Understanding the Authoring and Playthrough of Nonvisual Smartphone Tutorials*

André Rodrigues[1], André Santos[1], Kyle Montague[2], Hugo Nicolau[3], and Tiago Guerreiro[1]

[1] LASIGE, Faculdade de Ciências, Universidade de Lisboa
afrodrigues@fc.ul.pt ,abranco@lasige.di.fc.ul.pt, tjvg@di.fc.ul.pt
[2] Open Lab, Newcastle University
kyle.montague@newcastle.ac.uk
[3] INESC-ID, Instituto Superior Técnico, Universidade de Lisboa
hman@inesc-id.pt

Abstract. Mobile device users are required to constantly learn to use new apps, features, and adapt to updates. For blind people, adapting to a new interface requires additional time and effort. At the limit, and often so, devices and applications may become unusable without support from someone else. Using tutorials is a common approach to foster independent learning of new concepts and workflows. However, most tutorials available online are limited in scope, detail, or quickly become outdated. Also, they presume a degree of tech savviness that is not at the reach of the common mobile device user. Our research explores the democratization of assistance by enabling non-technical people to create tutorials in their mobile phones for others. We report on the interaction and information needs of blind people when following 'amateur' tutorials. Thus, providing insights into how to widen and improve the authoring and playthrough of these learning artifacts. We conducted a study where 12 blind users followed tutorials previously created by blind or sighted people. 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 do not align. We provide insights on how to support both authoring and playthrough of nonvisual smartphone tutorials. There is an opportunity to design solutions that mediate authoring, combine contributions, adapt to user profile, react to context and are living artifacts capable of perpetual improvement.

Keywords: Tutorials· Blind· Smartphones· Accessibility· Assistance.

* We thank Fundação Raquel and Martin Sain in Lisbon (Portugal) and all participants. This work was partially supported by Fundação para a Ciência e Tecnologia (FCT) through scholarship SFRH/BD/103935/2014, funding of project mIDR (AAC 02/SAICT/- 2017), INESC-ID research unit UID/CEC/50021/2019, and LASIGE Research Unit, ref. UID/CEC/00408/2019.

1 Introduction

Mobile devices are pivotal tools for inclusion and independence. The inability to operate them proficiently, and quickly adapt to new usages, is likely to have deep social and professional repercussions. Blind people encounter challenges when first adopting these devices and with each new application and update [23]. The wide variety of interface templates and workflows mandates users to create different mental models and constantly adapt to them. Moreover, users are often confronted with accessibility issues that, without assistance, render the app, feature, or even device, inaccessible. In prior work [23, 22], users reported to primarily rely on the assistance of others, when they felt helpless with their smartphones. However, depending on the issue, their support network might not be able to help effectively. Sighted users with no accessibility experience often could not cope with the interaction changes caused by the assistive technology [23]. In contrast, for issues relating to missing labels or hidden screens, sighted assistance is required. Nevertheless, there are situations whereby knowledgeable users, capable of assistance, might not always be available and co-located.

Another key strategy to overcome technology challenges is to browse the web for answers and guidance. Possible knowledge sources are online video tutorials (e.g. on YouTube), including channels dedicated to accessibility such as the Tech Accessibility Tutorial [28], where users can listen to tutorials on a variety of tasks. Alternatively, people can resort to dedicated fora and mailing lists, for example AppleVis[1] or Eyes-Free [6]. Despite the availability of tutorials, it is not always easy or possible for blind users to translate the content to their own devices or settings. Less tech savvy users often lack the initiative to search for content when its not where they expect it. Equally, they do not necessarily ask the right questions, or possess the ability to understand the answers and know how to apply them to their specific context [22].

It is commonplace for applications and OS to have onboarding tutorials that guide users through their core features, thus supporting only 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 tutorials 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 [10].

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 report on a 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.

Our contributions include: 1) understanding of the instructions given by sighted and blind people when creating nonvisual tutorials; and 2) the interaction and information needs of blind people when following tutorials created by other users.

2 Related Work

In this section, we discuss previous work in four topics: nonvisual mobile accessibility and its open challenges; current support mechanisms; in-context assistance; and attempts to democratize assistance in other contexts.

2.1 Nonvisual Mobile Accessibility

Researchers have quickly become aware of the opportunities and challenges smartphones could bring to blind people. There has been a large body of work focusing on the interaction challenges of smartphones. At first, there was a focus on enabling access to surfaces with no physical buttons, and novel interaction techniques emerged. Kane et al. [14] developed and evaluated a set of multi-touch interaction techniques to provide nonvisual access to multiple applications. Nowadays, the main smartphone OS come packaged with a native screen reader (e.g., VoiceOver on iOS, Talkback on Android). Users perform directional swipes to navigate through content or rely on Explore by Touch by dragging one finger around the screen while elements are read aloud; to select, users can double tap anywhere on the screen. Since one of the first challenges is learning how to perform gestures on touchscreens, prior work has investigated how sonification can be used to support gesture learning [20].

Despite the efforts, gaining proficiency with such devices is still a challenge. In an eight-week study of the smartphone adoption process by blind people [23], Rodrigues et al. found that learning how to use the device is still an arduous task that most often requires help from peers. Others have looked at the open challenges touchscreen interactions create for blind people [12], which go beyond touch, gestures, and text-entry. In Rodrigues et al. [22], a series of workshops on smartphones, featuring users of varied expertise levels, confirmed difficulties beyond touchscreen interaction and the need to rely on others to surpass them.

Challenges often spawn from the overwhelming number of apps and features users need to adapt to. A possible solution is to simplify the interaction by replacing the entire system with a single application (e.g. Mobile Accessibility [5]) guaranteeing consistency, coherence of layouts, and app behaviors. However, it has the severe drawback of negating the benefits of all other apps available, and to disallow people with disabilities to have access to the same applications as others [25]. Another option is to adapt how content is rendered and navigated to maximize its accessibility [30]. Zhang et al. [30] proposes the use of interactions proxies to allow third-party developers to address apps accessibility issues. However, for a successful adaptation one must be aware of each applications issues and adapt them to the end users 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 [23].

2.2 Supporting Smartphone Learning

Aware of the differences and difficulties of learning how to use a smartphone, both VoiceOver and Talkback feature training tutorials for their users. VoiceOver provides a training canvas for gestures and relies on customer service to get users started with the device. On the other hand, Talkback has five interactive lessons to cover the basics of screen reader usage: Basic Navigation, Scrolling, Talkback Menus, Text Navigation and Text Editing. The tutorials available are limited in their scope, only addressing gestures and screen reader control. There is a lack of assistance in guiding the users through their holistic needs from basic navigation to complex task guidance (e.g. 'Add a contact', 'Forward a message').

Previous work found that blind people prefer the assistance of peers who are familiar with assistive technologies (e.g., screen readers), and have overcome similar issues [22]. Despite such perception, some screen information is likely to be accessible solely via visual feedback (e.g. describing layout or identifying a missing tag). To our knowledge, instruction-giving has not been explored in a nonvisual smartphone usage setting.

2.3 In-Context Assistance

Providing inline or in-context interactive assistance can facilitate users learning process, as it has been shown in desktop applications [10, 16]. Kelleher et al. [16] investigated an interaction technique for presenting in-context tutorials. The proposed technique, now commonly used in smartphone onboarding tutorials, features an overlay to obscure the non-relevant content and restrict user interaction. This approach has showed to be more efficient than traditional tutorials, reducing errors, time, and required assistance [16].

Interactive tutorials can be applied to a variety of contexts and population. Hagiya et al. [13] reported a text-entry tutorial for older adults that detects errors and provides instructions to correct them. Also, it detects when the user is taking too long to type the next letter/word and provides instructions simultaneously through voice, text and finger animations. In a study with 28 elderly participants, the tutoring system significantly increased typing speed (by 17%) and reduced errors (by 59%).

Tutorials can be designed to be engaging experiences to boost user performance. In Fernquist et al. [7], the system guides users through the interface providing assistance while sketching. Using a step navigation dialog, in each step, the user is shown how, where, and when to change settings, as well as when and where to draw. Yet, interactive tutorials should not restrict users of following the steps. In Lieberman et al. [18], the authors argue that at various points during the tutorial, users may require different levels of assistance (e.g. let me do it, show me how, and guide me through it). At each step, users can delve into its particularities and freely navigate between steps.

As for interactive assistance, in-app onboarding has become more common with apps, guiding users on their first interactions. Unfortunately, they are limited in their scope, only supporting first usage, heavily relying on restricting users interactions, and conveying instructions through visual metaphors. In EverTutor [29], researchers have investigated how to broaden the reach of interactive tutorials by allowing the creation of system-wide tutorials from user demonstration. When a tutorial is played, it uses overlays with visual metaphors, and an obscuring overlay to convey the next target and gesture in-context; additionally, it prevents users from performing incorrect steps.

To our knowledge, there are no tools with the ability to create nonvisual interactive tutorials on smartphones nor studies on how blind people cope with the instructions provided. Moreover, there are no insights into their efficacy nor how to design them.

2.4 Democratizing Assistance

The limitations in the success of the available tutorials on commodity smartphones put forward a stereotypical and limited view on the challenges smartphones impose to blind people. The variety of applications, and the complexity of their interactions and workflows, require for support to be flexible. One solution found by tech savvy users is to resort to online sources, for example, posing a question in a dedicated forum. There have been efforts to allow for this type of flexible on-demand support building on contributions from volunteers and/or crowd workers [2, 22, 24, 26]. Bigham et al. [2] developed VizWiz, a mobile application that enabled users to ask visual questions by taking a photo and have crowd-workers answer in nearly real-time. In a follow-up study, researchers explored how blind users were taking advantage of the platform to ask for assistance with a variety of real-world tasks [3].

In the context of web accessibility, the Social Accessibility project [26] has shown how to take advantage of a network of domain knowledgeable sighted volunteers to provide document metadata (e.g. image labeling, document structure). After 20 months, the project had 350 volunteers that created about 19,000 metadata items for over 3000 web pages [24], revealing the potential of human powered approaches for assistance reliant on volunteers. In contrast with previous work, the Social Accessibility project can aggregate and reutilize the knowledge to assist others without the need of additional interventions by volunteers. Past work has explored how an in-context Q&A system relied on sighted volunteers to provide answers to support nonvisual access [22]. Insights from that work rallied for support tools that promote self-organized learning. Moreover, it 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 the need to better understand how those tools can be designed to be effective. In this paper, we examine how blind and sighted people provide instructions in this context and how they support, or fail in doing so, blind people in performing smartphone tasks.

3 Methodology

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 the intended audience were blind users. Sighted participants were given a set of tips (discussed in the following section) that were previously 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 not what was missing.

In a preliminary study [21], 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 that acted in place of an ideal interactive tutorial similar to the question-asking protocol introduced by Kato [15]. 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 learnibility issues[11]. We adopted the approach, and in addition, we observed and analyzed 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.

3.1 Task Design

Participants created tutorials for six tasks (T) during the authoring session. Two were training tasks (TT) and were designed by the research team. In an effort 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. The tasks apps and description were as follow: TT1 *SimpleNote* - Delete an existing note; TT2 *SimpleNote* - Share existing note on WhatsApp; T1 *Settings* - Clean data from an app; T2 *Messages* - Forward an SMS; T3 *WhatsApp* - Create a group chat; T4 *RadioNet* - Add a station to favorites.

Tips for Accessible Tutorial Authoring Sighted participants were informed the tutorials were 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 have the ability to ask questions and explain their requirements. On the other hand, for assistance provided through technology (e.g. tutorials, Q&A [22, 26]) there is an opportunity to inform helpers of the user requirements. To this end, we had a session with two blind IT instructors where we devised a description of a mobile 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, an accessibility software that allows blind people to interact with touchscreens. When active they change the way users interact with the device. To navigate, taps that used to select options now focus the tapped element and read it using text-to-speech technology. Alternatively, users can swipe left or right to change focus to the next or previous element. To select, instead of a tap, users need to 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;

3.2 Authoring Tutorials towards Democratization

Users created tutorials while performing the task, first they described a step and then they perform 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. 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[29, 17].

3.3 Participants

For the authoring session, we recruited five sighted participants with no previous screen reader experience, ages ranging between 19 and 23 (M=20.8, SD=1.64) years old, three Android users and two iOS, experience between 3 and 4 years; and 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 [22]. 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), 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 task requirements to be experienced users, and two novice users that currently only place/receive calls and receive messages. Only one participant had previously done the forward task in an Android device. There were three participants that had previously forwarded a message but on iOS devices. Additionally, one had previously created a group in WhatsApp. In the following sections we will refer to playthrough participants as Consumers.

3.4 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.

3.5 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 tips aforementioned. Prior to creating a tutorial for each task, users were instructed to explore and perform it. 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 users opinions about the authoring process.

Pre-Processing Content For the playthrough session, we discarded four tutorials for having missteps (i.e. a 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 answered 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 assitance. The only limitations imposed were: (1) the researcher could not physically assist in any way; (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. We had a debriefing session whereby participants could discuss the experienced challenges and provide insights on possible features. Finally, we asked Consumers what information they believed to be essential to a instruction.

3.6 Data & Analysis

We conducted a thematic analysis leveraging the flexibility of the method in reflecting over the data collected [4]. 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, and merged. Another set of ten tutorials were coded independently and reached a Cohens Kappa agreement of k=0.82. The final codebook is shown on Table 1 aggregated by theme.

During the playthrough session, a researcher was observing interactions, behaviors, 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 that 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.

Table 1. Code frequency in the tutorials instructions.

Codes	Type					Screen		
	Visual	**Text**	**Incorrect Text**	**Function**	**Type**	**Function Description**	**Layout Description**	
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 Explanation**	**Absolute**	**Relative**	**Hierarchical Functional**	**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

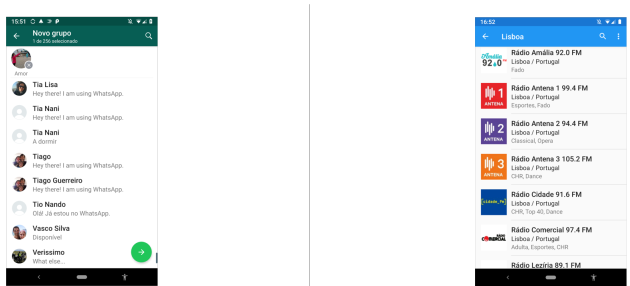

Fig. 1. Two examples of two steps described by four different authors. Two blind authors (BA) and two sighted (SA).

3.7 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 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, requesting 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 (Table 2). The discussion on each topic is also supported by the analysis conducted on the tutorial instructions (Table 1). 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.

Table 2. Code Frequencies of the major categories of issues found during playthrough. A single step may possess multiple overlapping categories.

Category	Instructions Content	Gesture & Navigation	Location & Layout	Tutorial Feedback
In steps (%)	21.3	13.3	15.8	17.5

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): the sighed Authors did not know what to call the confirmation button, a green arrow, 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 1. Blind Authors provided additional information about the current state of the tutorial and its overall goal (33% and 9% of instructions respectively)1. Despite being more verbose, only 8% of blind people's instructions referenced any kind of audio feedback. *"It will say in all of them the radio logo, radio followed by the name." [B1]* Moreover, none described any type of audio cue.

Target Description. Although most instructions had text descriptions, 19% of the ones provided by sighted people did not match the item label. Which is to be expected given the known issues with the variability of vocabulary used by people when interacting with systems [8]. 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 behaviors.

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 Authors contained a brief explanation on how to perform a particular action. *"I am going to keep it pressed" [S2].* On the other hand, 52% of blind authored instructions contained the additional information on how to perform a gesture. *"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].* However, how to instruct the user raised some questions.

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 Swipe Navigation). In multiple instances, Consumers tried using an unfamiliar method with no success. Moreover, they were convinced that since the tutorial instructed them to do so, it was the only way to reach the target. Thus, participants of both methods 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 double tap with a long press on the second tap; without one is 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 behaviors 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. Moreover, 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.

Location & Layout In 38 (15.8%) of steps, Consumers required some 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% (absolute 19%, relative 13%, and hierarchical 11%) of instructions by sighted Authors came with location, many of them were inadequate and even misleading *"I am going to click in the OK that is on the bottom right corner of the pop-up" [S4]*. In this example, while location was provided, the Consumer was unaware of the location and dimensions of the pop-up menu.

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 rely on mixed interaction methods wanted to optimize their navigation behaviors. 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 behaviors.

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 quickly grasp a screen structure. From the 128 instructions given by sighted Authors only two attempted to describe screen layout. In contrast, for a blind Author to describe a layout he/she must first explore all the interface. Even so, three of the 128 instructions, by blind Authors, contained layout descriptions.

Consumers asked how the content 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 has been previously reported in Vigo et al. [27] on the coping tactics employed by visually impaired people on the web, Consumers asked for reassurance and confirmation. Confirming 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 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 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 understand how 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; struggled to provide descriptions to visual elements; and at times, even to provide location. *"I will press again (hadnt mentioned or previously pressed that button) in the button on the right line below, in the bottom right corner." [S2].*

4 Discussion

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

4.1 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. [18] 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 the authoring process to be able to develop flexible playthrough tools. This may come with the cost of overburdening authors; thus, we must work towards solutions that support the authoring process.

4.2 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 in 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 provide answers to visual questions [2]. 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. However, we believe a more scalable approach can 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 in steps and prompt people (Authors) to provide different types of information in small tasks (e.g. Item location, Layout Structure), without any training or particular understanding of the underlying requirements; similar to what has been previously proposed and achieve by Gleason et. al [9] in enabling non-expert participation in the installation and maintenance of indoor localization infrastructure.

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. Moreover, 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 [17], consumers interactions with content can also be leveraged to enhance the content provided (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.

4.3 Account for user expertise, behaviors 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 [19] has already explored continuously updating models based on current behaviors, leveraging its information 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 behaviors (i.e. past and current); navigation pattern detection; content instructions and personalization.

4.4 Flexible instructions & App Modeling

If we can collect different types of information and develop solutions that take into consideration user interaction behaviors and immediate needs, we will be able to provide flexible instructions. As recognized by Lieberman et al. [18], at different steps users may require different levels of guidance. By default, users should be able to access all types of instructions during playthrough by request or based on triggering interactions. Moreover, 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. Moreover, 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 for third party assistive technologies. However, if we can model application structures and workflows we will be in a position to creater smarter assistive tools.

4.5 Enable dialog, a fallback mechanism.

The previous considerations stem from the unpredictability of the user 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 preemptively prepare for them. As our findings suggest, invariable instructions were not enough. With adapted solutions we might get closer to fully automatize assistance. However, we believe the only answer to byzantine problems is to rely on others once more. To do so, we can create solutions that take advantage of peers/crowdworkers/users 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 artifacts capable of perpetual improvement.

4.6 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 Conclusion

We inform future work on the design of solutions that rely on untrained individuals to provide asynchronous technical assistance. We believe both to be valuable contributions for the community in future efforts to design assistive technologies that empower and enable peer support.

We identified the different information required by users during playthrough when following instructions by others. 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. When following a tutorial, the differences in users expertise, interaction behaviors 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.

Future research should seek 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-specialists; 2) the limitations of rigid instructions by looking into novel playthrough mechanisms for nonvisual interactive tutorials.

References

1. AppleViz: iOS blind and low-vision support community (2018), accessed Jan 3 2019. https://www.applevis.com/
2. Bigham, J.P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R.C., Miller, R., Tatarowicz, A., White, B., White, S., Yeh, T.: Vizwiz: Nearly real-time answers to visual questions. In: Proc. of ACM Symposium on User Interface Software and Technology. pp. 333–342. UIST '10 (2010). https://doi.org/10.1145/1866029.1866080, http://doi.acm.org/10.1145/1866029.1866080
3. Brady, E., Morris, M.R., Zhong, Y., White, S., Bigham, J.P.: Visual challenges in the everyday lives of blind people. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 2117–2126. CHI '13, ACM, New York, NY, USA (2013). https://doi.org/10.1145/2470654.2481291, http://doi.acm.org/10.1145/2470654.2481291
4. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qualitative Research in Psychology **3**(2), 77–101 (2006). https://doi.org/10.1191/1478088706qp063oa, http://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa
5. CodeFactory: Mobile Accessibility (2011), accessed May 17 2018. http://codefactoryglobal.com/app-store/mobile-accessibility/
6. Eyes-Free: eyes free (2010), accessed January 3 2019. https://groups.google.com/forum/#!forum/eyes-free
7. Fernquist, J., Grossman, T., Fitzmaurice, G.: 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. pp. 373–382. UIST '11, ACM, New York, NY, USA (2011). https://doi.org/10.1145/2047196.2047245, http://doi.acm.org/10.1145/2047196.2047245
8. Furnas, G.W., Landauer, T.K., Gomez, L.M., Dumais, S.T.: The vocabulary problem in human-system communication. Commun. ACM **30**(11), 964–971 (Nov 1987). https://doi.org/10.1145/32206.32212, http://doi.acm.org/10.1145/32206.32212
9. Gleason, C., Ahmetovic, D., Savage, S., Toxtli, C., Posthuma, C., Asakawa, C., Kitani, K.M., Bigham, J.P.: 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 (Mar 2018). https://doi.org/10.1145/3191741, http://doi.acm.org/10.1145/3191741
10. Grossman, T., Fitzmaurice, G.: Toolclips: An investigation of contextual video assistance for functionality understanding. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1515–1524. CHI '10, ACM, New York, NY, USA (2010). https://doi.org/10.1145/1753326.1753552, http://doi.acm.org/10.1145/1753326.1753552
11. Grossman, T., Fitzmaurice, G., Attar, R.: A survey of software learnability: Metrics, methodologies and guidelines. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 649–658. CHI '09, ACM, New York, NY, USA (2009). https://doi.org/10.1145/1518701.1518803, http://doi.acm.org/10.1145/1518701.1518803
12. Grussenmeyer, W., Folmer, E.: Accessible touchscreen technology for people with visual impairments: A survey. ACM Trans. Access. Comput. **9**(2), 6:1–6:31 (Jan 2017). https://doi.org/10.1145/3022701, http://doi.acm.org/10.1145/3022701

13. Hagiya, T., Yazaki, T., Horiuchi, T., Kato, T.: Typing tutor: Automatic error detection and instruction in text entry for elderly people. In: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct. pp. 696–703. MobileHCI '15, ACM, New York, NY, USA (2015). https://doi.org/10.1145/2786567.2793690, http://doi.acm.org/10.1145/2786567.2793690
14. Kane, S.K., Bigham, J.P., Wobbrock, J.O.: 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. pp. 73–80. Assets '08, ACM, New York, NY, USA (2008). https://doi.org/10.1145/1414471.1414487, http://doi.acm.org/10.1145/1414471.1414487
15. Kato, T.: What question-asking protocols can say about the user interface. International Journal of Man-Machine Studies **25**(6), 659 – 673 (1986). https://doi.org/https://doi.org/10.1016/S0020-7373(86)80080-3, http://www.sciencedirect.com/science/article/pii/S0020737386800803
16. Kelleher, C., Pausch, R.: Stencils-based tutorials: Design and evaluation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 541–550. CHI '05, ACM, New York, NY, USA (2005). https://doi.org/10.1145/1054972.1055047, http://doi.acm.org/10.1145/1054972.1055047
17. Lafreniere, B., Grossman, T., Fitzmaurice, G.: Community enhanced tutorials: Improving tutorials with multiple demonstrations. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1779–1788. CHI '13, ACM, New York, NY, USA (2013). https://doi.org/10.1145/2470654.2466235, http://doi.acm.org/10.1145/2470654.2466235
18. Lieberman, H., Rosenzweig, E., Fry, C.: Steptorials: Mixed-initiative learning of high-functionality applications. In: Proceedings of the 19th International Conference on Intelligent User Interfaces. pp. 359–364. IUI '14, ACM, New York, NY, USA (2014). https://doi.org/10.1145/2557500.2557543, http://doi.acm.org/10.1145/2557500.2557543
19. Montague, K., Hanson, V.L., Cobley, A.: Designing for individuals: Usable touch-screen interaction through shared user models. In: Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility. pp. 151–158. ASSETS '12, ACM, New York, NY, USA (2012). https://doi.org/10.1145/2384916.2384943, http://doi.acm.org/10.1145/2384916.2384943
20. Oh, U., Kane, S.K., Findlater, L.: 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. pp. 13:1–13:8. ASSETS '13 (2013). https://doi.org/10.1145/2513383.2513455, http://doi.acm.org/10.1145/2513383.2513455
21. Rodrigues, A., Camacho, L., Nicolau, H., Montague, K., Guerreiro, T.: Aidme: Interactive non-visual smartphone tutorials. In: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct. pp. 205–212. MobileHCI '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3236112.3236141, http://doi.acm.org/10.1145/3236112.3236141
22. Rodrigues, A., Montague, K., Nicolau, H., Guerreiro, J., Guerreiro, T.: In-context Q&A to support blind people using smartphones. In:

Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility. pp. 32–36. ASSETS '17, ACM, New York, NY, USA (2017). https://doi.org/10.1145/3132525.3132555, http://doi.acm.org/10.1145/3132525.3132555
23. Rodrigues, A., Montague, K., Nicolau, H., Guerreiro, T.: Getting smartphones to talkback: Understanding the smartphone adoption process of blind users. In: Proc. of ACM SIGACCESS Conference on Computers and Accessibility. pp. 23–32. ASSETS '15 (2015). https://doi.org/10.1145/2700648.2809842, http://doi.acm.org/10.1145/2700648.2809842
24. Sato, D., Takagi, H., Kobayashi, M., Kawanaka, S., Asakawa, C.: Exploratory analysis of collaborative web accessibility improvement. ACM Trans. Access. Comput. **3**(2), 5:1–5:30 (Nov 2010). https://doi.org/10.1145/1857920.1857922, http://doi.acm.org/10.1145/1857920.1857922
25. Shinohara, K., Wobbrock, J.O.: In the shadow of misperception: Assistive technology use and social interactions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 705–714. CHI '11, ACM, New York, NY, USA (2011). https://doi.org/10.1145/1978942.1979044, http://doi.acm.org/10.1145/1978942.1979044
26. Takagi, H., Kawanaka, S., Kobayashi, M., Itoh, T., Asakawa, C.: Social accessibility: Achieving accessibility through collaborative metadata authoring. In: Proc. of ACM SIGACCESS Conference on Computers and Accessibility. pp. 193–200. Assets '08 (2008). https://doi.org/10.1145/1414471.1414507, http://doi.acm.org/10.1145/1414471.1414507
27. Vigo, M., Harper, S.: Coping tactics employed by visually disabled users on the web. International Journal of Human-Computer Studies **71**(11), 1013 – 1025 (2013). https://doi.org/https://doi.org/10.1016/j.ijhcs.2013.08.002, http://www.sciencedirect.com/science/article/pii/S1071581913001006
28. Vyshnavi: Tech Accessibility Youtube Channel (2018), accessed Jan 3 201. https://www.youtube.com/channel/UCOVwNKy4c_jDllBXUntbh5Q/about
29. Wang, C.Y., Chu, W.C., Chen, H.R., Hsu, C.Y., Chen, M.Y.: Evertutor: Automatically creating interactive guided tutorials on smartphones by user demonstration. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 4027–4036. CHI '14, ACM, New York, NY, USA (2014). https://doi.org/10.1145/2556288.2557407, http://doi.acm.org/10.1145/2556288.2557407
30. Zhang, D., Zhou, L., Uchidiuno, J.O., Kilic, I.Y.: Personalized assistive web for improving mobile web browsing and accessibility for visually impaired users. ACM Trans. Access. Comput. **10**(2), 6:1–6:22 (Apr 2017). https://doi.org/10.1145/3053733, http://doi.acm.org/10.1145/3053733