improved success rate to an improvement in efficiency.

Effective assistance appears to have an impact beyond success rate. Our results suggest an effect on how users measure their own ability to perform tasks, and how they perceived apps accessibility and complexity. In turn, the increase in self-efficacy may be what is affecting participants perceptions of task difficulty. Similarly, Agarwal et al. (Agarwal and Karahanna 2000) showed a strong relationship between application specific self-efficacy and ease of use.

A Rich Smartphone Assistant that can cover a wide range of apps, and functions as a third-party service, has the potential to impact user's overall smartphone use and exploration behaviours. In this study, we observed how having the assistant impacted how participants reacted to inaccessible content, complexity, and self-reported their confidence to control the device. We believe the results are promising, revealing how effective assistance can support users to learn and explore independently.

Limitations

RISA was assessed in a comparative study, in a laboratory setting, with a particular device, and with researcher designed tasks. Although, the results are promising, an effective smartphone assistant should be able to adapt to the wild, and provide support for tasks that users request, which might differ from the ones created.

In this study, we controlled the tasks, and thus did not assess if RISA was able to successfully tackle the issues associated with dynamic content - described in the previous chapter. Although we designed RISA to be able to tackle those issues, future work needs to be conducted to ensure third party services have the necessary access to the content and its properties, enabling the creation of adaptable assistive solutions.

7.2 Summary

In this chapter, we evaluated the effectiveness of our built smartphone task assistant in a comparative study. In a first session, sighted people with limited accessibility knowledge, created assistive content for six different tasks in six different applications. Then 16 blind people participated in two playthrough sessions: one where they did tasks with assistance; and another where the other tasks were completed without any help. Our results show RISA significantly affected participants' task success rate. With RISA, participants completed on average 70% of the tasks against the 10% without it. RISA was an effective smartphone task assistant that showed we can leverage non-experts to create effective content that impacts user's success.

We found RISA had a positive effect on perceptions of self-efficacy and perceived accessibility, suggesting future work can leverage this type of tools to improve performance and both internal and external perceptions.

Chapter 8

Conclusion

All the previous work presented in this dissertation, led to the user study reported in the previous chapter, which allowed us to demonstrate our thesis:

Human-powered smartphone assistance by non-experts is effective and impacts perceptions of self-efficacy.

In this last chapter, we start by highlighting the major contributions and results of this dissertation. We then discuss the implications for the design of smartphone assistive solutions that rely on non-experts. While we were able to provide evidence for our thesis, this work has limitations, which we discuss in detail. Lastly, we discuss a set of actionable directions for future research.

8.1 Contributions and Major Results

This dissertation thesis, and contributions, are supported by our findings in six user studies, designed to characterize smartphone challenges, and to explore human-powered assistance. Below we summarized our major contributions.

First, we conducted a two-month longitudinal study with blind participants, where we observed and assessed the smartphone adoption process. We found that the **current getting started mechanisms are not effective for newcomers**. The major challenges come from the **mental model mismatch,** created from the transition between a rather sequential feature phone with mostly one interface type, to the perpetual changing and intertwined workflows of smartphones. These issues have **forced people to rely on the support of others frequently or give up on the transition when no help is available**.

Our next two studies focused on identifying issues at a larger scale and assess their pervasiveness throughout different levels of user experience and expertise. First, we conducted a user study where we ran seven workshops locally with a total of 42 participants, with a variety of expertise levels. Then, to assess challenges at a larger scale, we analysed the top content of the largest Android and iOS forums dedicated to visually impaired people. Results show **the need for assistance is prevalent among all expertise and experience levels**. The issues with understanding the underlying interface remain long after the adoption process has been completed; with **users often struggling with mapping out new interfaces and being aware of all the available options**.

Then, we set out to explore if human-powered assistance could be the solution to the identified need for additional support and availability. First, we conducted a user study with an in-context Q&A service, powered by an all-knowing volunteer, which successfully assisted users in completing all tasks. Using the service as a probe, we were able to explore the perceived usefulness and acceptance of human-powered services for smartphone assistance. Furthermore, we derived a **set of implications for the design of solutions in human-powered smartphone assistance,** which were leveraged in our following approaches. This study relied on a single volunteer with full context and domain knowledge, and raised the question of how one could mimic its results without one. Thus, in the following study we explored how people without any context and/or domain knowledge (i.e. sighted and blind people) create instructions, to understand how one can begin to leverage them for future solutions. In this study, we **characterized the information provided by volunteers and identified the required information and feedback needs for end-users**. We extend the insights provided in the previous study informing future solutions on **authoring support, providing flexible instructions, and highlight the need for a fallback mechanism**.

Finally, we conducted a **comparative assessment of the effectiveness of a human-powered nonvisual task assistant**. In this last study, we assessed if the assistive content created by non-experts was effective in assisting end-users in completing a variety of tasks. We found success rate was positively affected by the use of the assistant. Thus, we provide evidence that **human-powered smartphone assistance by non experts is effective.** We assessed if the assistant had any impact on end-users perceptions, with a particular focus on self-efficacy. We found the **assistant enabled users to have a different type of learning experience, with a task centred approach that was seen as beneficial, particularly for novice users**. Lastly, we showed the measured **self-efficacy was affected by our effective assistant**, enabling us to validate our thesis.

8.2 Implications for Designing Human-Powered Non-Visual Technology Assistance

8.2.1 Identify relevant types of information required and provided

The first step one should take when designing novel assistive technology, for a particular domain, is to identify and characterize the types of information that are required by end-users. With an understanding of the information needs, we can start to design solutions that cater to them. There are contexts where it will be possible to create assistance that solely relies on automatic solutions. However, many complex problems are still better addressed by people. In such cases, human-powered solutions can be the answer. When designing technology that relies on people's contributions, identifying what information can be provided becomes an essential part of the solution. Thus, one should always start by characterizing both what is required, and what we can expect to be able to be retrieved, from individual contributions.

8.2.2 Dissect the contribution process

With an understanding of the information provided, and required, we will be in a position to design effective tasks. Request each type of information independently, whenever possible, to focus contributors on a single problem. Provide examples and suggestions of what can represent an adequate contribution, particularly when relying on non-experts on the domain, or people without full context awareness. Avoid free form contributions to minimize the variability. Minimize authoring efforts by each individual user; people are more likely to contribute if tasks are simpler(Rogstadius et al. 2011). Whenever possible, rely on automatic mechanisms to not overburden contributors.

8.2.3 Ensure delivery consistency

By dissecting the contribution process and relying on both automatic and human-powered sources, we are also diversifying the sources of information. However, when providing information to an end-user, it becomes crucial to ensure the assistance provided is a single entity, recognizable, with consistency in the delivery of the information, and therefore predictable regardless of the information source. Avoid confounding end-users with multiple feedback sources and provide information through a single distinguishable feedback channel.

8.2.4 Design flexible instructions

Updates and changes are no longer yearly affairs. Interfaces, apps and platforms have quick and small update cycles. Changes are often tested, made and even reversed in a short time span. Furthermore, there is a vast number of apps and services that although interfaces remain relatively stable, their contents are highly dynamic (e.g. Twitter feed). An effective assistant must be able to **adapt to the high volatility of current interfaces** and their content. And it should be able to ensure the information collected is enough for the developing solution to be able to adapt effectively; or, at the very least, to create mechanisms to deprecate outdated information.

Particularly for third-party assistive technologies that have limited control over apps' content and structure, it is important to **model the underlying structure** that it intends to support. Solutions can be designed to go further and **model application** workflows. With that understanding, we can be in a position to adapt instructions that are able to deal with the volatility caused by dynamic content.

Instructions provided need to be flexible, to not only deal with the volatility of apps structure, and content; but also, with the variety of different users' expertise, behaviours, and requirements. The information required by each user, at each point in time, when interacting with mobile interfaces, is highly variable. One way to predict information needs, and adapt instructions, to a given user, is by pre-emptively **monitoring and modelling user behaviours**. The more we know about an app and user, the better we can adapt instructions to be effective.

8.2.5 Support recovery

Inevitably, users will stray from the intended path, purposefully or not. Assistance will be requested when users are in unexpected states for the tasks they are trying to achieve. Therefore, it is necessary to create mechanisms that not only guide through an optimal path, but are also capable of conveying current state and provide assistance for users to recover and resume their intended tasks. One could argue that to avoid recovery one could try to force users down a single path. However, being able to understand and recover interface states is an essential part of becoming a proficient user.

8.2.6 Always have fallback mechanisms

If we are able to design solutions that take into account all previous considerations, we will be able to provide effective assistance, as we have shown in this work. However, without any prior knowledge, we will not be able to predict every issue or question users will have. Thus, it is essential to design fallback mechanisms that allow end-users to request new, or additional information, about whatever they are trying to accomplish. To do so, create solutions that enable a dialog mediated by technology. This channel should be designed such that discussions further fuel the accessibility and adaptability of the assistance provided.

8.2.7 Support different learning behaviours

Nowadays, when interacting with a smartphone interface, blind users typically start by navigating all elements available on a screen, to create a mental model of the underlying interface. As we have described during our work, this is a long and at times demanding process. The task based alternative, where users learn by performing specific tasks while ignoring the rest, was perceived to be a less demanding approach, that would be ideal for novice users and when first exploring new apps with a particular purpose. However, for experts, who are comfortable with free form exploration, task-based assistance does not support their learning behaviours. For a solution that is able to cater to the variety of expertise, and learning behaviours, the assistant provided must support both. Facilitating use is not just about overcoming challenges; it is also about promoting serendipitous discovery of new features. Assistance must allow users to have control of their learning process in order to learn at their own pace, and in their own terms.

8.2.8 Aggregate knowledge

Human-powered solutions should not waste user contributions in a single use, but instead leverage them to iteratively build a shared knowledge base. Crowdsource in human-powered access technology has been leveraged most often in a one to one basis (BeMyEyes, VizWiz), where a single user and a single volunteer are part of the process and all knowledge created is ephemeral. Although not all problems/solutions can be leveraged from one person to the next, in the specific domain of tech assistance, they will most often be representative. Thus, we should design solutions similar to Social Accessibility (Takagi et al. 2008), LemonAid (Chilana, Ko, and Wobbrock 2012), RISA, among others, ensuring we are leveraging every contribution to improve the overall performance of our assistive technology. Moreover, we must look for opportunities to pre-emptively generate knowledge, enabling better coverage and availability.

8.2.9 Respect privacy

We know mobile devices have become increasingly more personal and private. Our studies suggest that in human-powered solutions, when end-users do not engage in real-time, directly with volunteers, they do not manifest any privacy or security concerns. In instances where users were engaging directly, it became clear the need to devise transparent mechanisms that enable users to be aware of what they are sharing, and with whom. Furthermore, although we did not find concerns about the reverse (what is being shared with me and from whom) there needs to be a provenance to every contribution, particularly when technology is designed to guide users through tasks.

8.3 Limitations

While we were able to develop, and assess effective human-powered assistance on smartphones, our work has several limitations that should not be disregarded.

We conducted extensive analysis identifying, and characterizing mobile challenges, with data collected both in laboratory settings and in-the-wild. However, to validate our hypothesis, we conducted a controlled laboratory study, with fabricated challenges, that may not represent users' interests, or trigger natural learning behaviours. Further studies are needed to understand how assistance in-the-wild may or not be effective; what challenges arise from this deployment; and how to accommodate an ever-growing set of supported tasks.

Although we created a solution that should be able to handle the nature of dynamic content on smartphones, it has not been validated. Furthermore, the workarounds used to overcome the technical limitations imposed to third party services on smartphones have added additional workload to the authoring process, which is far from ideal.

We first set out to provide effective smartphone assistance to all users regardless of expertise level. However, RISA appears to be perceived as particularly helpful for novice users, while only supporting experts occasionally when trying to accomplish a specific task. On the other hand, Hint Me is by design only able to support freeform exploration, and relies on end-users being able to express their issues, thus it is more suitable for knowledgeable users. There is one expertise level that we did not cover with any of our solutions, and that future work should seek to improve upon: newcomers. We identify the major issues with getting started with a smartphone, but our solutions target the next step, after the user is, at the very least, comfortable with basic navigation.

Lastly, we conducted multiple studies and assessed the performance on a limited number of tasks. We tried to cover a wide variety of categories in our last study to ensure our solution, and implications, can be useful for other contexts. However, we focused our work in apps and tasks that did not require complex gesture interactions such as games, which would require different solutions and are interesting subjects of future work.

8.4 Future Work

The findings presented in this dissertation should be of interest to researchers and practitioners in the fields of mobile accessibility and human-powered assistance. Below we discuss some of the open challenges and posit areas of attention for future research.

8.4.1 Developing with Accessibility

Despite the vast amount of work that has targeted basic accessibility challenges, they are still prevalent and relevant. In our work we have focused on providing assistance for when everything else fails, but accessibility by default must remain an open challenge that we should be working towards. There have been recent efforts in the standardization with the W3C release of the Web Content Accessibility Guidelines 2.1. One of the focus of the new release was to improve accessibility guidance in respect to mobile devices. Other efforts have also been industry led, particularly for native applications (e.g. Google[22], Apple[23], BBC[24]). However, the smartphone ecosystem is highly fragmented with no guidelines currently imposed on developers. Recent work by (Ross et al. 2017) proposed an accessibility framework based on epidemiology, seeking to explore the accessibility issues of mobile apps at a population level, within the context of the entire mobile (and software development) ecosystem over time. There is work to be done in how technology can intervene at development time, to ensure compliance with accessibility requirements. For instances, common building blocks for interfaces should be accessible, as Apple has done for some of their pervasive controls and widgets.

8.4.2 Forced Interfaces

The decline of physical buttons is still a concern for some who wish to turn back the clock. For many people, adapting to touchscreen devices is a need that was forced upon by the shift in the phone market. Although people recognize the benefits of touchscreens, for most, they are still not efficient or as easy to handle when compared with physical keys. Providing assistance can be beneficial for people who struggle to interact with touchscreens, but it won't replace the need for better interaction methods. There has been some work in augmenting current devices with cases for multiple purposes, from having a t9 (X. Zhang et al. 2018), to editing text (Trindade et al. 2018). However, blind people do not require to have the smartphone out to receive any feedback. Is a touchscreen really the best input method we can design for blind people to interact with technology on the go? We have this artificial restriction imposed solely because sighted people require a screen for feedback. For desktops, televisions, entertainment systems and other technology, it seems that often touchscreens are the least accessible option. What can we design if we think about taking a 'controller' out of the pocket instead of the smartphone? We believe there is an opportunity to design novel interaction controllers that enable users to reap the benefits of smartphones, while not being restrained by the touchscreen.

---

[22] Material Design Accessibility Guidelines (https://material.io/design/usability/accessibility.html)

[23] iOS Guidelines (developer.apple.com/design/human-interface-guidelines/accessibility/overview/introduction/)

[24] Mobile Accessibility Guidelines (http://www.bbc.co.uk/guidelines/futuremedia/accessibility/mobile)

8.4.3 Ubiquitous Accessibility Information

During our studies people expressed the lack of available information regarding the accessibility of different devices and applications. Discovering and exploring new apps is always a daunting task, where users are often faced with completely inaccessible applications (e.g. no labels). Similarly, as we have stressed throughout this dissertation, updates and changes are frequent. When they happen, users have to re-adapt to already known interfaces. Past research has already started to investigate how the choice of technology, or the uptake of new ones, can impact the overall accessibility in the context of web pages (Richards, Montague, and Hanson 2012; Duarte et al. 2016). We believe there is an entire field of underexplored research going beyond understanding how app accessibility evolves, and how it affects each and all individuals. We need to start acting on this knowledge and develop solutions that are designed to address the volatility of today, providing the relevant information to the users at the different stages.

For example, users struggle to navigate the device and app market with no knowledge available of their accessibility features or compliance. In both instances, users have to rely on third parties (e.g. friends, forums) to know about the device or apps. For mobile applications, users can leverage ratings, app stores descriptions and comments, to try and pre-assess their accessibility. A dedicated accessibility rating and other metrics could allow users to make informed decisions. Automatic evaluations and new metrics for accessibility may be needed to accommodate the dynamic nature of most mobile apps, that frequently cause awareness, and mental model challenges.

8.4.4 Supporting Newcomers

During our studies we found newcomers face challenges that are often only overcome with the assistance of others. However, the solutions we present in this dissertation tackle the next step in a user learning process. Although there has been some work on how to learn (Oh, Kane, and Findlater 2013) and perform gestures on touchscreens (Kane, Wobbrock, and Ladner 2011; Buzzi et al. 2017), it appears the challenges are not yet surpassed. Gestures are not discoverable, nor easy to learn based on some of the given descriptions, and users have no fallback mechanism to rely on, except for their support network. It is time to think about how gesture discovery and practice can be embedded into everyday interactions, how can we track performance, hint at corrections, adapt recognizers or even develop entertainment apps whose sole goal is gesture practice and discovery. For instance, one could develop a game for users to learn and train screen reader gestures, like what Microsoft did for the mouse with the release of Minesweeper for Windows 3.1 (Kapp 2013).

The shift from single thread applications to an OS and apps with a variety of complex workflows has been largely ignored. Future work can explore how we can best convey this new interaction paradigm in order to accelerate the adoption process.

8.4.5 Pervasive Assistance

In this work we highlighted the importance of always available assistance and its impact on user's perceptions of self-efficacy. Assistance should come by default with any technology and users should be able to rely on it to successfully master the device. For example, voice assistants have been integrated in modern mobile devices (e.g. Siri), and have recently started to permeate into our homes (e.g. Alexa) and workplaces. They are designed to assist in our daily affairs, from sending texts, to mark a restaurant, making a search or call someone without any other interaction other than voice. However, they have yet to be leveraged to facilitate the user adoption of a device or to provide assistance when interacting with a particular interface. We can imagine what we could achieve if such assistants in addition to their current abilities, were able to leverage the crowd and provide tech assistance. Further work is needed to understand how these voice-assistants can provide pervasive assistance relevant to the user context, facilitating the use of technology that is not directly associated with predefined commands or tasks.

8.4.6 Data Donors

We envision a paradigm shift where interactions, and contributions by knowledgeable users can assist others beyond what the app and OS provide. By exploiting this knowledge, we can establish support networks where there were none, enabling the creation of services and tools that tackle accessibility challenges informed by usage data. RISA could be an example of such a service that gathers information about application workflows and available tasks. We have previously presented Data Donors (Rodrigues, Montague, and Guerreiro 2018), a conceptual framework proposing the enablement of users with the capacity to help others to do so, by donating their mobile interaction data and knowledge. We believe there is an opportunity to explore how to establish global data donation programs, that could be leveraged in the creation of smarter accessibility services. Such programs should carefully consider how to maintain user engagement and design to promote intrinsic and or extrinsic motivation (Rogstadius et al. 2011). If we consider *Youtube* as an example, people share videos and create content for multiple reasons, but there are mechanisms in place to reward the contributions to the platform. Videos have likes, channels have subscriptions and paid adds. *Authors* can be rewarded with prestige, recognition and even monetarily for producing appealing content. The rating systems in place also serve as filters to make sure *Consumers* are presented with relevant and appealing content. Similarly, for data donation programs there must be clear benefits to both stakeholders in order to make the solution sustainable. In some blood donation programs in Europe donors are notified whenever their blood is used and by which hospital. To promote donors' intrinsic motivation to help, one can create backtracking mechanisms notifying and quantifying how, and when, their data is being used, and for what purpose. Pre-emptively generating knowledge about application interfaces and workflows is an exciting venue of research that could drive the design of novel assistive solutions. While we do not argue it can save lives, it has the potential to change many for the better.

8.4.7 Evaluating accessibility research

Accessibility research has been mostly conducted in single laboratory assessments, with prefabricated scenarios, in neat control environments, often with small user numbers; and there is undeniable value in its contributions. However, we believe we should start to work towards evaluation at scale, not in size, but in-context. As technology became ubiquitous so have the challenges. It is increasingly harder to mimic real-world conditions in lab assessments due to the variability of contexts where technology is used. It is not even just about the where, it's about with whom, for what, at what time, in what mood and so many other variations in-context. In our work we developed TBB in an effort to understand the adoption progress with in-the-wild ecological valid data. The current panorama of available tools for this sort of study is limited, and even the ones that are available can only capture a fraction of the richness of the user context (e.g. subjective measures of mood and stress might reveal new insights). The concept of a Data Donor program, described in the previous section, can be expanded to think how we can scale evaluations where users willingly participate with their time and data.

We started this dissertation by arguing that although mobile devices seem accessible, they are not. The belief that accessibility is good enough because some can use it to its full potential should be deprecated. We need to start expanding our accessibility metrics beyond if a user can or cannot, is faster or more accurate. Rather than solely focus on technology performance, we need to shift our focus to assess its impact on people. In our last study we assessed the effect of assistance on self-efficacy, and our results suggest it could impact users experience with their devices. There is a gap in current work in exploring and validating new metrics that put the focus on people rather than technology.

Bibliography

Agarwal, Ritu, and Elena Karahanna. 2000. ‘Time Flies When You’re Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage’. *MIS Quarterly* 24 (4): 665–94. https://doi.org/10.2307/3250951.

Anthony, Lisa, YooJin Kim, and Leah Findlater. 2013. ‘Analyzing User-Generated Youtube Videos to Understand Touchscreen Use by People with Motor Impairments’. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1223–1232. ACM. http://dl.acm.org/citation.cfm?id=2466158.

Azenkot, Shiri, and Nicole B. Lee. 2013. ‘Exploring the Use of Speech Input by Blind People on Mobile Devices’. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS ’13*, 1–8. ACM. https://doi.org/10.1145/2513383.2513440.

Azenkot, Shiri, Kyle Rector, Richard Ladner, and Jacob Wobbrock. 2012. ‘PassChords: Secure Multi-Touch Authentication for Blind People’. In *Proc. of ACM SIGACCESS Conference on Computers and Accessibility*, 159–166. ASSETS ’12. https://doi.org/10.1145/2384916.2384945.

Azenkot, Shiri, Jacob O. Wobbrock, Sanjana Prasain, and Richard E. Ladner. 2012. ‘Input Finger Detection for Nonvisual Touch Screen Text Entry in Perkinput’. In *Proc. of Graphics Interface*, 121–129. GI ’12. http://dl.acm.org/citation.cfm?id=2305276.2305297.

Bandura, Albert. 1998. ‘Personal and Collective Efficacy in Human Adaptation and Change’. In *Advances in Psychological Science, Vol. 1: Social, Personal, and Cultural Aspects*, 51–71. Hove, England: Psychology Press/Erlbaum (UK) Taylor & Francis.

Bandura, Albert. 2006. ‘Guide for Constructing Self-Efficacy Scales’. *Self-Efficacy Beliefs of Adolescents* 5 (1): 307–337.

Bandura, Albert. 2010. ‘Self-Efficacy’. In *The Corsini Encyclopedia of Psychology*, 1–3. American Cancer Society. https://doi.org/10.1002/9780470479216.corpsy0836.

Bateman, Daniel Robert, Erin Brady, David Wilkerson, Eun-Hye Yi, Yamini Karanam, and Christopher M. Callahan. 2017. ‘Comparing Crowdsourcing and Friendsourcing: A Social Media-Based Feasibility Study to Support Alzheimer Disease Caregivers’. *JMIR Research Protocols* 6 (4): e56. https://doi.org/10.2196/resprot.6904.

Bigham, Jeffrey P., Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, et al. 2010. ‘VizWiz: Nearly Real-Time Answers to Visual Questions’. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology*, 333–342. ACM. http://dl.acm.org/citation.cfm?id=1866080.

Bigham, Jeffrey P., Richard E. Ladner, and Yevgen Borodin. 2011. ‘The Design of Human-Powered Access Technology’. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*, 3–10. ACM. http://dl.acm.org/citation.cfm?id=2049540.

Bigham, Jeffrey P., Irene Lin, and Saiph Savage. 2017. 'The Effects of "Not Knowing What You Don'T Know" on Web Accessibility for Blind Web Users'. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 101–109. ASSETS '17. New York, NY, USA: ACM. https://doi.org/10.1145/3132525.3132533.

Bonner, Matthew N., Jeremy T. Brudvik, Gregory D. Abowd, and W. Keith Edwards. 2010. 'No-Look Notes: Accessible Eyes-Free Multi-Touch Text Entry'. In *Pervasive Computing*, edited by Patrik Floréen, Antonio Krüger, and Mirjana Spasojevic, 409–26. Lecture Notes in Computer Science. Springer Berlin Heidelberg.

Brady, Erin L., Yu Zhong, Meredith Ringel Morris, and Jeffrey P. Bigham. 2013. 'Investigating the Appropriateness of Social Network Question Asking As a Resource for Blind Users'. In *Proc. of Conference on Computer Supported Cooperative Work*, 1225–1236. CSCW '13. https://doi.org/10.1145/2441776.2441915.

Brady, Erin, Meredith Ringel Morris, Yu Zhong, Samuel White, and Jeffrey P. Bigham. 2013. 'Visual Challenges in the Everyday Lives of Blind People'. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2117–2126. CHI '13. New York, NY, USA: ACM. https://doi.org/10.1145/2470654.2481291.

Braun, Virginia, and Victoria Clarke. 2006. 'Using Thematic Analysis in Psychology'. *Qualitative Research in Psychology* 3 (2): 77–101. https://doi.org/10.1191/1478088706qp063oa.

Brungart, Douglas S., and Brian D. Simpson. 2005. 'Improving Multitalker Speech Communication with Advanced Audio Displays'. AIR FORCE RESEARCH LAB WRIGHT-PATTERSON AFB OH. https://apps.dtic.mil/docs/citations/ADA454531.

Buzzi, Maria Claudia, Marina Buzzi, Barbara Leporini, and Amaury Trujillo. 2014. 'Designing a Text Entry Multimodal Keypad for Blind Users of Touchscreen Mobile Phones'. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility*, 131–136. ASSETS '14. New York, NY, USA: ACM. https://doi.org/10.1145/2661334.2661354.

Buzzi, Maria Claudia, Marina Buzzi, Barbara Leporini, and Amaury Trujillo. 2017. 'Analyzing Visually Impaired People's Touch Gestures on Smartphones'. *Multimedia Tools and Applications* 76 (4): 5141–5169. https://doi.org/10.1007/s11042-016-3594-9.

Chilana, Parmit K., Andrew J. Ko, and Jacob O. Wobbrock. 2012. 'LemonAid: Selection-Based Crowdsourced Contextual Help for Web Applications'. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1549–1558. ACM. http://dl.acm.org/citation.cfm?id=2208620.

Cohen, Jacob. 2013. *Statistical Power Analysis for the Behavioral Sciences*. Routledge. https://doi.org/10.4324/9780203771587.

Damaceno, Rafael Jeferson Pezzuto, Juliana Cristina Braga, and Jesús Pascual Mena-Chalco. 2018. 'Mobile Device Accessibility for the Visually Impaired: Problems Mapping and Recommendations'. *Universal Access in the Information Society* 17 (2): 421–35. https://doi.org/10.1007/s10209-017-0540-1.

Duarte, Carlos, Inês Matos, João Vicente, Ana Salvado, Carlos M. Duarte, and Luís Carriço. 2016. 'Development Technologies Impact in Web Accessibility'. In *Proceedings of the 13th Web for All Conference*, 6:1–6:4. W4A '16. New York, NY, USA: ACM. https://doi.org/10.1145/2899475.2899498.

Fernquist, Jennifer, Tovi Grossman, and George Fitzmaurice. 2011. 'Sketch-Sketch Revolution: An Engaging Tutorial System for Guided Sketching and Application Learning'. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, 373–382. ACM. http://dl.acm.org/citation.cfm?id=2047245.

Fleiss, Joseph L., Bruce Levin, and Myunghee Cho Paik. 2013. *Statistical Methods for Rates and Proportions*. John Wiley & Sons.

Furnas, G. W., T. K. Landauer, L. M. Gomez, and S. T. Dumais. 1987. 'The Vocabulary Problem in Human-System Communication'. *Commun. ACM* 30 (11): 964–971. https://doi.org/10.1145/32206.32212.

Gleason, Cole, Dragan Ahmetovic, Saiph Savage, Carlos Toxtli, Carl Posthuma, Chieko Asakawa, Kris M. Kitani, and Jeffrey P. Bigham. 2018. 'Crowdsourcing the Installation and Maintenance of Indoor Localization Infrastructure to Support Blind Navigation'. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2 (1): 9:1–9:25. https://doi.org/10.1145/3191741.

Gleason, Cole, Anhong Guo, Gierad Laput, Kris Kitani, and Jeffrey P. Bigham. 2016. 'VizMap: Accessible Visual Information Through Crowdsourced Map Reconstruction'. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, 273–274. ASSETS '16. New York, NY, USA: ACM. https://doi.org/10.1145/2982142.2982200.

Grossman, Tovi, and George Fitzmaurice. 2010. 'ToolClips: An Investigation of Contextual Video Assistance for Functionality Understanding'. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1515–1524. ACM. http://dl.acm.org/citation.cfm?id=1753552.

Grossman, Tovi, George Fitzmaurice, and Ramtin Attar. 2009. 'A Survey of Software Learnability: Metrics, Methodologies and Guidelines'. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 649–658. ACM. http://dl.acm.org/citation.cfm?id=1518803.

Grussenmeyer, William, and Eelke Folmer. 2017. 'Accessible Touchscreen Technology for People with Visual Impairments: A Survey'. *ACM Trans. Access. Comput.* 9 (2): 6:1–6:31. https://doi.org/10.1145/3022701.

Guerreiro, João, and Daniel Gonçalves. 2014. 'Text-to-Speeches: Evaluating the Perception of Concurrent Speech by Blind People'. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility*, 169–176. ASSETS '14. New York, NY, USA: ACM. https://doi.org/10.1145/2661334.2661367.

Guerreiro, João, André Rodrigues, Kyle Montague, Tiago Guerreiro, Hugo Nicolau, and Daniel Gonçalves. 2015. 'TabLETS Get Physical: Non-Visual Text Entry on Tablet Devices'. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 39–42. CHI '15. New York, NY, USA: ACM. https://doi.org/10.1145/2702123.2702373.

Guerreiro, Tiago, Hugo Nicolau, and Joaquim A. Jorge. 2008. 'From Tapping to Touching: Making Touch Screens Accessible to Blind Users'. *IEEE Multimedia*, 48–50.

Guo, Anhong, Xiang 'Anthony' Chen, Haoran Qi, Samuel White, Suman Ghosh, Chieko Asakawa, and Jeffrey P. Bigham. 2016. 'VizLens: A Robust and Interactive Screen Reader for Interfaces in the Real World'. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, 651–664. UIST '16. New York, NY, USA: ACM. https://doi.org/10.1145/2984511.2984518.

Gurari, Danna, Qing Li, Abigale J. Stangl, Anhong Guo, Chi Lin, Kristen Grauman, Jiebo Luo, and Jeffrey P. Bigham. 2018. 'VizWiz Grand Challenge: Answering Visual Questions From Blind People'. In , 3608–17. http://openaccess.thecvf.com/content_cvpr_2018/html/Gurari_VizWiz_Grand_Challenge_CVPR_2018_paper.html.

Hagiya, Toshiyuki, Tomonori Yazaki, Toshiharu Horiuchi, and Tsuneo Kato. 2015. 'Typing Tutor: Automatic Error Detection and Instruction in Text Entry for Elderly People'. In , 696–703. ACM Press. https://doi.org/10.1145/2786567.2793690.

Hailpern, Joshua, Loretta Guarino Reid, and Richard Boardman. 2009. 'DTorial: An Interactive Tutorial Framework for Blind Users in a Web 2.0 World'. In *Human-Computer Interaction–INTERACT 2009*, 5–18. Springer. http://link.springer.com/chapter/10.1007/978-3-642-03655-2_3.

Holman, Jonathan, Jonathan Lazar, Jinjuan Heidi Feng, and John D'Arcy. 2007. 'Developing Usable CAPTCHAs for Blind Users'. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, 245–246. ACM. http://dl.acm.org/citation.cfm?id=1296894.

Hong, Hwajung, Jennifer G. Kim, Gregory D. Abowd, and Rosa I. Arriaga. 2012. 'Designing a Social Network to Support the Independence of Young Adults with Autism'. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, 627–636. CSCW '12. New York, NY, USA: ACM. https://doi.org/10.1145/2145204.2145300.

Kane, Shaun K., Jeffrey P. Bigham, and Jacob O. Wobbrock. 2008. 'Slide Rule: Making Mobile Touch Screens Accessible to Blind People Using Multi-Touch Interaction Techniques'. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*, 73–80. Assets '08. New York, NY, USA: ACM. https://doi.org/10.1145/1414471.1414487.

Kane, Shaun K., Chandrika Jayant, Jacob O. Wobbrock, and Richard E. Ladner. 2009. 'Freedom to Roam: A Study of Mobile Device Adoption and Accessibility for People with Visual and Motor Disabilities'. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility*, 115–122. Assets '09. New York, NY, USA: ACM. https://doi.org/10.1145/1639642.1639663.

Kane, Shaun K., Jacob O. Wobbrock, and Richard E. Ladner. 2011. 'Usable Gestures for Blind People: Understanding Preference and Performance'. In *Proc. of SIGCHI Conference on Human Factors in Computing Systems*, 413–422. CHI '11. https://doi.org/10.1145/1978942.1979001.

Kapp, Karl M. 2013. *The Gamification of Learning and Instruction Fieldbook: Ideas into Practice*. 1st ed. Pfeiffer & Company.

Kato, Takashi. 1986. 'What "Question-Asking Protocols" Can Say about the User Interface'. *International Journal of Man-Machine Studies* 25 (6): 659–73. https://doi.org/10.1016/S0020-7373(86)80080-3.

Kelleher, Caitlin, and Randy Pausch. 2005. 'Stencils-Based Tutorials: Design and Evaluation'. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 541–550. ACM. http://dl.acm.org/citation.cfm?id=1055047.

Lafreniere, Benjamin, Tovi Grossman, and George Fitzmaurice. 2013. 'Community Enhanced Tutorials: Improving Tutorials with Multiple Demonstrations'. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1779–1788. ACM. http://dl.acm.org/citation.cfm?id=2466235.

Lakhani, Karim R., and Eric von Hippel. 2004. 'How Open Source Software Works: "Free" User-to-User Assistance'. In *Produktentwicklung Mit Virtuellen Communities: Kundenwünsche Erfahren Und Innovationen Realisieren*, edited by Cornelius Herstatt and Jan G. Sander, 303–39. Wiesbaden: Gabler Verlag. https://doi.org/10.1007/978-3-322-84540-5_13.

Lasecki, Walter S., Phyo Thiha, Yu Zhong, Erin Brady, and Jeffrey P. Bigham. 2013. 'Answering Visual Questions with Conversational Crowd Assistants'. In *Proc. of ACM SIGACCESS Conference on Computers and Accessibility*, 18:1–18:8. ASSETS '13. https://doi.org/10.1145/2513383.2517033.

Lazar, Jonathan, Alfreda Dudley-Sponaugle, and Kisha-Dawn Greenidge. 2004. 'Improving Web Accessibility: A Study of Webmaster Perceptions'. *Computers in Human Behavior*, The Compass of Human-Computer Interaction, 20 (2): 269–88. https://doi.org/10.1016/j.chb.2003.10.018.

Leporini, Barbara, Maria Claudia Buzzi, and Marina Buzzi. 2012. 'Interacting with Mobile Devices via VoiceOver: Usability and Accessibility Issues'. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*, 339–348. OzCHI '12. New York, NY, USA: ACM. https://doi.org/10.1145/2414536.2414591.

Li, Wei, Tovi Grossman, and George Fitzmaurice. 2012. 'Gamicad: A Gamified Tutorial System for First Time Autocad Users'. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, 103–112. ACM. http://dl.acm.org/citation.cfm?id=2380131.

Lieberman, Henry, Elizabeth Rosenzweig, and Christopher Fry. 2014. 'Steptorials: Mixed-Initiative Learning of High-Functionality Applications'. In , 359–64. ACM Press. https://doi.org/10.1145/2557500.2557543.

Lynch, Tony, and Joan Maclean. 2000. 'Exploring the Benefits of Task Repetition and Recycling for Classroom Language Learning'. *Language Teaching Research* 4 (3): 221–50. https://doi.org/10.1177/136216880000400303.

MacKenzie, I. Scott, and Shawn X. Zhang. 2001. 'An Empirical Investigation of the Novice Experience with Soft Keyboards'. *Behaviour & Information Technology* 20 (6): 411–18. https://doi.org/10.1080/01449290110089561.

Maes, Pattie. 1995. 'Agents That Reduce Work and Information Overload'. In *Readings in Human–Computer Interaction*, edited by RONALD M. Baecker, JONATHAN Grudin, WILLIAM A. S. Buxton, and SAUL Greenberg, 811–21. Interactive Technologies. Morgan Kaufmann. https://doi.org/10.1016/B978-0-08-051574-8.50084-4.

Mahendru, Aroma, Viraj Prabhu, Akrit Mohapatra, Dhruv Batra, and Stefan Lee. 2017. 'The Promise of Premise: Harnessing Question Premises in Visual Question Answering'. *ArXiv:1705.00601 [Cs]*, May. http://arxiv.org/abs/1705.00601.

Manduchi, R., J. Coughlan, and V. Ivanchenko. 2008. 'Search Strategies of Visually Impaired Persons Using a Camera Phone Wayfinding System'. In *Computers Helping People with Special Needs*, edited by Klaus Miesenberger, Joachim Klaus, Wolfgang Zagler, and Arthur Karshmer, 1135–40. Lecture Notes in Computer Science. Springer Berlin Heidelberg.

Marques, Diogo, Luís Carriço, and Tiago Guerreiro. 2015. 'Assessing Inconspicuous Smartphone Authentication for Blind People'. *ArXiv:1506.00930 [Cs]*, June. http://arxiv.org/abs/1506.00930.

Martins, João, José Carilho, Oliver Schnell, Carlos Duarte, Francisco M. Couto, Luís Carriço, and Tiago Guerreiro. 2014. 'Friendsourcing the Unmet Needs of People with Dementia'. In *Proceedings of the 11th Web for All Conference,* 35. ACM. http://dl.acm.org/citation.cfm?id=2596716.

McGookin, David, Stephen Brewster, and WeiWei Jiang. 2008. 'Investigating Touchscreen Accessibility for People with Visual Impairments'. In *Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges,* 298–307. NordiCHI '08. New York, NY, USA: ACM. https://doi.org/10.1145/1463160.1463193.

Montague, Kyle. 2012. 'Interactions Speak Louder than Words: Shared User Models and Adaptive Interfaces'. In *Adjunct Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology,* 39–42. ACM. http://dl.acm.org/citation.cfm?id=2380315.

Montague, Kyle, Hugo Nicolau, and Vicki L. Hanson. 2014. 'Motor-Impaired Touchscreen Interactions in-the-wild'. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility,* 123–130. ASSETS '14. New York, NY, USA: ACM. https://doi.org/10.1145/2661334.2661362.

Montague, Kyle, André Rodrigues, Hugo Nicolau, and Tiago Guerreiro. 2015. 'TinyBlackBox: Supporting Mobile In-The-Wild Studies'. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility,* 379–380. ASSETS '15. New York, NY, USA: ACM. https://doi.org/10.1145/2700648.2811379.

Naftali, Maia, and Leah Findlater. 2014. 'Accessibility in-context: Understanding the Truly Mobile Experience of Smartphone Users with Motor Impairments'. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility,* 209–216. ASSETS '14. New York, NY, USA: ACM. https://doi.org/10.1145/2661334.2661372.

Narasimhan, Priya, Rajeev Gandhi, and Dan Rossi. 2009. 'Smartphone-Based Assistive Technologies for the Blind'. In *Proceedings of the 2009 International Conference on Compilers, Architecture, and Synthesis for Embedded Systems*, 223–232. CASES '09. New York, NY, USA: ACM. https://doi.org/10.1145/1629395.1629427.

Nicolau, Hugo, João Guerreiro, Tiago Guerreiro, and Luís Carriço. 2013. 'UbiBraille: Designing and Evaluating a Vibrotactile Braille-Reading Device'. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '13*, 1–8. Bellevue, Washington: ACM Press. https://doi.org/10.1145/2513383.2513437.

Nicolau, Hugo, Kyle Montague, Tiago Guerreiro, João Guerreiro, and Vicki L. Hanson. 2014. 'B#: Chord-Based Correction for Multitouch Braille Input'. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1705–1708. CHI '14. New York, NY, USA: ACM. https://doi.org/10.1145/2556288.2557269.

Nicolau, Hugo, Kyle Montague, Tiago Guerreiro, André Rodrigues, and Vicki L. Hanson. 2015a. 'HoliBraille: Multipoint Vibrotactile Feedback on Mobile Devices'. In *Proceedings of the 12th Web for All Conference*, 30:1–30:4. W4A '15. New York, NY, USA: ACM. https://doi.org/10.1145/2745555.2746643.

Nicolau, Hugo, Kyle Montague, Tiago Guerreiro, André Rodrigues, and Vicki L. Hanson. 2015b. 'Typing Performance of Blind Users: An Analysis of Touch Behaviors, Learning Effect, and In-Situ Usage'. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, 273–280. ASSETS '15. New York, NY, USA: ACM. https://doi.org/10.1145/2700648.2809861.

Nicolau, Hugo, Kyle Montague, Tiago Guerreiro, André Rodrigues, and Vicki L. Hanson. 2017. ‘Investigating Laboratory and Everyday Typing Performance of Blind Users’. *ACM Trans. Access. Comput.* 10 (1): 4:1–4:26. https://doi.org/10.1145/3046785.

Oh, Uran, Shaun K. Kane, and Leah Findlater. 2013. ‘Follow That Sound: Using Sonification and Corrective Verbal Feedback to Teach Touchscreen Gestures’. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, 13:1–13:8. ASSETS ’13. New York, NY, USA: ACM. https://doi.org/10.1145/2513383.2513455.

Oliveira, Joao, Tiago Guerreiro, Hugo Nicolau, Joaquim Jorge, and Daniel Gonçalves. 2011. ‘Blind People and Mobile Touch-Based Text-Entry: Acknowledging the Need for Different Flavors’. In *Proc. of ACM SIGACCESS Conference on Computers and Accessibility*, 179–186. ASSETS ’11. https://doi.org/10.1145/2049536.2049569.

Oliveira, João, Tiago Guerreiro, Hugo Nicolau, Joaquim Jorge, and Daniel Gonçalves. 2011. ‘BrailleType: Unleashing Braille over Touch Screen Mobile Phones’. In *Human-Computer Interaction – INTERACT 2011*, edited by Pedro Campos, Nicholas Graham, Joaquim Jorge, Nuno Nunes, Philippe Palanque, and Marco Winckler, 100–107. Berlin, Heidelberg: Springer Berlin Heidelberg.

Pal, Joyojeet, Anandhi Viswanathan, Priyank Chandra, Anisha Nazareth, Vaishnav Kameswaran, Hariharan Subramonyam, Aditya Johri, Mark S. Ackerman, and Sile O’Modhrain. 2017. ‘Agency in Assistive Technology Adoption: Visual Impairment and Smartphone Use in Bangalore’. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 5929–5940. CHI ’17. New York, NY, USA: ACM. https://doi.org/10.1145/3025453.3025895.

Postill, John, and Sarah Pink. 2012. ‘Social Media Ethnography: The Digital Researcher in a Messy Web’. *Media International Australia* 145 (1): 123–34. https://doi.org/10.1177/1329878X1214500114.

Prabhu, Neiman Stern. 1987. *Second Language Pedagogy*. Vol. 20. Oxford University Press Oxford.

Quinn, Alexander J., and Benjamin B. Bederson. 2011. ‘Human Computation: A Survey and Taxonomy of a Growing Field’. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1403–1412. CHI ’11. New York, NY, USA: ACM. https://doi.org/10.1145/1978942.1979148.

Richards, John T., Kyle Montague, and Vicki L. Hanson. 2012. ‘Web Accessibility As a Side Effect’. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, 79–86. ASSETS ’12. New York, NY, USA: ACM. https://doi.org/10.1145/2384916.2384931.

Rodrigues, André, Leonardo Camacho, Hugo Nicolau, Kyle Montague, and Tiago Guerreiro. 2018. ‘Aidme: Interactive Non-Visual Smartphone Tutorials’. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, 205–212. MobileHCI ’18. New York, NY, USA: ACM. https://doi.org/10.1145/3236112.3236141.

Rodrigues, André, Kyle Montague, and Tiago Guerreiro. 2018. ‘Data Donors: Sharing Knowledge for Mobile Accessibility’. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, LBW539:1–LBW539:6. CHI EA ’18. New York, NY, USA: ACM. https://doi.org/10.1145/3170427.3188627.

Rodrigues, André, Kyle Montague, Hugo Nicolau, João Guerreiro, and Tiago Guerreiro. 2017. 'In-context Q&A to Support Blind People Using Smartphones'. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 32–36. ASSETS '17. New York, NY, USA: ACM. https://doi.org/10.1145/3132525.3132555.

Rodrigues, André, Kyle Montague, Hugo Nicolau, and Tiago Guerreiro. 2015. 'Getting Smartphones to Talkback: Understanding the Smartphone Adoption Process of Blind Users'. In *Proc. of ACM SIGACCESS Conference on Computers and Accessibility*, 23–32. ASSETS '15. https://doi.org/10.1145/2700648.2809842.

Rodrigues, André, Hugo Nicolau, Kyle Montague, Luís Carriço, and Tiago Guerreiro. 2016. '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*, 47–52. MobileHCI '16. New York, NY, USA: ACM. https://doi.org/10.1145/2935334.2935376.

Rodrigues, André, André Santos, Kyle Montague, and Tiago Guerreiro. 2017. 'Improving Smartphone Accessibility with Personalizable Static Overlays'. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 37–41. ASSETS '17. New York, NY, USA: ACM. https://doi.org/10.1145/3132525.3132558.

Rodrigues, André, André Santos, Kyle Montague, Hugo Nicolau, and Tiago Guerreiro. 2019. 'Understanding the Authoring and Playthrough of Nonvisual Smartphone Tutorials'. In *Human-Computer Interaction – INTERACT 2019*, edited by David Lamas, Fernando Loizides, Lennart Nacke, Helen Petrie, Marco Winckler, and Panayiotis Zaphiris, 42–62. Lecture Notes in Computer Science. Springer International Publishing.

Rogstadius, Jakob, Vassilis Kostakos, Aniket Kittur, Boris Smus, Jim Laredo, and Maja Vukovic. 2011. 'An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets'. In *Fifth International AAAI Conference on Weblogs and Social Media*. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2778.

Romero, Mario, Brian Frey, Caleb Southern, and Gregory D. Abowd. 2011. 'BrailleTouch: Designing a Mobile Eyes-Free Soft Keyboard'. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, 707–709. MobileHCI '11. New York, NY, USA: ACM. https://doi.org/10.1145/2037373.2037491.

Ross, Anne Spencer, Xiaoyi Zhang, James Fogarty, and Jacob O. Wobbrock. 2017. 'Epidemiology As a Framework for Large-Scale Mobile Application Accessibility Assessment'. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 2–11. ASSETS '17. New York, NY, USA: ACM. https://doi.org/10.1145/3132525.3132547.

Sato, Daisuke, Hironobu Takagi, Masatomo Kobayashi, Shinya Kawanaka, and Chieko Asakawa. 2010. 'Exploratory Analysis of Collaborative Web Accessibility Improvement'. *ACM Transactions on Accessible Computing* 3 (2): 1–30. https://doi.org/10.1145/1857920.1857922.

Seedhouse, Paul. 1999. 'Task-Based Interaction'. *ELT Journal* 53 (3): 149–56. https://doi.org/10.1093/elt/53.3.149.

Seltman, Howard J. 2012. 'Experimental Design and Analysis', 428.

Shinohara, Kristen, and Jacob O. Wobbrock. 2011. 'In the Shadow of Misperception: Assistive Technology Use and Social Interactions'. In *Proceedings of the SIGCHI*

Conference on Human Factors in Computing Systems, 705–714. CHI ’11. New York, NY, USA: ACM. https://doi.org/10.1145/1978942.1979044.

Skehan, Peter. 2003. ‘Task-Based Instruction’. *Language Teaching* 36 (1): 1–14. https://doi.org/10.1017/S026144480200188X.

Southern, Caleb, James Clawson, Brian Frey, Gregory Abowd, and Mario Romero. 2012. ‘An Evaluation of BrailleTouch: Mobile Touchscreen Text Entry for the Visually Impaired’. In *Proc. of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, 317–326. MobileHCI ’12. https://doi.org/10.1145/2371574.2371623.

Takagi, Hironobu, Shinya Kawanaka, Masatomo Kobayashi, Takashi Itoh, and Chieko Asakawa. 2008. ‘Social Accessibility: Achieving Accessibility through Collaborative Metadata Authoring’. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*, 193–200. ACM. http://dl.acm.org/citation.cfm?id=1414507.

Trindade, Daniel, André Rodrigues, Tiago Guerreiro, and Hugo Nicolau. 2018. ‘Hybrid-Brailler: Combining Physical and Gestural Interaction for Mobile Braille Input and Editing’. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 27:1–27:12. CHI ’18. New York, NY, USA: ACM. https://doi.org/10.1145/3173574.3173601.

Vega, Julio. 2019. ‘Unobtrusive and Personalised Monitoring of Parkinson’s Disease Using Smartphones’. PhD Thesis. https://doi.org/10.5281/zenodo.3242597.

Vigo, Markel, and Simon Harper. 2013. ‘Coping Tactics Employed by Visually Disabled Users on the Web’. *International Journal of Human-Computer Studies* 71 (11): 1013–25. https://doi.org/10.1016/j.ijhcs.2013.08.002.

Wang, Cheng-Yao, Wei-Chen Chu, Hou-Ren Chen, Chun-Yen Hsu, and Mike Y. Chen. 2014. ‘EverTutor: Automatically Creating Interactive Guided Tutorials on Smartphones by User Demonstration’. In , 4027–36. ACM Press. https://doi.org/10.1145/2556288.2557407.

Weber, Gerhard, Helen Petrie, Dirk Kochanek, and Sarah Morley. 1994. ‘Training Blind People in the Use of Graphical User Interfaces’. In *Computers for Handicapped Persons*, edited by Wolfgang L. Zagler, Geoffrey Busby, and Roland R. Wagner, 25–31. Berlin, Heidelberg: Springer Berlin Heidelberg.

Widdowson, H. G. 1978. *Teaching Language as Communication*. OUP Oxford.

Wobbrock, Jacob O., and Brad A. Myers. 2006. ‘Analyzing the Input Stream for Character- Level Errors in Unconstrained Text Entry Evaluations’. *ACM Transactions on Computer-Human Interaction (TOCHI)* 13 (4): 458–89. https://doi.org/10.1145/1188816.1188819.

Yi, Mun Y, and Yujong Hwang. 2003. ‘Predicting the Use of Web-Based Information Systems: Self-Efficacy, Enjoyment, Learning Goal Orientation, and the Technology Acceptance Model’. *International Journal of Human-Computer Studies*, Zhang and Dillon Special Issue on HCI and MIS, 59 (4): 431–49. https://doi.org/10.1016/S1071-5819(03)00114-9.

Zhang, Dongsong, Lina Zhou, Judith O. Uchidiuno, and Isil Y. Kilic. 2017. ‘Personalized Assistive Web for Improving Mobile Web Browsing and Accessibility for Visually Impaired Users’. *ACM Trans. Access. Comput.* 10 (2): 6:1–6:22. https://doi.org/10.1145/3053733.

Zhang, Xiaoyi, Anne Spencer Ross, Anat Caspi, James Fogarty, and Jacob O. Wobbrock. 2017. ‘Interaction Proxies for Runtime Repair and Enhancement of Mobile Application Accessibility’. In *Proceedings of the 2017 CHI Conference on*

Human Factors in Computing Systems, 6024–6037. CHI ’17. New York, NY, USA: ACM. https://doi.org/10.1145/3025453.3025846.

Zhang, Xiaoyi, Tracy Tran, Yuqian Sun, Ian Culhane, Shobhit Jain, James Fogarty, and Jennifer Mankoff. 2018. ‘Interactiles: 3D Printed Tactile Interfaces to Enhance Mobile Touchscreen Accessibility’. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, 131–142. ACM.

Zhong, Yu, T. V. Raman, Casey Burkhardt, Fadi Biadsy, and Jeffrey P. Bigham. 2014. ‘JustSpeak: Enabling Universal Voice Control on Android’. In *Proceedings of the 11th Web for All Conference*, 36:1–36:4. W4A ’14. New York, NY, USA: ACM. https://doi.org/10.1145/2596695.2596720.