Exploring Visually Impaired People's Gesture Preferences for Smartphones

Maria Claudia Buzzi[1], Marina Buzzi[1], Barbara Leporini[2], Amaury Trujillo[1]
[1] IIT-CNR; [2] ISTI-CNR
Via Moruzzi
Pisa, Italy
{claudia.buzzi, marina.buzzi, amaury.trujillo}@iit.cnr.it; barbara.leporini@isti.cnr.it

ABSTRACT
In this study we investigated how visually impaired people perform gestures on touch-screen smartphones. To this end, we recruited 36 visually impaired participants to explore differences and preferences in carrying out a set of gestures, selected according to certain characteristics (e.g., shape, number of fingers or strokes, etc.). For this purpose, we developed a system to collect gestures from several participants interacting with mobile smartphones at the same time. Results confirm previous research regarding the preference of visually impaired users for simple gestures, made with one finger and a single stroke. Moreover, rounded shapes were greatly preferred to angular ones.

Categories and Subject Descriptors
H.5.2. **[Information Interfaces and Presentation]:** User Interfaces – *input devices and strategies*; K.4.2. **[Computers and Society]**: Social Issues – *assistive technologies for persons with disabilities*.

Keywords
Accessibility, visually impairment, touch gestures, mobile devices, multimodal interfaces

1. INTRODUCTION
Smartphones are the most common mobile devices; in the third quarter of 2014 they accounted for 66% of the global mobile market [18]. Most smartphones use touchscreen technology, so touch-based user interfaces have become the main mobile interaction paradigm. However, touch-based interfaces are a challenge for most visually impaired people, especially for blind users [8]. Some gestures may require complicated finger movement or too many fingers to combine in order to carry out an action such as text editing. Furthermore, like their desktop counterparts, assistive technologies for mobile devices (e.g., VoiceOver and TalkBack) may require many commands, with numerous gestures to be learned. For this reason, other input methods such as automatic speech recognition and external keyboards are accessible alternatives, but they also have significant usability issues, (e.g., the small size of the keys, difficulty using text-to-speech feedback or dictation in noisy environments) [3; 7]. Visually impaired people should have the same opportunity to choose gestures or alternative input modalities as sighted users. Therefore, the main focus of this study is to understand whether and how gestures could be improved, in order to make the touch-based interaction by visually impaired users more accessible and usable. Indeed, we feel that usable touch gestures can significantly enrich the mobile user experience of visually impaired people, despite the inherent difficulties in touch-based interfaces on a flat screen.

CHItaly 2015, September 28-30, 2015, Rome, Italy
© 2015 ACM. ISBN 978-1-4503-3684-0/15/09…$15.00
DOI: http://dx.doi.org/10.1145/2808435.2808448

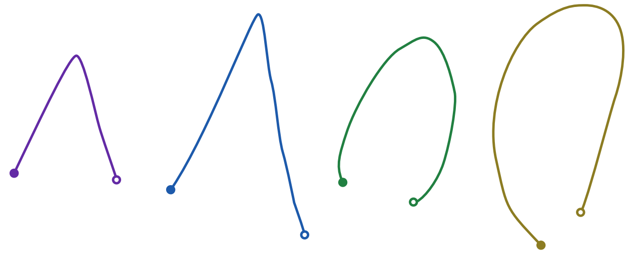

Figure 1. Four sample variations of a chevron gesture made by people with different levels of visual impairment

Starting from the main difficulties experienced by visually impaired people, in our study we would like to investigate and discuss their gesture preferences on smartphone-sized screens, based on their feedback and data of gestures' input data. For example, visually impaired people, especially the blind, can experience problems understanding the shape or outline of the gesture. Thus, we selected and categorized a set of gestures in order to investigate their main usability problems as well as to detect those preferred – and more suitable – by visually impaired people. We based several of these reference gestures on those available in VoiceOver and TalkBack since some users might be already familiar with them.

In the following sections, we present a brief discussion of the related work in the field. We then describe our methodology and capture system, which we developed to collect the gesture samples of the 36 participants during the study. We then present results, discussion and conclusions regarding the collected data.

2. RELATED WORK
A gesture is defined by a set of features called descriptors. Many studies use a combination of absolute geometric and kinematic features as descriptors [2; 10; 17], particularly those defined by Rubine in his seminal work for gesture recognition [19]. Rubine describes a set of thirteen gesture features that can be used to recognize a given gesture. He based this selection based on three criteria for the set of features: 1) each feature should be incrementally computable in constant time per input unit; 2) each feature should be meaningful so that it can be used in gesture semantics and recognition; 3) there should be enough features to provide differentiation between all gestures that might be expected within reasonable expectations and efficiency. Among the simplest features of Rubine's selection are the following:

- *Length:* the cumulative sum of the Euclidean distance, in pixels, between adjacent points of the gesture's path.
- *Area*: the area, in square pixels, of the bounding box that encloses the gesture, defined by its maximum and minimum points along the x-axis and the y-axis.
- *Aspect ratio*: the dimensionless ratio between the width and height of the gesture's bounding box.
- *Duration:* the difference, in milliseconds, of the timestamps of the gesture's first point and last points.
- *Speed:* the rate of movement of the gesture's pointers, defined by the length and duration of the gesture.

Nonetheless, Vatavu et al. have proposed the use of relative accuracy measures for stroke gestures [25]. They argue that traditional absolute measures do not describe in fine-grained detail the features of a gesture path. To compensate for this lack of subtle details, they proposed twelve new measures to describe the geometry, kinematics, and articulation of stroke gestures. In a related study, the same authors later proposed the use of color-rich gesture heat maps to better understand gesture performance [26].

Regarding the general research on touch-based gesture performance and preferences, Morris et al. suggested using gesture elicitation, a participatory design approach, to discover good gestures through participants' consensus [13]. They defined as *good* gestures those that meet usability criteria such as discoverability, ease-of-performance, memorability, and reliability. Incidentally, gesture elicitation has been also used to study spatial gestures using smartphone motion sensors [20].

Concerning the *goodness* of touch-based gestures, Yosra et al. carried out a study to better understand perceived difficulty of multi-touch gesture articulation [17]. They found a significant impact of finger count, stroke count and synchronicity on participants' perceived difficulty. They also presented a set of guidelines with fourteen recommendations. However, given that this study used a 32-inch multi-touch display, we think some of the proposed guidelines are not suitable for smartphone displays, especially for users with visual impairments. For instance, the use of more than one finger for more expressive gestures or bi-manual gesture synchronicity would be difficult on smartphone screens. In another study, Vatavu et al. estimated the user-perceived scale of stroke gestures (small, medium, or large) and found a significant consensus among users [24]. For this reason, they suggested using gesture scale to simplify gesture set design and function mappings.

On the other hand, touchscreen accessibility research is much more limited for people with visual impairments. Moreover, sometimes researchers do not have access to visually impaired participants, so they perform their studies with blindfolded participants [11; 21] or by blocking the device display [14]. However, such practices might give misleading results [23]. In addition, blind people have more difficulties learning touch-based gestures, given their intrinsic graphical nature [22].

Besides the traditional raised paper diagrams, alternative computer-based training methods to teach and train blind people to draw have been proposed, such as sonification [14] and multimodal pens [16]. Nonetheless, learning new gestures still remains a challenging task for blind people. A study by Sandness et al. explored the use one-finger, one-stroke touch-based directional gestures [21] on self-service devices, like an automated teller machine. Their proposed interface element was a menu with a set of options in the main border landmarks of the screen (top, bottom, left, etc.). The user would then select a given option by performing a simple gesture in any part of the screen in the direction of the desired option. The prototype was tested and well received by people with and without visual impairments.

Kane, Wobbrock, and Ladner did two studies on preference and performance of touch gestures made by bind people on a tablet PC, compared to sighted people [10]. The first study concerned gesture elicitation by both blind and sighted people, in which participants were asked to invent gestures for a given set of tasks. The second study was to determine whether there were significant differences between these two groups of participants. Based on these two studies they proposed a series of guidelines for usable touch-based interaction. These guidelines suggest not using print symbols, to favor screen landmarks, reduce location accuracy demand, limit time based-gestures, and use familiar layouts when possible. Based on these studies, we aim to improve the understanding of gesture performance and preference of visually impaired people in smaller touch-screen devices.

3. GESTURES IN VOICEOVER AND TALKBACK

Since Android and iOS are the most popular smartphone operating systems worldwide [18], we also looked for inspiration in the use of gestures on their respective screen readers: TalkBack and VoiceOver. Despite the significant market share gap between the most popular Android and the second place iOS, we would like to note that blind users strongly prefer iOS as their mobile system [12]. Given that VoiceOver was released before and it is derived from its desktop counterpart, the more mature accessibility features of VoiceOver compared to TalkBack, may explain this preference. In addition, people's resistance to change is a significant factor in the adoption of new or alternative user interfaces [4].

Both VoiceOver and Talkback use tap and double tap gestures to select and activate items, and swipe left or swipe right to navigate within the user interface. Both also offer a similar feature of user interface exploration, in which the user can touch or drag a finger over the screen to hear the items present in the interface. Beyond this set of common tasks, the number of gestures and features available in each platform is very different. TalkBack includes around a dozen of specific gestures to navigate on the device. Most of these require one finger; they are two part swipes at a right angle. Only a couple of gestures require the use of two fingers. On the other hand, VoiceOver has more than twenty gestures available. This set of gestures consists primarily of taps and straight swipes. The number of fingers for these two types of gesture varies from one to four. In addition, taps can be single, double or triple. Besides navigation and reading, VoiceOver offers gestures to perform specific actions that are application-dependent. For instance, a two-finger double tap could start or pause music playback and voice recording, or it could take a picture within the camera application.

The two screen readers also offer system-wide mechanisms to access more features. TalkBack has global and local context menus. Both of these context menus are radial, and they offer functions such as reading screen items, accessing controls, and setting the granularity for reading (e.g., page, word, etc.). VoiceOver uses a rotor gesture to set the function of swipe up or down according to a given application, as well as to select special input methods. For instance, the rotor can be used to set the effect on granularity.

We also initially thought that it would be interesting to see how the default iOS and Android gesture recognizers interpret participants' gestures. However, given that the main scope of this study is to describe how visually impaired people perform gestures on touchscreen smartphones and their preferences, we did not consider it suitable to expand on gesture recognition. In addition, diverse gestures require diverse recognition mechanisms. For this reason different platforms use and offer distinct libraries and approaches. Besides, the default gesture recognizers would have not worked for all of the gestures, because each screen reader prefers some gesture types to others, as previously mentioned. In addition, given that VoiceOver and TalkBack are system-wide features and take over the touchscreen's input, we would not had been able to process the gestures because the screen readers filter them. Therefore, we think these aspects would be better suited for a future apposite study.

4. Capture System Design

Due to the time constraints of the participants and local centers, we realized that it would be more convenient, if not necessary, to work with several participants at the same time. After an analysis of the available author's resources, we decided that we should be able to capture the gestures of three participants contemporaneously. Consequently, we needed to perform the capture sessions with identical mobile devices while being able to monitor and register all of the session's participants and their gestures.

Starting from the idea of using multiple mobile devices at the same time, we designed and developed a simple wireless capture gesture system, based on the client-server model. For the implementation of the capture system, we primarily used web technologies. This decision was based on the possibility of implementing such a system across different mobile platforms with relative ease, compared to native solutions [5].

4.1 System Architecture

The capture system's architecture is composed of three main nodes: 1) touchscreen mobile client devices, smartphones into which the users would input their gestures via a custom application; 2) web application, a dashboard in which the authors would adjust the settings for the capture and client devices, monitor the capture session, and visualize participants' gestures recorded previously; 3) web server, which would act as the intermediary between the mobile clients and the web application, as well as the data repository of the reference gestures and participants' captures. We would like to note that although they could have been in separate machines, we hosted both the web application and the web server on a single personal computer (a MacBook Pro laptop) for simplicity.

A wireless access point via a Wi-Fi local network established the connection between the client devices and the personal computer hosting the web server and application. Such an access point could be a wireless router, the hosting personal computer, or one of the smartphone client devices. We chose the latter because of the ease of creating an access point through a smartphone and to avoid the use of an additional device (a wireless router).

We used the WebSockets protocol in order to obtain full-duplex communication between all of the three nodes of the system. WebSockets allow opening an interactive communication session between a client (usually a web browser) and a server. Therefore, this protocol allows sending messages to a server and receiving event-driven responses (e.g., participant's gestures) without having to poll either the server or client for a reply. We chose JSON (JavaScript Object Notation) as the data interchange format between the capture system's nodes. JSON is a lightweight open standard that is machine and human readable, and is widely used for web services and applications.

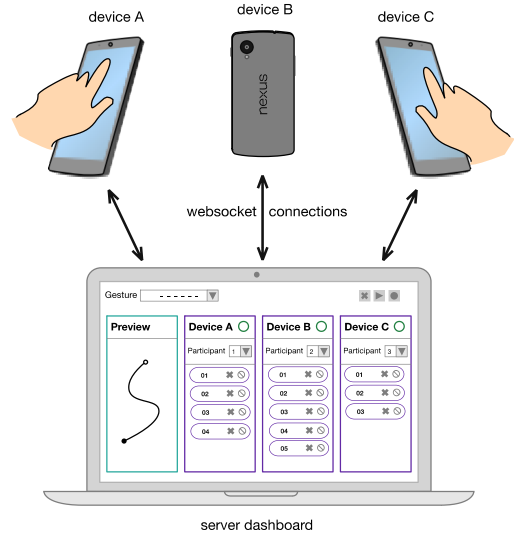

Figure 2. Client-server architecture used to wireless capture and monitor participant's gestures over a local network

4.2 Touchscreen mobile client devices

We used three identical Nexus 5 smartphones (with Android v4.4) as client devices to capture the participants' touch gestures. The Nexus 5 model has a slab format and it measures 137.9×69.2×8.6 mm. The 4.95 inches touchscreen display of the Nexus 5 has a resolution of 1080×1920 pixels (445 ppi), equivalent to the full High Definition (HD) video image format. The entire front of the device is glass with a uniform and thin bevel around the edge of the device, although the display itself does not have tactile edges. Relative to the edge of the device, the display has a 15-mm top margin, an 11-mm bottom margin and 3-mm lateral margins.

Figure 3. Dimensions of the Nexus 5 smartphone model used as touchscreen mobile client for the experiment

We developed and installed in each smartphone an Android application (app) that would connect to the web server and capture the participants' gestures. Such a mobile app consists of three main app activities, that is to say three application components that provide a screen with which users can interact in order to do something. The initial app activity would display the device ID (either A, B or C) and the current Internet Protocol (IP) address of the web server. This app activity also provides access to both the configuration and capture activities. In the configuration activity the authors are able to change the device's ID and the web server IP address. Finally, the capture app activity allows registering all of the participant's touch gestures made on the whole area of the device's display.

The graphical user interface of the capture app activity is an empty black background in full screen mode. Small uniquely colored circles indicate independent pointers (or fingers) upon contact with the display. These colored circles are only a visual reference for the authors to allow them to directly inspect the touch interaction of the participants with the screen during the user test. Because of the use of the whole display area for the perform and the capture of each gesture and the lack of navigation physical buttons on the Nexus 5, in order to close the capture app activity to return to the initial app activity, we used a moderate shake gesture of the device as the activity's close action.

4.3 Web application

The web application allowed us to register the participants, retrieve and visualize previous recorded gesture data, and manage the capture session. Using the application's capture manager dashboard, the authors are able to verify the connection to the client devices, select the reference gesture to capture (e.g., swipe left), assign a participant to each one of the connected devices, start or stop the capture of the gestures, and monitor the participants' gestures. In addition, the web application automatically marked gesture iterations made by the participants that did not match the reference gesture characteristics regarding the number or strokes or fingers used. It was also possible to manually mark gesture iterations as accidental (e.g., the participant touched the screen with the hand holding the device).

We implemented the capture system's web application using AngularJS, based on the JavaScript programming language. AngularJS is a rich web application framework that uses the model-view-controller (MVC) architectural paradigm and is mainly oriented for single-page web applications within a web browser. Based on the MVC conventions established by AngularJS, we implemented two kinds of models for the management of data, logic and rules of the application depending of the context. For data coming and going to the app of the touchscreen mobile devices through the web server, we implemented a straightforward model using WebSockets. On the other hand, to retrieve and save the participants' gestures and other information stored in the web server's database we used models based on a RESTful (Representational State Transfer) API. REST is a stateless protocol for hypermedia systems that follows an architectural style based on web services.

4.4 Web server

The web server acts as the intermediary between the web application and the mobile client devices, and it also is the gateway to retrieve and store the participant's data and reference gesture information. The web server is divided into two main components. The first component controls the data that passes between the other two system's nodes through WebSockets in the context of a capture session. The other is a RESTFul component that interacts only with the web application to save and retrieve data from the database. The web server was implemented in the Ruby programming language using the minimalist Sinatra web framework. The database was stored in an SQLite file, with the participants' gestures stored as serialized JSON.

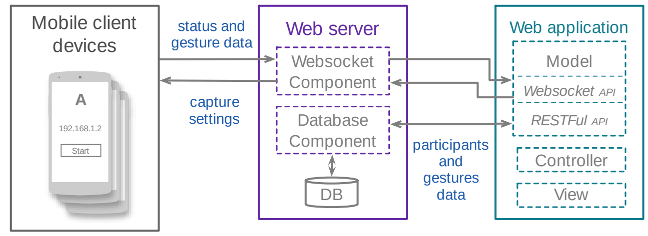

Figure 4. Conceptual model of the gesture capture system

5. STUDY DESIGN

5.1 Participants

We recruited 36 participants (14 female, 22 male) from four different local centers for blind and low-vision people in Tuscany. The mean age of the participants was 45 years for females (SD = 14.3) and 50 years for males (SD = 16.8). We classified the 36 participants into four categories: severe low vision (11), blind since birth (7), blind since adolescence (6), and those who became blind in adulthood (12). Low-vision participants had varying degrees of legal blindness; they were able to perceive light and shapes (at a relatively close distance). Therefore, they were able to perceive the general form and borders of the smartphone in their hands, although they could not discern the display edge very well (we used a black background for our capture application). The female-to-male ratio is equal or slightly higher in all categories except severe low vision, which includes only one female participant. Regarding participants' self-reported handedness, 25 were right-handed, 10 mixed-handed and only one was left-handed.

Some kind of touchscreen mobile device had previously been used by 26 of the participants – mostly smartphones, but also tablets or music players. None of the participants reported having experienced touchscreen devices before the onset of their impairment. The most popular platform was iOS, with 27 mentions out of 47. All of the participants received a USB flash drive of 8 GB for their involvement in the study.

5.2 Reference Gestures Set

We employed 25 gesture types for our study. We selected the set of gestures based on three main characteristics: pointer count, stroke count, and direction. We use p*ointer count* to refer to the number of fingers (i.e., pointers) that touch the screen uninterruptedly, from an initial touch *down* event to a final touch *up* event. For instance, a single swipe has one pointer; a pinch-in gesture has two pointers. We use *stroke count* to refer to the number of pointer sets in a gesture. A multi-stroke gesture has several pointer sets carried out in succession, with a brief period of time in between (usually around 500 ms). For example, the plus sign has two strokes of one pointer (finger) each. The term *direction* is dependent on the gesture. For swipes it can refer to a line going from one side of the screen (top, right, bottom, left, or their combination) to another. On the other hand, pinch gestures are defined by the divergence or convergence of the fingers from opposite directions, pinch-out and pinch-in respectively. Both rotor and circle gestures can have a clockwise or counter-clockwise direction. In the case of tap gestures, direction is not relevant.

Many of the gesture types we used in our study were taken from those used in VoiceOver and TalkBack. For instance, swipe down

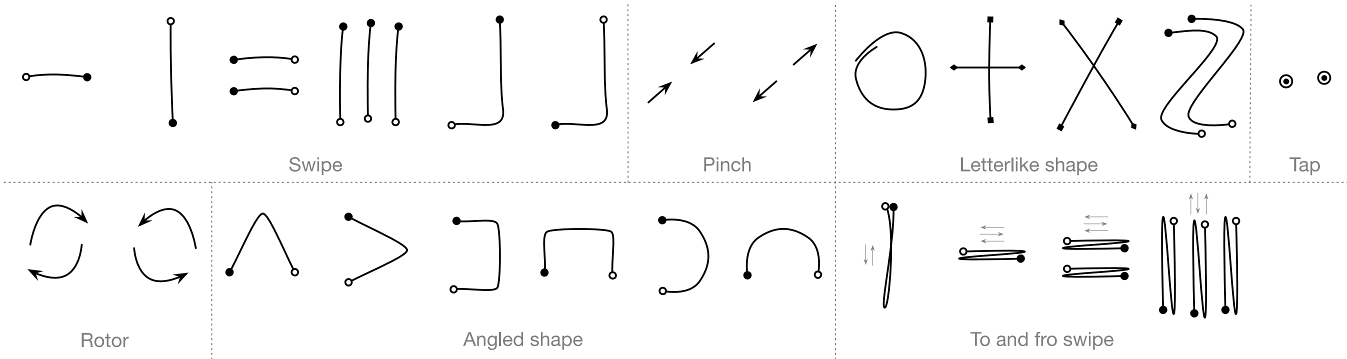

Figure 5. Set of reference gestures used in the study

then left, and swipe right then up, which are used in TalkBack but not in VoiceOver. Other gestures two-finger double tap, rotors, and the two-finger letter Z (also called scrub) are used in VoiceOver but not in Talkback The rest of the gestures were proposed by the authors to study the viability of gestures with more complex features or to study other specific aspects, such as angle preference. We classified our reference gestures in seven groups as follows:

- *Swipe:* swipe left, swipe up, two-finger swipe right, three-finger swipe down, swipe down then left, and swipe right then up
- *Pinch*: pinch-in, and pinch-out
- *Letterlike shape*: circle, plus sign, letter X, and letter Z (with two fingers)
- *Tap:* two-finger double tap
- *Rotor:* rotor clockwise, and rotor counterclockwise
- *Angled shape:* chevron open at the bottom, chevron open at the left, arch open at the bottom, arch open at the left, square open at the bottom, square open at the left
- *To and fro swipe:* swipe down then up, swipe left then right then left, two-finger swipe left then right then left, three-finger swipe up then down then up.

Regarding the use of letter-like shapes, we concur with Kane et al. in not using letter symbols because many blind people have not learned print characters or symbols [10]. However, in order to analyze multi-stroke and multi-touch gestures, we decided to test a few letter-like gestures (plus sign, letter X, and letter Z). These letters are composed of simple strokes (unlike more complex letters like P, R, or K) and may easily be described to the participants as diagonal and/or cardinal directions. We refer to them as letter-like because it is easier to describe their shape with this term. However, we did not rely on the participants' knowledge of letters, especially considering the study's participants who were blind since birth.

We also added three pairs of angled shape gestures: two chevrons, two arches and two square brackets opening in the same directions. We thought these groups of gestures would allow us to compare how different groups of visually impaired people perform related gestures. The gesture group of different angled shapes is of particular interest to us, because it would allow us to better understand the preferences and characteristics of similar gestures with steep, rounded and right angle variations. We would also like to note that when we described the gesture types to the participants, we did not associate a semantic behavior to the gestures. Instead, we described a given gesture in terms of its shape or how it is performed. For instance, we did not associate a pinch-in to a zoom-in action. Instead, we described it as the gesture one makes to pick up something small between the thumb and the index fingers (e.g., a pinch of salt).

5.3 Procedure

Our study was designed to allow up to three participants per session. In addition, we ordered the gestures by increasing difficulty, according to our perception, in order to avoid frustration for the participants. Each session lasted approximately 75 minutes and it consisted of two parts: the capture itself, and a questionnaire. In the first part, we captured the gestures of the participants. We asked the 36 participants to perform each of the 25 gestures six times, aiming to have 5400 gesture samples.

Five researchers were involved in the capture sessions, to manage the capture system, give participants instructions, and assist them. First, one of the researchers would tell the name of the gesture (in Italian), the she would explain orally how to perform each gesture, using analogies where relevant, as previously mentioned. Then, we asked the participants to make some preliminary trials to perform the gesture. Three researchers (one for each user) directly observed these non-recorded executions of the gesture to verify that the participants had understood it. In some instances, if the gesture was still not clear, we provided a cardboard cutout with the given gesture shape as a tactile representation (Fig. 6). These cardboards were rectangular and measured about 12×9 cm. As an example, the cardboard for the plus sign gesture would have this shape cut out of its center, allowing the participants to perceive the structure of the gesture by running their fingers over it. In a few instances, if the gestures were still not clear, one of the researchers would take the hand of the participant and gently guide the movement of their fingers following the given gesture shape.

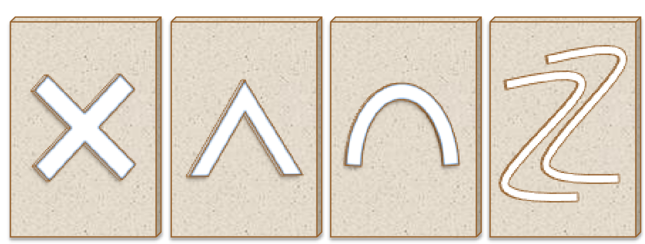

Figure 6. Gesture cardboard cutout illustrations for letter X, chevron, arch open at the bottom, and two-finger letter Z

To ease incorrect gesture detection, we also integrated an automated mechanism to mark participant's gestures in case of an incorrect number of simultaneous pointers or consecutive strokes, according to the gesture type. However, valid gesture iterations do not imply correct recognition by a gesture recognizer. For instance, a rounded "chevron" would be a valid gesture (one finger, one stroke, open at the bottom), but most likely would not be recognized as a chevron by a gesture library. We would let the participants know if they had performed the given gesture incorrectly (i.e., wrong number of stokes, fingers, direction), either marked automatically by the system, or on visual inspection in the dashboard (e.g., the gesture was performed to the right instead of to the left). In that case, we asked the participants to repeat the gestures until one of the following conditions was achieved: they completed six valid iterations; they reached ten incorrect iterations; they started to become frustrated for not correctly performing the given gesture type. We decided to move forward if the participants became frustrated by having to repeat a given gesture too many times, because we did not want to discourage nor upset them. For this reason, we do not have six iterations per participant for all of the gesture types.

Once all the participants in the session declared they understood the given gesture, we proceeded to the actual capture. After all the participants in the session had finished recording a given gesture type (or after we decided to move forward in the case of too many incorrect iterations), we asked them to rate its difficulty, using a five-point Likert rating scale, from 1 (very easy) to 5 (very difficult). We used a simple dashboard to visually manage and monitor the gestures' capture.

The second part of the procedure consisted in a web-based questionnaire in which we gathered data about the characteristics, mobile device use and gesture preferences of the participants. A researcher guided each participant in answering this online questionnaire. The questions were on participants' age, visual impairment, previous touchscreen experience, use of touchscreen mobile devices, and gesture preferences (i.e., shape, number of fingers and strokes). The researcher, using a laptop, would read aloud the questions and their respective multiple choice answers to the participant, then the researcher would input the participant's answer into the form. In addition, for each question, the participant could also add a comment to better explain the given answer. At the end of the study, the participants' responses were processed and then inserted into our database.

6. RESULTS

We would like to note that we had some difficulty capturing all of the individual samples. In certain gesture groups, some participants were unable to correctly perform all of the six iterations per gesture type in the available time. In fact, the participants perceived these groups as the most difficult: letter-like shapes, rotor, and to and fro swipes. We describe in more detail some of the participants' difficulties in capturing the gestures in the *Discussion* section. Because of these issues, not all of the 5400 individual samples collected are adequate for comparisons among groups for the moment. For these reasons, we limit the results of our preliminary analysis mainly to swipes, angled swipes, and taps.

Regarding the discarded gesture types, we have 84 incomplete captures. More than a third of these partial captures correspond to only three gestures: three-finger swipe up then down then up, two-finger swipe left then right then left, and two-finger letter Z. In addition, the number of incomplete captures is close among the participant groups (ranging from 19 to 22 by group). These gesture types were not taken into account for the following results because the samples are partial, their examination would require different and more sophisticated analyses, and it would be difficult to make an appropriate comparison with the other gesture types.

6.1 Self-reported results

Despite some issues while capturing a part of the gesture samples, we still could collect participants' prior knowledge and difficulty for all gesture types, as well as participants' preferences.

6.1.1 Perceived gesture difficulty

Overall, the participants perceived the difficulty level of most of the gesture types as low (mean = 1.49, median = 1). As illustrated in Listing 1, according to the participants the most difficult gesture groups in our set were: to and fro, and rotor. Incidentally, the top five most difficult gestures use two or more fingers, and they require several changes in direction. The easiest group of gestures was swipes, mainly with straight or right-angled strokes and taps. We would like to note that despite the similar characteristics between swipe left and swipe up, only the former was rated by all of the participants as very easy. Therefore, we confirmed previous results that the longer the gesture the more difficult it is perceived to be [17].

Listing 1. Gesture patterns sorted by participants' perceived difficulty (in increasing order)

1. Swipe left
2. Two-finger double tap
3. Swipe up
4. Two-finger swipe right
5. Swipe down then left
6. Swipe down then up
7. Circle
8. Chevron open at the bottom
9. Chevron open at the left
10. Arch open at the bottom
11. Arch open at the left
12. Square open at the left
13. Three-finger swipe down
14. Swipe right then up
15. Plus sign
16. Square open at the bottom
17. Pinch-in
18. Pinch-out
19. Letter X
20. Swipe left then right then left
21. Rotor counterclockwise
22. Two-finger swipe then right
23. Rotor clockwise
24. Two-finger letter Z
25. Three-finger swipe up then down then up

6.1.2 Gesture preferences and prior knowledge

Half of the 36 participants preferred rounded gestures, ten had no preference, six preferred steep-angle shapes, and only two participants preferred right-angled gestures. In each of the four visual impairment groups, most of the participants preferred rounded gestures. Concerning the number of simultaneous pointers per gesture, 22 participants preferred to use one finger, seven had no preference, five preferred to use no more than two fingers, and only two expressed they preferred to use at most three fingers. For the number of strokes per gesture, 19 preferred one-stroke gestures, 12 had no preference, and five preferred at most two strokes. The group of blind people since birth was the only group in which no participant expressed preference for two-stroke gestures. The majority of reference gestures we used in our study were not previously known to most of the participants. Participants reported prior knowledge of the gesture in 151 out of 900 instances (36 participants × 25 gesture types). The most popular gestures were swipe left and swipe up, which half of the participants declared to have known or used prior to the study.

6.2 Group Gesture Features

To describe the gestures made by the participants, we mainly use a subset of Rubine's features listed in the Related Work section. Based on this set of features for the swipe gesture group, we found that as we mentioned before, the mean length of the swipe depended on the direction and orientation of the gesture. Straight gestures in a vertical direction were longer (by 49%) and faster (by 21%) than those in horizontal direction. We also found shape differences in the graphical representation of the participants' mean angled swipes on steep angle gestures (Figure 7) between people with severe low vision or blind in adulthood, and people blind since early in life.

Figure 7. Participants' average gestures for the chevron gesture. From left to right: severe low vision, blind in adulthood, since adolescence, since birth, all groups

7. DISCUSSION

7.1 Overall results

Although our results are derived from an initial analysis, they suggest noteworthy differences regarding angled gestures among different groups of visually impaired people. However, so far we cannot conclude that these differences are significant. With respect to the participants' perceived difficulty of the gesture references, we think the overall low rating was due primarily to minimum feedback on their gesture performance accuracy. We let the participants know if they had used (usually unknowingly) the incorrect number of pointers, number of strokes or wrong gesture direction. However, they did not have feedback on the shape or steadiness of the gesture. We presume that if we had used a gesture template and a gesture recognition mechanism, the participants would have reported a higher level of difficulty in case their gestures were not recognized.

In general, participants rated gestures with multiple fingers, strokes or greater length as more difficult. These results are consistent with previous studies, both for sighted [17] and visually impaired people [10]. Regarding prior knowledge of gestures, although 26 participants said they had used some kind of mobile touchscreen, only half of them already knew the most popular gestures: swipe left and swipe up. Perhaps this is due to other kinds of interaction available in smartphones (e.g., speech recognition), or due to the low familiarity with touchscreen devices of some participants. Besides, we also think this result is not surprising because of the exploratory nature of the study and the fact that many blind people have not learned print characters or symbols, as remarked in the discussion of a related study [10].

7.2 Capture Issues

As mentioned in the Results section, some of the participants had issues with certain gesture groups. In most cases, the participants would repeat and successfully complete the captures after realizing they made a mistake, but in other cases this was not possible. Given the time constraints we had (each session lasted 75 minutes), and to avoid the frustration of the participants after a few repetitions, in the latter cases we decided to move to the following gesture type. One of the main difficulties was the lack of tactile edges on the devices' display. Participants would sometimes perform a gesture outside the boundaries of the display and our application would incorrectly consider the gesture as finished. This was most notorious in the case of gestures with several changes in their direction (e.g., to-and-fro swipes group). We had anticipated this issue, but decided not to use a screen overlay [9], so we could better understand how some gestures are performed around the boundaries of the screen in the real situations. In other cases, we noticed the participants' fingernails (not the fingertips) would make contact with the screen, thus the gesture was not registered as intended. This problem was more prevalent in women with long fingernails. Differences in individual attributes and abilities also affected the samples' capture. Both sensory and cognitive capabilities have a significant impact on touch-based interaction performance [15]. For example, some participants would invert the directions of the gesture while capturing it. As an instance, they would make a swipe up and down in one iteration, and a swipe down and up in the next one.

7.3 Study Limitations

Although we have gathered valuable data in this study, we realize that the average age of our participants, 48 years (SD = 15.8), implies a certain underrepresentation of the younger chunk of the visually impaired population. This issue is further exacerbated by the significant differences related to age regarding touch-based interaction [1; 6]. Additionally, younger generations have grown up in an era where touchscreen devices are commonplace; therefore, younger people are more likely to be familiar with touch-based interaction. Our case was an example of the difficulties encountered in user representation in accessibility research [23]. We hope that our future research efforts will be able to overcome these issues, working with more comprehensive groups of participants. Given the positive effect of practice in gesture consistency [2], the lack of extensive practice sessions by the participants before the capture of the sample gestures may also be considered a limitation of our study.

8. CONCLUSION

We have presented a novel approach to capturing touch-based gestures from several users at the same time on mobile devices using web technologies. We have also discussed the initial results of a study that explores touch gesture performance and preferences in smartphones among people blind from different stages of life or with severe low vision.

We found that visually impaired people generally prefer simple gestures: one finger, one stroke, in one direction preferably or with round angles otherwise. We also found noteworthy differences between various visually impaired groups regarding gesture performance, but they require further analysis before concluding they are significant.

Finally, we hope that our approach, methodology, and results, as well as the issues we encountered during the study will help and inspire future research and projects on accessible mobile touch-based interaction.

9. REFERENCES

1. Anthony, L., Brown, Q., Nias, J., Tate, B., and Mohan, S. Interaction and recognition challenges in interpreting children's touch and gesture input on mobile devices. In *Proc. Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*, ACM (2012), 225-234.
2. Anthony, L., Vatavu, R.-D., and Wobbrock, J.O. Understanding the consistency of users' pen and finger stroke gesture articulation. In *Proc. Proceedings of Graphics Interface 2013*, Canadian Information Processing Society (2013), 87-94.
3. Azenkot, S. and Lee, N.B. Exploring the Use of Speech Input by Blind People on Mobile Devices. In *Proc. 15th International ACM SIGACCESS Conference on Computers and Accessibility*, ACM (2013), 11:11-11:18.

4. Benyon, D., 1993. Accommodating individual differences through an adaptive user interface. *Human Factors in Information Technology 10*, 149-149.
5. Charland, A. and Leroux, B., 2011. Mobile application development: web vs. native. *Commun. ACM 54*, 5, 49-53.
6. Findlater, L., Froehlich, J.E., Fattal, K., Wobbrock, J.O., and Dastyar, T. Age-related differences in performance with touchscreens compared to traditional mouse input. In *Proc. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2013), 343-346.
7. Guerreiro, T., Lagoá, P., Nicolau, H., Santana, P., and Jorge, J. Mobile text-entry models for people with disabilities. In *Proc. 15th European conference on Cognitive ergonomics*, (2008), 39.
8. Kane, S.K., Bigham, J.P., and Wobbrock, J.O., 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* ACM, 73-80.
9. Kane, S.K., Morris, M.R., and Wobbrock, J.O. Touchplates: low-cost tactile overlays for visually impaired touch screen users. In *Proc. 15th International ACM SIGACCESS*, (2013), 22.
10. Kane, S.K., Wobbrock, J.O., and Ladner, R.E. Usable gestures for blind people: understanding preference and performance. In *Proc. SIGCHI Conference*, (2011), 413-422.
11. McGookin, D., Brewster, S., and Jiang, W. Investigating touchscreen accessibility for people with visual impairments. In *Proc. 5th Nordic conference on Human-computer interaction*, (2008), 298-307.
12. Morris, J. and Mueller, J., 2014. Blind and Deaf Consumer Preferences for Android and iOS Smartphones. In *Inclusive Designing*, P.M. LANGDON, J. LAZAR, A. HEYLIGHEN and H. DONG Eds. Springer International Publishing, 69-79.
13. Morris, M.R., Danielescu, A., Drucker, S., Fisher, D., Lee, B., schraefel, m.c., and Wobbrock, J.O., 2014. Reducing legacy bias in gesture elicitation studies. *interactions 21*, 3, 40-45.
14. Oh, U., Kane, S.K., and Findlater, L. Follow That Sound: Using Sonification and Corrective Verbal Feedback to Teach Touchscreen Gestures. In *Proc. 15th International ACM SIGACCESS*, ACM (2013), 13:11-13:18.
15. Oliveira, J., Guerreiro, T., Nicolau, H., Jorge, J., and Gonçalves, D. Blind People and Mobile Touch-based Text-entry: Acknowledging the Need for Different Flavors. In *Proc. 13th International ACM SIGACCESS Conference on Computers and Accessibility*, ACM (2011), 179-186.
16. Plimmer, B., Crossan, A., Brewster, S.A., and Blagojevic, R. Multimodal collaborative handwriting training for visually-impaired people. In *Proc. SIGCHI Conference*, (2008), 393-402.
17. Rekik, Y., Vatavu, R.-D., and Grisoni, L., 2014. Understanding users' perceived difficulty of multi-touch gesture articulation. In *Proceedings of the 16th International Conference on Multimodal Interaction* ACM, 232-239.
18. Rivera, J. and van der Meulen, R., 2014. Gartner Says Sales of Smartphones Grew 20 Percent in Third Quarter of 2014 Gartner, Egham, UK.
19. Rubine, D. Specifying gestures by example. In *Proc. Proceedings of the 18th annual conference on Computer graphics and interactive techniques*, ACM (1991), 329-337.
20. Ruiz, J., Li, Y., and Lank, E. User-defined motion gestures for mobile interaction. In *Proc. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2011).
21. Sandnes, F., Tan, T., Johansen, A., Sulic, E., Vesterhus, E., and Iversen, E., 2012. Making touch-based kiosks accessible to blind users through simple gestures. *Universal Access in the Information Society 11*, 4 (2012/11/01), 421-431.
22. Schmidt, M. and Weber, G., 2009. Multitouch Haptic Interaction. In *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments*, C. STEPHANIDIS Ed. Springer Berlin Heidelberg, 574-582.
23. Sears, A. and Hanson, V. Representing users in accessibility research. In *Proc. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2011), 2235-2238.
24. Vatavu, R.-D., #233, Casiez, r., and Grisoni, L. Small, medium, or large?: estimating the user-perceived scale of stroke gestures. In *Proc. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2013), 277-280.
25. Vatavu, R.-D., Anthony, L., and Wobbrock, J.O. Relative accuracy measures for stroke gestures. In *Proc. Proceedings of the 15th ACM on International conference on multimodal interaction*, ACM (2013), 279-286.
26. Vatavu, R.-D., Anthony, L., and Wobbrock, J.O. Gesture Heatmaps: Understanding Gesture Performance with Colorful Visualizations. In *Proc. Proceedings of the 16th International Conference on Multimodal Interaction*, ACM (2014), 172-179.