AidMe: Interactive Non-Visual Smartphone Tutorials

André Rodrigues
LASIGE, Faculdade de Ciências, Universidade de Lisboa
Lisboa, Portugal
afrodrigues@fc.ul.pt

Kyle Montague
Open Lab,
Newcastle University
Newcastle upon Tyne, UK
kyle.montague@newcastle.ac.uk

Leonardo Camacho
LASIGE, Faculdade de Ciências, Universidade de Lisboa
Lisboa, Portugal
lcamacho@lasige.di.fc.ul.pt

Tiago Guerreiro
LASIGE, Faculdade de Ciências, Universidade de Lisboa
Lisboa, Portugal
tjvg@di.fc.ul.pt

Hugo Nicolau
INESC-ID, Instituto Superior Técnico, Universidade de Lisboa
Lisboa, Portugal
hman@inesc-id.pt

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single-spaced in Verdana 7 point font. Please do not change the size of this text box.
Each submission will be assigned a unique DOI string to be included here.

Abstract

The constant barrage of updates and novel applications to explore creates a ceaseless cycle of new layouts and interaction methods that we must adapt to. One way to address these challenges is through in-context interactive tutorials. Most applications provide onboarding tutorials using visual metaphors to guide the user through the core features available. However, these tutorials are limited in their scope and are often inaccessible to blind people. In this paper, we present AidMe, a system-wide authoring and playthrough of non-visual interactive tutorials. Tutorials are created via user demonstration and narration. Using AidMe, in a user study with 11 blind participants we identified issues with instruction delivery and user guidance providing insights into the development of accessible interactive non-visual tutorials.

Author Keywords

Tutorials; Blind; Smartphone; Accessibility; Assistance.

ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces - Training, help, and documentation.

Introduction

Smartphones built-in accessibility services have enabled blind people to take full advantage of the device (e.g. screen reader). However, adopting the device, learning a new task or adapting to an update is still a challenge [7,8]. The variety of applications and layouts can be overwhelming when one wants to be effective using its smartphone. Nowadays, when a user starts, for the very first time, interacting with a mobile application they are often led through an onboarding process. Developers create these onboarding tutorials to guide users through the core features of their applications. Unfortunately, many rely solely on visual metaphors to guide users (e.g. overlay to obscure non-relevant content). Moreover, they are limited in their scope covering only features deemed relevant, supporting only first interactions and limited by the bounds of each application. Thus, they have limited coverability and are not accessible to blind people.

To overcome these issues, users could eventually browse the web to find answers (e.g. YouTube accessibility channels [8]) and/or resort to dedicated fora and mailing lists, such as *AppleVis* [1] or *Eyes-Free* [2], to ask questions and find answers. However, less tech savvy users will likely lack the expertise to make the adequate queries, or to understand how the answers translate to their specific context. Moreover, relying on a secondary device, and having to constantly switch between context, will increase the cognitive workload by requiring the user to split attention. Providing in-context assistance can be part of the solution as it has been shown to facilitate users' learning process in other domains [3,5].

In Evertutor [10], researchers have tackled the coverability issue by creating a tool that allows the authoring of system-wide interactive tutorials based on user demonstrations. However, it also relies solely on visual metaphors to guide the user and is not available on unmodified operating systems. Our prior work in smartphone assistance for blind people, has explored how to provide in-context Q&A allowing users to rely on a network of sighted volunteers to ask and browse questions [8]. The study revealed the need to support users beyond a single step, supporting complex tasks.

To our knowledge, there are no tools with the ability to create non-visual interactive tutorials on smartphones nor studies on how blind people cope with the instructions provided. In this paper we present AidMe, a tool for the creation and playthrough of system-wide non-visual tutorials. Using AidMe, in a user study with 11 blind participants, allowed us to understand behaviors and difficulties during playthrough, as well as assess the overall effectiveness of the tutorials created.

Our results revealed participants struggled to understand tutorial instructions either due to poor speech performance by authors, or confusion of text-to-speech instructions with interaction feedback. Coping with the variety of instructors and learners that this approach enables calls out for the need for both authoring and playthrough dynamic mechanisms that improve and adapt the quality of the dialogue with users. Only 30% of the tutorials were successfully completed by participants. Users often deviated from the path and were unable to understand how to reach their target. Participants had a variety of navigational behaviors that influenced their ability to follow specific tutorial instructions.

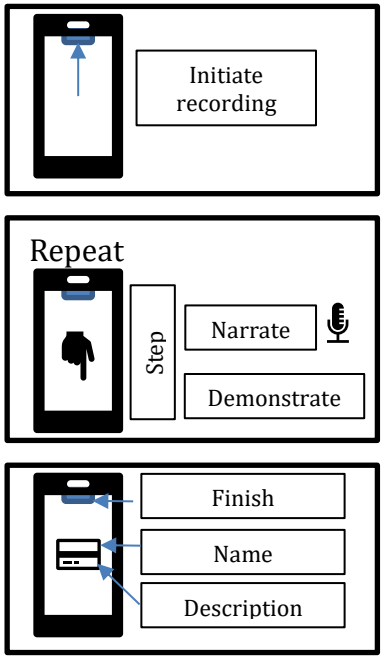

Figure 1 - Authoring workflow

Aid Me

To explore interactive non-visual tutorials we developed Aid Me, a system-wide tutorial service for Android. AidMe is an accessibility service that users can take advantage to record and play tutorials in any application and across applications. AidMe enables both sighted and blind people to create non-visual tutorials. Created tutorials can then be shared and followed by anyone.

Authoring

We developed AidMe to support tutorial authoring by demonstration and audio recording. Users first select an overlay button on the top of the screen to start recording from any screen in any application. The overlay button is easy to access, as it is close to the edges and maps to the shortcut of the first element; and unobtrusive since it does not overlay with any focusable item on screen, only partially over the status bar. Where the users start determines the screen from which the tutorial will later be available. For example, if users start recording from the home screen how to add a contact, then the tutorial will be available to start from the home screen. When users start the recording process they can narrate each step while demonstrating how to perform it (Figure 1). While recording, AidMe is gathering in addition to the audio recording, the elements with which the user is interacting, how the user is interacting and every information available in each screen (e.g. layouts, descriptions, sizes). To finish the tutorial, users select again the overlay button on top of the screen, and either discard it or give the tutorial a name and a description to save it. The end-result is a tutorial segmented into steps that are defined by the target element (i.e. extracted from the user interaction) and associate with the audio excerpt, an application and the corresponding screen. Tutorials are stored in a shared cloud database. The authoring process is accessible to both sighted and visually impaired tutorial creators.

Playthrough

When AidMe detects there is an available tutorial that starts or has a step in the current screen, an overlay button appears that when pressed shows a list of available tutorials. When following a tutorial, in each step, the user will first hear the audio instruction of the author (e.g. "*Open the app*".) followed by the announcement of the target element by the screen reader (e.g. "*Next Step WhatsApp*"). Similarly, to AidMe authoring, playthrough was designed to be unobtrusive and to not interfere with app layouts or user interactions. The only added element to every screen was the ability to select the overlay button to repeat the last instruction given. When playing through a tutorial, if the user deviates from the recorded path, an announcement is made that he or she has left the intended path. If the user navigates to any of the screens associated with any of the steps, the tutorial is resumed.

Following Tutorials

We were interested in understanding if blind people could take advantage of the non-visual interactive tutorials created through AidMe. To do so we first conducted a tutorial authoring session with 10 participants. Then we conducted a laboratory study to understand users' behaviors and needs during playthrough assessing how different content and delivery impacts the user's ability to follow. To do so, we recruited 11 blind smartphone users and had each of them follow three tutorials with AidMe.

Authoring Session

We recruited five sighted users and five blind expert smartphone users. Users were considered experts if they were able to do a wide variety of advanced tasks (e.g. configure accessibility services, install applications). Sighted participants had a short briefing on mobile screen readers, how blind people interact with smartphones and given a set of tips on how to provide instructions (e.g. refer element function not its visual appearance "search" vs. "magnifying glass"). Prior to creating a tutorial for each task, users were instructed to explore and perform it. Each participant was tasked with creating five tutorials. Two were used for training. Overall, participants found AidMe unobtrusive and easy to use. Details on the differences found between the tutorials created by sighted and blind participants will be reported elsewhere.

Pre-Processing Tutorials

Six tutorials were discarded due to path deviations and audio issues. Since audio was recorded during the demonstration, there were long periods of silence. Thus, we removed them from all recordings. To train participants in the playthrough feature, we created a sample tutorial for SimpleNote on how to delete a note. In total we had 33 unique tutorials for the three tasks.

Participants

We recruited eleven blind participants, ages ranging between 25 to 63 (M=45.36, SD=14.85) years old, three females, six Android users and five iOS, experience with smartphones between one month and three years. None of the participants participated in the authoring session.

Apparatus

We used a Xiaomi Redmi 3 running Android 7.1.2, with AidMe installed. The Android default screen reader (*Talkback*) and the default keyboard (*GBoard - Google Keyboard*) were used; participants had a set of headphones available, if they wished to use. All applications were accessible from the device home screen. For posterior analysis, an interaction logger was added to AidMe that collected navigational behaviors during the reproduction of a tutorial (e.g. focused, selected elements).

Procedure

Participants were informed the purpose of the study was to understand how we could facilitate smartphone usage by providing interactive tutorials, and that they would be trying tutorials that had been previously recorded by other people - both sighted and blind. First, participants completed a demographics questionnaire. Then, to allow users to get accustomed to the device and the tutorial playthrough, they completed the training tutorial created by the research team. We explained how they could start and stop the tutorial, repeat the instruction, and how the tutorial behaved: first, they would hear the author's instruction followed by the screen reader announcing the target they needed to find and select in that tutorial step. During the training period, they were asked to use the overlay button to repeat an instruction. Once they completed the task and felt comfortable navigating the device, we asked participants to follow three tutorials, one for each task (i.e. T1 - "Clean Data from an app", T2 - "Create a group chat in *Whatsapp*", T3 -"Add a station to favorite in RadioNet").

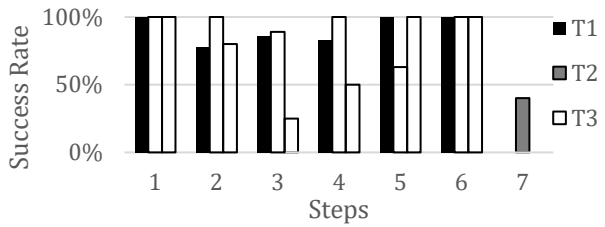

Figure 2 - Step Success Rate by Task (a participant is only included in the success rate calculation of a step, if successfully completed the previous one).

Prior to starting each tutorial, they were informed what was the task they would be learning. Each participant followed tutorials from both groups but did not repeat tasks, i.e. they would follow a random (Sighted participant - SP or blind participant - BP) tutorial for each of the tasks performed. Order of the tasks was counterbalanced and every tutorial was followed at least once. The participants could not ask questions during the task. If participants explicitly informed us they were unable to complete the step, or if they were stuck in one for more than three minutes, the trial would be considered unsuccessful and they would proceed to the next one. After each tutorial, they completed a short questionnaire about that specific tutorial. After all tasks had been completed, we had a debriefing session where we asked about their experiences with AidMe and how it could be improved.

Results

In total the participants performed 33 trials, 17 with tutorials created by SP and 16 by BP; 10 of them were completed successfully (30.3%). Participants successfully completed 6 tutorials created by SP (35.29% success rate) against 4 (25%) created by BP. Breaking it down by task, T1 had a success rate of 54.55%, T2 45.45% and T3 9.09%. Successful trials took on average 396.20s (SD=267.56). Although, success rate was low for completing a tutorial, 65.43% of all required steps were completed successfully.

Steps were not equally demanding (as shown in Figure 2) that represents the success rate of each step attempted, divided by task. All users successfully completed the first step of every task (i.e. opening the app on the home screen) and the second one of T2. The difficulties with T1 and T3 started on the second step. The most difficult step was the third step of T3 with a success rate of 25% (Figure 2) where users had difficulties interacting with the device keyboard. Below, we report on the issues identified with the playthrough process and users' behaviors when following tutorials.

PLAYTHROUGH ISSUES

One voice for multiple purposes can mislead the user. The target element was announced by the screen reader after the audio recorded for that particular step. The announcement had the same voice and audio characteristics as any other interaction with the screen reader. Some participants interacted with the device before the target had been announced which led users to mistakenly believe, when the announcement was made, it was a result of their last interaction. In three trials, users selected their current focus target when the announcement was made and were unable to recover from this incorrect step. In another case, a user inadvertently interacted with the screen and the announcement was overlapped with interaction feedback; the participant was unable to discern between the two.

Author's speech performance is relevant. We did not control the audio quality of the instructions from

the authors, nor trained them on how to narrate a tutorial. Thus, participants reported having trouble understanding some of the instructions due to authors lowering their volume during the instruction or speaking too fast.

User Behaviors

People often deviate from the intended path. A Spearman's rank-order correlation between the success rate and the number of deviations revealed a statistically significant negative correlation between trial success rate and path deviation ($r_s = -.402$, $p = .034$). In 77% of the unsuccessful trials, there were deviations to the tutorial path from which participants were unable to recover. In some cases, participants were unaware they had selected the incorrect target and were confused to why suddenly the tutorial announced they were off track. Others unintentionally selected an element and the lack of control over the device prevented them from recovering without losing track of the task at hand.

Deviating from the path did not prevent some users from completing the tutorial. In four instances users were able to recover after having deviated from the intended path. In total, 40% of the successful tutorials have a deviation in its playthrough. In all but one case participants backed into a previous state of the tutorial to resume from, most often to the start (e.g. home screen). In one case, a participant took an alternative path and ended up skipping one of the tutorial steps after two path deviations, finding himself an alternative path.

Participants focused more elements before deviating from the tutorial path. On average, users focused 11.87 elements (SD=13.77) before performing the correct selection. On steps where users deviated from the tutorial path they focused on average 19.70 elements (SD=29.39). The high standard deviation value was caused by the steps where the announcement of the target led the user to mistakenly select the first focused item.

At times people could not find the target element. The remaining unsuccessful trials are due to the user not finding the target element on a particular screen and either giving up without selecting anything or reaching the time limit defined for each step (3 minutes). However, most participants only gave up finding an element after significant navigation. On average, on the last step of an unsuccessful tutorial, users focused 41.73 elements (SD=50.01) before giving up. The high standard deviation is due to, first, the unsuccessful steps with the keyboard where the focused keys cannot be tracked, and, second, to the high variety of user's behaviors. Some participants interacted quickly with the device and did not wait for the screen reader to fully read elements, focusing as many as 136 elements. Others explore the screen meticulously and slowly through explore by touch, focusing as little as 8 elements giving up after believing to have navigated all the screen elements. Notice that this happened even with the target element being announced in each step.

Unexpected audio feedback prevented success. On the third step of T3 no author preemptively informed the user music would start to play nor how to turn off the music. Users were unable to understand the tutorial and what to do next and, as such, only one was able to complete the task.

Discussion

With AidMe users were guided based on the interactions performed on a predefined path. Despite many of the instructions seeming to be complete and useful, some participants still failed to complete them. Below, we discuss the lessons learned which should be of interest to researchers and practitioners working on non-visual mobile accessibility.

Ensure accuracy.

Users expect target descriptions in instructions to be a precise match. Furthermore, if the initial syllables do not match, users skimming through content will often discard the target. To ensure the accuracy of target descriptions one can either rely on data collected during the authoring process (e.g. user demonstration) and enrich tutorials with accurate descriptions; or alternative one must create mechanisms to ensure authors are aware of the target descriptions, this is of relevance when relying on unknowledgeable authors.

Provide guidance.

Users lacked overall awareness where they were in the application, how they had to navigate to reach their goal, and in which direction. While sighted authors lack the knowledge on how to perform navigational actions; blind authors cannot anticipate which navigation method is preferred by the end user (i.e. explore by touch or directional swiping). Moreover, participants wanted information that allowed them to be guided in real time. They wished to know how much and in which direction they needed to explore. This need calls out for dynamic tutorials that assess the user performance and support him/her, as argued in previous work for other contexts and user groups [5].

Supporting tutorial authoring by non-professionals

Although some users reported to have enjoyed the human voices on the tutorials others pointed out how the different performance of the authors impeded their understanding of the task. It is relevant to add additional control over the audio tracks created. This can be achieved by either improving the creation process through additional audio control options (e.g. mic sensitivity, pause/re-record step, and replay), which creates awareness over the quality of what is being produced, or providing post-editing tools.

Instructions delivery

One must look for alternative solutions to differentiate audio instructions from standard feedback and user interactions. A restrictive solution would be to lock user interactions while instructions are being given. Alternatively, one could use concurrent speech techniques [3] and allow users to freely explore the screen while hearing the instructions.

Safeguarding playthrough

Deviations from the tutorial path were one of the main causes of unsuccessful trials. Prior work with sighted people has argued over the benefits of allowing the user to freely explore an application while following a tutorial [6]. However, for novice users who are not able to recover from unwarranted deviations, a similar approach to Evertutor [10] where incorrect interactions are prevented might prove to be more effective. One must also consider users with a higher expertise level that benefit from a free exploration of the task or that users might be tempted to find alternative paths.

Limitations

Participants did not use their own device. Therefore, some of the faced difficulties (e.g. using the keyboard) cannot be discerned if were caused by lack of familiarity or if they were long standing issues. We assessed three tasks and although they are all from different applications they may not be representative of the multitude of smartphone tasks.

Conclusion

We presented the first iteration of AidMe, the first non-visual interactive tutorial authoring and playthrough tool on smartphones. We report on a study where 11 blind participants followed tutorials that revealed issues with instruction delivery and guidance. There is an opportunity to explore feedback mechanisms (e.g. sonification and multiple audio sources) to provide in-context assistance on smartphones. In future work we will explore how to support the accuracy, distinguishability and personalization of tutorial instructions that adapt to each user need in each step.

Acknowledgments

We thank FRMS in Lisbon (Portugal) and all participants. This work was supported by FCT through funding of the scholarship, ref. SFRH/BD/103935/2014, INESC-ID Research Unit UID/CEC/50021/2013 and LASIGE Research Unit, ref. UID/CEC/00408/2013.

References

1. AppleViz, iOS blind support community, August 22, 2017 from https://www.applevis.com/
2. Eyes-Free, Android blind and low-vision forum, August 22, 2017 from https://goo.gl/uiib1F
3. Tovi Grossman 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 (CHI '10). ACM, New York, NY, USA, 1515-1524.
4. João Guerreiro and Daniel Gonçalves. 2014. Text-to-speeches: evaluating the perception of concurrent speech by blind people. In Proc. ASSETS '14. ACM, New York, NY, USA, 169-176.
5. Toshiyuki Hagiya, Tomonori Yazaki, Toshiharu Horiuchi, and Tsuneo Kato. 2015. Typing Tutor: Automatic Error Detection and Instruction in Text Entry for Elderly People. In Proc. (MobileHCI '15). ACM, New York, NY, USA, 696-703.
6. Henry Lieberman, Elizabeth Rosenzweig, and Christopher Fry. 2014. Steptorials: mixed-initiative learning of high-functionality applications. In Proc. IUI '14. ACM, New York, NY, USA, 359-364.
7. André Rodrigues, Kyle Montague, Hugo Nicolau, and Tiago Guerreiro. 2015. Getting Smartphones to Talkback: Understanding the Smartphone Adoption Process of Blind Users. In Proc. ASSETS '15. ACM, New York, NY, USA, 23-32.
8. André André Rodrigues, 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 (ASSETS '17). ACM, New York, NY, USA, 32-36.
9. Vyshnavi, Tech Accessibility Youtube Channel, August 22, 2017 from https://goo.gl/3ijYhp
10. Cheng-Yao Wang, 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 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14). ACM, New York, NY, USA, 4027-4036.