Multimed Tools Appl (2017) 76:5141–5169
DOI 10.1007/s11042-016-3594-9

Analyzing visually impaired people's touch gestures on smartphones

Maria Claudia Buzzi[1] **· Marina Buzzi**[1] **·**
Barbara Leporini[2] **· Amaury Trujillo**[1]

Received: 14 December 2015 / Revised: 4 May 2016 / Accepted: 6 May 2016 /
Published online: 17 May 2016

Abstract We present an analysis of how visually impaired people perform gestures on touch-screen smartphones and report their preferences, explaining the procedure and technical implementation that we followed to collect gesture samples. To that end, we recruited 36 visually impaired participants and divided them into two main groups of low-vision and blind people respectively. We then examined their touch-based gesture preferences in terms of number of strokes, multi-touch, and shape angle, as well as their execution in geometric, kinematic and relative terms. For this purpose, we developed a wireless system to simultaneously record sample gestures from several participants, with the possibility of monitoring the capture process. Our results are consistent with previous research regarding the preference of visually impaired users for simple gestures: with one finger, a single stroke, and in one or two cardinal directions. Of the two groups of participants, blind people are less consistent with multi-stroke gestures. In addition, they are more likely than low-vision people to go outside the bounds of the display in the absence of its physical delimitation of, especially with multi-touch gestures. In the case of more complex gestures, rounded shapes are greatly preferred to angular ones, especially by blind people, who have difficulty performing straight gestures with steep or right angles. Based on these results and on previous related research, we offer suggestions to improve gesture accessibility of handheld touchscreen devices.

Keywords Accessibility · Visually impairment · Touch gestures · Mobile devices · Multimodal interfaces

✉ Maria Claudia Buzzi
claudia.buzzi@iit.cnr.it

[1] IIT – CNR, v. Moruzzi 1, 56124 Pisa, PI, Italy

[2] ISTI – CNR, v. Moruzzi 1, 56124 Pisa, PI, Italy

1 Introduction

Multimedia tools and applications have greatly benefited from multimodal interfaces such as speech recognition and touch-based surfaces, which allow more transparent, flexible, efficient and expressive interaction with machines [36]. In particular, interaction on touch-sensitive screens is considered to be a direct form of user interface, where information control and display are unified in an intuitive manner [1]. Furthermore, nowadays most smartphones use multi-touch screen technology; thus, given the popularity of mobile devices, touch-based user interfaces have become the most pervasive interaction paradigm between people and computers. Nonetheless, touch-based interfaces may be a challenge for most visually impaired people, especially for users who have been totally blind since infancy [18]. For example, some gestures may require complicated finger movements or the coordination of several fingers in order to carry out an action (i.e., text editing). These gestures can be difficult to perform on a handheld device screen, especially when there is no feedback other than visual cues. Still, with the replacement of physical keys by virtual keys, touch-based interaction has become the main means of interaction with most smartphones. For instance, assistive technologies such as integrated screen readers for iOS and Android may require many touch-based commands, with numerous gestures to be learned. Unlike interaction with desktop platforms, visually impaired users do not have easy access to a mouse or a QUERTY-like physical keyboard to control their smartphone screen readers.

Besides touch-based interaction, visually impaired people and blind users in particular mainly use automatic speech recognition as an input method (e.g., Siri for iOS). In addition, visually impaired people may use the few physical buttons on the device, or an external physical keyboard via Bluetooth (QUERTY-like or Braille). Although these other multimodal input methods may be accessible alternatives, they also have significant usability issues. These include the small keys on portable keyboards, difficulty using text-to-speech feedback, or dictation in noisy environments; in addition, Braille keyboards are perceived as expensive or cumbersome to carry [6, 15]. We firmly believe that visually impaired people should have the same opportunity to choose gestural or alternative input modalities as sighted users do. Therefore, the main focus of this study is to understand whether and how gestures could be improved, in order to make touch-based interaction by visually impaired users more accessible and usable. Indeed, we think that usable touch gestures can significantly enrich the daily mobile user experience of visually impaired people, despite its inherent difficulties.

In this extended and revised work of a previous exploratory study [11], we present a detailed analysis of 36 visually impaired people's performance of touch-based gestures. We expand our previous work by describing in greater detail the underlying principles of touch-gestures and their recognition, the main difficulties experienced by the visually impaired in this regard, and their preferences on smartphone-sized screens, by examining the geometric and kinematic characteristics of the sampled gestures, both absolute and relative. For example, visually impaired people, especially blind people, can experience problems understanding the shape or outline of the gesture, which are visual by nature. 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 often 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.

First, we present a brief discussion of the related work in mobile touch-based interaction and gesture interaction, as well as an overview of accessible touch gestures. We also describe the use of gestures in the most popular mobile screen readers. We then describe our methodology and the capture system we developed to collect the gesture samples of the 36 participants during the study. We then present detailed results, discussion and conclusions concerning the collected participants' data and our experience and suggestions regarding the study.

2 Related work

The usability and accessibility of touch-based gestures rely on their characteristics and the recognition techniques utilized. These techniques take into account a range of gesture characteristics that continually increases thanks to the availability of new screen technology and algorithms. For instance, the precursors of smartphones, the personal digital assistants (PDAs), were handheld devices that already had single-touch interaction via a stylus and handwritten recognition, although with mixed results [7]. The handwritten recognition used by most PDAs used simplified alphabets based on stroke gestures, like the alphabets *Unistrokes* [14] and *Graffiti,* intended to be easy to learn and perform [27]. Over time, smartphones overshadowed the PDA market, particularly after the release of the first iPhone, which popularized multi-touch interaction on smartphones that was based on research by Westerman, Elias, and Hedge [58]. Incidentally, the demand for touch-based smartphones increased significantly thanks to the greater industrialization of capacitive sensing technologies [57]. This demand also prompted more research on the topic, such as touch-based interaction in mobile environments [9] and usability of multi-touch mobile applications [23]. Nonetheless, despite the significant changes in mobile touchscreen technology, the principles of gesture recognition are decades old and are shared across different types of surfaces and devices.

2.1 Gesture features and recognition

Touchscreen gestures are basically a series of timed points made by a pointer in a two-dimensional plane, which is defined by a set of features called descriptors. In addition, certain touchscreen technologies, such as capacitive screens used in smartphones, also allow measuring the pressure at each point of the gesture [57]. For touchscreen-based interactions, the pointers are usually the user's fingers, although a stylus could also be used. Most gesture studies and technologies use a combination of absolute geometric and kinematic features as descriptors [3, 20, 41], such as 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.

Among the most common features used to describe gestures on a surface are those specified by Rubine in his seminal work on gesture recognition [45]. Rubine describes a set of thirteen gesture features that can be used to recognize a given gesture. He made 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. The popularity of Rubine's features may be explained by their relative simplicity, as they were originally defined focusing mainly on gestures made with one stroke and one finger. However, multi-touch and multi-stroke gestures require a more complex specification of gestures' features. For this reason, different authors have proposed the use of others descriptors, such as gesture curviness or structure, that allow recognition of multi-touch and multi-stroke gestures [25, 59], or even their formalization [17]. Nevertheless, none of these composite gesture implementations or formalizations has yet become a commonly used standard.

Gesture and handwriting recognition are particular cases of pattern recognition, as both focus on the transformation of a language represented in a spatial form through graphic marks into its symbolic representation [38]. The choice of gesture recognition algorithms depends on the context of the application, and it is based on the characteristics of the target marks, such as timing, number of pointers and number of strokes. The techniques for recognition of gestures and handwriting vary from the traditional feature-based Rubine's statistical classifier [45] and the Dynamic-Type Warping programing [32], to Hidden Markov Models [50] and neural networks [37]. Because these techniques require special libraries or training, simpler and lighter algorithms have been proposed for prototyping and mobile platforms, such the $-family of gesture recognizers developed at the University of Washington. For instance, the $1 recognizer was designed for rapid prototyping for gesture-based interaction and it uses geometric template matching [61], based on a previous computer pen-based shorthand writing system [22]. Another template recognizer is *Protractor*, but it has different handling of orientation and scaling, in addition to a novel similarity measure based on angular distance on training samples [24]. Still, both $1 and *Protractor* are only for the recognition of single-stroke gestures. For multi-stoke gestures the $N recognizer was initially developed as an extension for $1 [4], then a more streamlined version was developed, $N-Protractor, based on *Protractor* [5]. Still, since $N-Protractor has significant resource costs due to its combinatoric approach, the $P recognizer was developed [53]. This multi-stroke recognizer discards the gesture articulations details by considering the gestures as clouds of points, so it is faster and lighter.

To ease the development of multi-touch applications, smartphone platforms such as iOS and Android provide recognizer libraries at the operating system level for multi-touch gestures for commonly used gestures (e.g., swipes and taps). For instance the *Protractor* recognizer was incorporated into the core Android code. Still, these libraries offer limited flexibility for higher abstraction models or different reference gesture models that need to be recognized with ad hoc implementation with the raw data from the devices' touch sensors [51]. For this reason, Vatavu et al. have proposed the use of relative accuracy measures for more complex composite gestures in a template approach [54]. 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, extending the features used in the $1 and $P recognizers. Moreover, the same authors propose the analysis of these relative features through color-rich gesture heat maps, to better study and understand gesture performance [55].

2.2 Related studies on gesture usability and accessibility

The majority of research on the characteristics that make touch-based gestures usable has been conducted with sighted users. Although not all of the suggestions of these studies are suited for visually impaired people, many of the results are still applicable to conduct studies on accessible gestures, as there are many commonalities in actions and preferences between the two kinds of users. For example, to better perform research on touch-based interaction, Morris et al. suggested using gesture elicitation, a participatory design approach, to discover good gestures through participants' consensus [30]. They defined as *good* gestures those that meet usability criteria such as discoverability, ease-of-performance, memorability, and reliability. For instance, gesture elicitation can be used to better understand the challenges of gesture customization [33]. In addition, the elicitation can be applied to study other gestural input mechanisms, such as spatial gestures using the smartphone motion sensors [46]. Situational impairments may also affect the usability of touch screen devices, particularly in a mobile context. Bradgon et al. discovered that uni-stroke gestures that used the bevel of the display as a departing point were more reliable in these situations, compared to gestures located elsewhere on the screen [9].

Concerning the *goodness* of touch-based gestures, Yosra et al. carried out a study to better understand the perceived difficulty of multi-touch gesture articulation [41], and they found a significant impact of finger count, stroke count and synchronicity on participants' perceived difficulty. An increase in one or a combination of these features increases the users' perceived difficulty of the given gesture. Based on their findings, 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. Nonetheless, the use of consecutive distant taps is a viable novel type of touch-gestures that can be used conflict-free with existing touch gestures on handheld devices [16]. 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 [52]. For this reason, they suggest using gesture scale to simplify gesture set design and function mappings.

On the other hand, touchscreen accessibility research for visually impaired people is relatively narrow and presents different challenges, although in the last few years there has been an increased interest on the topic. For instance, researchers are not always able to recruit visually impaired participants, particularly blind people, so they conduct their studies with blindfolded participants [28, 47] or by covering the device display [34]. However, such practices might result in misleading outcomes [49]. Additionally, the gesture learning strategies of blind people are another aspect of the challenge in studying accessibility of gesture-based interfaces, given that to use a gesture people first need to learn how to perform it. Blind people have more difficulty learning touch-based gestures, a task that is intrinsically graphical [48]. 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 [34] and multimodal pens [39]. 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 [47] on self-service devices such 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 blind people on a tablet PC, compared to sighted people [20]. 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. Taking advantage of multimodal features of smartphones, it has also been proposed to use gestures, audio and vibration at the same time for simple tasks, like telephone dialing [56]. Although successful, this approach might be less convenient given the current simple haptic technology of smartphones for more complicated tasks such as text entry and editing [10]. Based on these studies, we aim to improve the understanding of gesture performance and preference of visually impaired people in smaller touch-screen devices.

2.3 Gestures in mobile screen readers

Since Android and iOS are the most popular smartphone operating systems worldwide [42], 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 [31]. Moreover, in gesture elicitation blind participants perform their gestures heavily influenced by the tap and flick gestures used in VoiceOver [44]. VoiceOver was released before TalkBack and it has more accessibility features since it is derived from its desktop counterpart. In addition, people's resistance to change is a significant factor in the adoption of new or alternative user interfaces [8]. 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 on 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 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, primarily taps and straight swipes with one to four fingers. 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. Although VoiceOver offers many more gestures and features than TalkBack, blind users might be overwhelmed by the amount of gestures and functionalities to learn and eventually will use a small subset of commands [26]. Besides, despite the great accessibility benefit of both platforms, they still have accessibility issues in terms of speech [6, 63] and gesture [26] recognition. Moreover, outside of the laboratory the adoption process of these accessibility services by blind people is an arduous and lengthy task [43].

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 have 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.

3 Methodology

3.1 Participants

We recruited a total of 36 participants (14 female and 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 sub-groups based on impairment or age of onset: 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 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 (our capture application had a black background). In addition we categorized the participants in two main groups. The first group includes all the three sub-groups of blind participants (25) for specific comparisons with the second group of low-vision participants. The female-to-male ratio was equal or slightly higher in all sub-groups except low-vision, which had only one female participant. Concerning participants' handedness, 25 were right-handed, 10 mixed-handed and only one was left-handed.

Most of the participants (26) had used some kind of touchscreen mobile device at least once, by the time they participated in the study. These were mostly smartphones, but also tablets or music players as seen in Table 1. None of the participants reported having used touchscreen devices before the onset of their impairment. The most popular platform was iOS, with 27 mentions out of 47, particularly iPhones. All of the participants received a USB flash drive of 8 GB for their involvement in the study.

3.2 Reference gesture set

We used 25 gesture references classified in seven groups based on three main characteristics: pointer count, stroke count, and direction. We use *finger count* to refer to the number of fingers (i.e., pointers) that continuously touch the screen, from an initial touch *down* event to a final

Table 1 Distribution of participants who had used some kind of touchscreen mobile device. Percentages are rounded

Group	*iOS*		*Android*		*Other*
	iPhone iPod	iPad	Phone	Tablet	Phone MP3
Low-vision	45 %	18 %	55 %	9 %	18 %
Blind since birth	57 %	14 %	29 %	14 %	0 %
Blind since adolescence	83 %	0 %	17 %	17 %	0 %
Blind in adulthood	58 %	25 %	8 %	0 %	8 %
All of the participants	*57 %*	*17 %*	*28 %*	*8 %*	*8 %*

touch *up* event. For example, a single swipe consists of one pointer; a pinch-in gesture consists of two pointers. We use *stroke count* to refer to the number of each continuous screen touches by set of pointers (one finger or more) in a gesture. A multi-stroke gesture has several pointer sets carried out in succession with brief interruptions (e.g., 500 milliseconds), in which no pointer makes contact with the screen. For instance, a plus shape is performed with 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 to another (to left, to the right, upward and downward). 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. However direction is not always relevant, as in the case of tap gestures. Regarding finger count and direction, we use abbreviations for reference gesture names in tables and figures to achieve a better layout. These abbreviations are: two-finger (2F) and three-finger (3F) for finger count; down (D), left (L), up (U), right (R), clockwise (CW) and counter-clockwise (CCW) for gesture direction.

We selected some of these gestures from those with a distinct use in VoiceOver and TalkBack. For example, right angle swipes (e.g., swipe down then left or swipe right then up) are used in TalkBack but not in VoiceOver. Gestures used in VoiceOver but not in Talkback are two-finger double tap, rotors, and the two-finger letter Z (also called scrub). Other gestures were proposed by the authors to study gestures with more complex features or to study other specific aspects like similar shapes with angle variations. In particular, we have three pairs of similar shape gestures that are angled differently: two chevrons, two semicircles and two square brackets. Each pair has the same shape orientation (open at the bottom and open at the left, respectively) with three angle variations: steep, rounded and right angle. We included these groups of gestures to compare preferences and performances of the different groups of visually impaired regarding shape angle.

In addition, in order to analyze multi-stroke and multi-touch gestures, we decided to test a few gestures that we classified as letter-like. These gestures consists of simple strokes and may be described to the participants with simple diagonal and/or cardinal directions, but we use the letter-like term because we think it is more apt for describing their shape. Nonetheless, we agree with the guideline of Kane et al. to avoid letter symbols, as some blind people may have not learned print characters or symbols [20], and we did not expect participants blind since birth to know the shape of these letters. We classified these reference gestures, illustrated in Fig. 1, into the following seven groups:

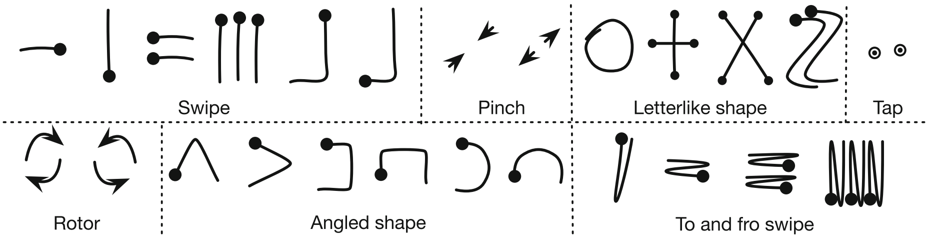


Fig. 1 Gesture reference groups

- *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
- *Letter-like shape*: circle, plus shape, x shape, and z shape (with two fingers)
- *Tap:* two-finger double tap
- *Rotor:* rotor clockwise, and rotor counterclockwise
- *Angled shape:* chevron up (open at the bottom) and chevron right (open at the left) for steep angles; semicircle up and semicircle right for rounded angles; bracket up and bracket right for right angles.
- *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.

When we described the gesture types to the participants, we did not associate a semantic behavior with the gestures (e.g., scroll user interface elements with swipe 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). We also used cardboard cutouts as tactile representations of reference gestures. These cards were rectangular and measured about 12 × 9 cm. As an example, the card for the plus shape 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.

3.3 Capture system

Taking into consideration the participant's schedule constraints across different visually impaired local centers and the available resources, we decided that it was more convenient to be able to work with several participants contemporarily. Therefore, we designed a wireless capture system, based on client–server architecture, which would allow us to work with up to three participants with identical smartphones in a single capture session. We then implemented this system mainly using web technologies because of the relative ease of cross-platform compatibility compared to native mobile solutions [12]. The capture system's architecture is composed of three main nodes: 1) client devices, touchscreen 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 The wireless connection between the client devices and the personal computer hosting both the web server and application is established through a Wi-Fi local access point. We used the WebSockets protocol to obtain full-duplex communication between all three nodes of the system. We chose the lightweight JSON (JavaScript Object Notation) as the data interchange format between the capture system's nodes.

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 $137.9 \times 69.2 \times 8.6$ mm. The 4.95 inch touchscreen display of the Nexus 5 has a resolution of 1080×1920 pixels (445 ppi) and the whole front consists of a glass panel with a 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. We developed an Android application that would allow each smartphone to connect to the web server and capture the participants' gestures. The graphical user interface of the capture display is an empty black background in full screen mode. The system's web application was implemented with AngularJS and the model-view-controller (MVC) architectural paradigm. Through the web application's capture manager dashboard, the authors were able to verify the connection to the client devices, select the reference gesture to capture, assign a participant to each connected device, start or stop the capture of the gestures, and monitor the participants' gestures. Finally, the web server acted as the intermediary between the web application and the client devices, and also as the gateway to retrieve and store the participant's data and reference gesture information (in a SQLlite database). The web server was implemented with the Sinatra web framework for Ruby and it is divided into two main components. The first one controls the data transmitted between the other two system's nodes through WebSockets during a capture session. The second is a RESTFul component that interacts only with the web application to save and retrieve data from the database (Fig. 2).

3.4 Procedure

We designed the capture session to work with one, two or three participants at the same time (Fig. 3). Each session lasted approximately 75 min in total 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, for a total 5400 gesture samples or iterations. We would like to note that we ordered the gestures by increasing difficulty according to our own perception in order to avoid frustration for the participants; this order is illustrated in Fig. 1 from left to right and top to bottom.

Five researchers (including the authors) were involved in the capture sessions, to manage the capture system, give participants instructions, and assist them. First, one of the researchers

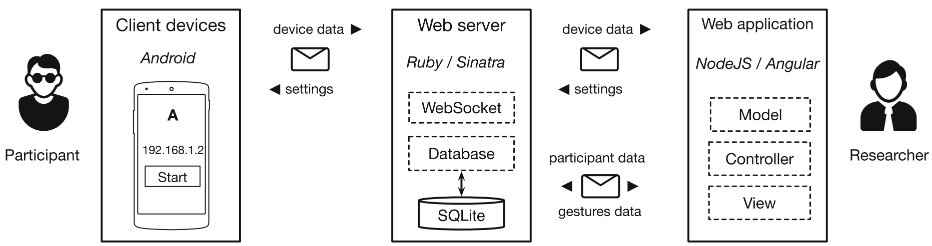


Fig. 2 Conceptual model of the gesture capture system

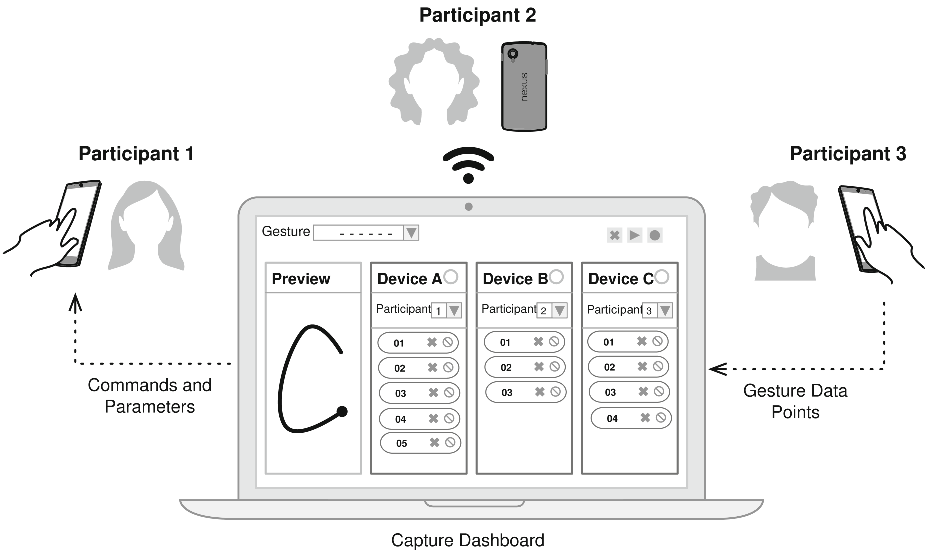


Fig. 3 Example of a capture session with three participants

would say the name of the gesture (in Italian), then she would explain orally how to perform each gesture, using analogies where relevant, as previously mentioned. Next, we would ask the participants to perform the given gesture a couple of times as preliminary trials. Three researchers (one for each user) would directly observe 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 would provide a cardboard cutout (about 12 × 9 cm) with the given gesture shape as a tactile representation. As an example, the cardboard for the plus shape 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 (cf. Fig. 6 of our exploratory study [11]). In the few instances that the gestures were still not clear, one of the researchers would take the participant's hand and gently guide the movement of their fingers following the given gesture shape.

Once all the participants in the session declared they understood the given gesture, we proceeded to the actual capture. 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, this mechanism was not able to filter out all kinds of invalid gestures, such as accidental touch of the screen that corresponded with the aforementioned of the given reference gestures. In addition, 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.

Therefore, we would let the participants know if they had performed the given gesture incorrectly (i.e., wrong number of strokes, 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 was achieved: they completed six iterations; they reached ten incorrect iterations; they began to become frustrated by 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 or upset them. For this reason, at the end of the study we did not have six valid iterations per participant for all of the gesture types. This is further discussed in the section 4. After all of 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.

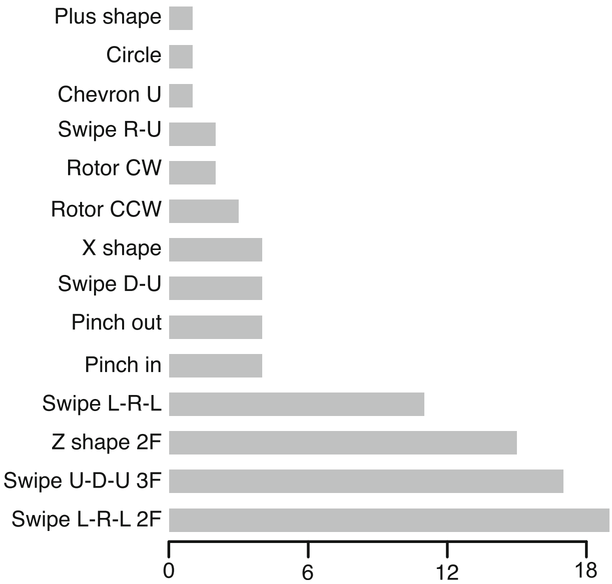
The second part of the procedure consisted of a web-based questionnaire in which we gathered data about the characteristics, mobile device use and gesture preferences of the participants. A researcher guided each participant in answering this online questionnaire. The questions were on participants' 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.

4 Results

We were expecting to have 900 captures (36 participants performed 25 gesture references) with a total of 5400 iterations (6 iterations by capture), but we realized that not all iterations could be used for analysis because some participants could not perform certain gestures within the allotted session time. We detail the participant's capture issues later in the section 5. Therefore, we decided to review all of the captures and iterations to mark down those that could be used for further analysis. We tagged as invalid those captures that had three or more recorded iterations in which the participant could not perform the reference gesture. For captures that had one or two iterations that did not correspond with the reference gesture and the rest of the iterations, we tagged the capture as valid but tagged the corresponding iterations as invalid. For instance, many captures were invalidated because one of two fingers would go outside of the borders of the display. After this review, we had 812 valid captures, of which 77 had one invalid iteration and 13 two invalid iterations. Although only 10 of the 25 gestures are multi-touch (i.e., they require two or more contemporary fingers), 64 of the 88 invalid captures belonged to multi-touch gestures.

Blind participants had 10.5 % of their captures invalidated, and low-vision participants 8 %. Female participants had a total of 43 of 350 invalid captures (12.28 %), male participants 45 of 550 (8.54 %). Of the 36 participants, only three participants had all of their captures marked as valid: all three male, two with low vision (aged 42 and 33 years) and one who became blind in adulthood (age 66). The participant with the most invalid captures (eight) was a 29-year-old female blind since birth, and the only one who stated being left-handed (and not mixed-handed). The second participant with the most invalid captures (six) was also female but blind since adolescence, age 75. Fourteen reference gestures had at least one invalid capture, but only the top four represented more than 70 % of invalid captures, as illustrated in Fig. 4. All of these four gesture references represent swipe-like gestures that have three directions, with one or multiple fingers.

Fig. 4 Number of invalid captures per gesture reference



Once we had marked down participants' valid captures and iterations, we performed a series of statistical analyses to study if there were potential differences between groups and sub-groups of participants. In the following paragraphs we present the results of these analyses. We first present the self-reported results of the participants on the perceived difficulty, prior knowledge and preferences of the participants. We then present an analysis of the valid gestures, using the R Statistical Project software, in terms of absolute geometric and kinetic features. Finally, we present the results of the analysis on relative features, in which the template gesture of reference is the mean gesture of all the valid gestures. For this we used three tools for gesture analysis released by the multimedia and accessible design laboratory (MAD lab) of the University of Washington.

4.1 Self-reported results

4.1.1 Gesture difficulty

Overall, participants perceived most of the gesture types as very easy in the five-point Likert scale (mean level = 1.49, SD = 0.91). Blind participants felt slightly more confident about how they performed gestures (mean level = 1.45, SD = 0.86) and those with low vision (mean level = 1.56, SD = 1.02), although there was no significant difference in perceived difficulty based on the Wilcoxon rank sum test with continuity correction ($W = 88474$ $p = 0.3785$). In addition, the level of agreement on perceived gesture difficulty among all of the participants was low. The overall Kendall's coefficient of concordance [21], which goes from zero (no agreement) to one (complete agreement), was $W = .265$, $\chi^2(24) = 229$, $p < .001$. On the other hand, the concordance was higher within visual impairment groups. People who became blind in adulthood had the lowest group agreement level: $W = .309$, χ^2 $(24) = 89$, $p < .001$, followed closely by people blind since adolescence $W = .31$, χ^2 $(24) = 44.6$, $p < .007$. For people blind since birth agreement was $W = .4$, χ^2 $(24) = 67.2$, $p < .001$. Finally, people with low vision had the highest level of agreement: $W = .415$, χ^2 $(24) = 110$, $p < .001$. As indicated in Listing 1, according to the participants the most difficult gesture groups in our set were: to-and-fro swipe, and rotor. Incidentally, the top five most difficult gestures use two or more fingers, and they require several changes in direction, with the most difficult gesture, swipe up then down then up with three fingers, having mean level of 2.66 (SD = 1.49). This corresponds to the reference gestures that had the most invalid captures. The easiest gestures were swipes, mainly with straight or right-angled strokes, as well as taps, with swipe left being perceived the easiest gesture (mean level = 1, SD = 0).

Regarding the swipe group of gestures, we would like to note that despite the similar characteristics between swipes left and up, only the former was rated by all of the participants as very easy; swipe up had a difficulty mean level of 1.14 (SD = 0.42). The difference is related to the higher space availability on the vertical axis of the screen to perform these gestures, as the smartphone devices that we used have an aspect ratio of 9:16 in portrait mode. Consequently, vertical swipes were longer than horizontal swipes. For instance, the mean path length for swipe left was 843 pixels (SD = 236), and for swipe up it was 1275 pixels (SD = 561). Likewise, the perceived difficulty of swipe right with two fingers (mean level = 1.19, SD = 0.47) was lower than that of swipe down with three fingers (mean level = 1.5, SD = 0.81). In addition, we noticed that both multi-touch swipes were perceived as slightly more difficult than one-finger swipes. In the same way, the multi-stroke gestures plus shape (mean level = 1.25, SD = 0.65) and x shape (mean level = 1.44, SD = 0.88) differ significantly in length, with the latter being 23 % longer on average than the former. These results are consistent with related studies on increased perceived gesture difficulty in terms of greater gesture length and number of fingers, both with sighted [41] and visually impaired participants [62].

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

1. Swipe L	8. Chevron U	15. Plus shape	22. Swipe L-R-L (2F)
2. Double tap (2F)	9. Chevron R	16. Bracket U	23. Rotor CW
3. Swipe U	10. Semicircle U	17. Pinch in	24. Z shape (2F)
4. Swipe R (2F)	11. Semicircle R	18. Pinch out	25. Swipe U-D-U (3F)
5. Swipe D-L	12. Bracket R	19. X shape	
6. Swipe D-U	13. Swipe D (3F)	20. Swipe L-R-L	
7. Circle	14. Swipe R-U	21. Rotor CCW	

4.1.2 Gesture prior knowledge and preferences

Most of the reference gestures we used were not previously known to most of the participants. At least one participant reported prior knowledge on 14 reference gestures (Table 2), and participants reported prior knowledge in 151 out of 900 instances (36 participants × 25 gesture types). The best-known gestures were swipe left and swipe up, which half of the participants declared to have known or used prior to the study. Eleven gestures were not known or used by any of the participants, particularly those with multiple fingers, strokes, or angles. Regarding the number of simultaneous fingers 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 said they preferred to use at most three fingers. Of the 25 blind participants, 15 preferred one-finger gestures, while those with low-vision were divided between no preference (4) and one-finger gestures (7). For the number of strokes per gesture, 19 preferred one-stroke gestures, 12 had no

Table 2 Gestures that at least one of the 36 participants declared knowing prior to the study

Gesture	Freq. of prior knowledge	Gesture	Freq. of prior knowledge)
Swipe L	18	*Swipe D (3F)*	13
Swipe U	18	*Pinch in*	11
Double tap (2F)	16	*Z shape (2F)*	5
Rotor CCW	15	*Swipe D-L*	4
Rotor CW	15	*Swipe D-U*	3
Swipe R (2F)	15	*Swipe R-U*	3
Pinch out	13	*Circle*	2

preference, and five preferred at most two strokes. In both low-vision and blind participant groups, single-stroke gestures were preferred.

Sixteen participants preferred rounded gestures, ten had no preference, six preferred steep-angle shapes, and only two participants preferred right-angled gestures. Most of the participants preferred rounded gestures in all of the four participant sub-groups. In the case of the angled swipe gesture group, we noticed a difference in perceived difficulty between low-vision and blind participants. The perceived difficulty mean level of participants with low vision for angled swipes was: 1.18 (SD = 0.5) for steep angles; 1.45 (SD = 0.67) for rounded angles; and 1.63 (SD = 1.05) for right angles. However, the perceived mean difficulty level of blind participants for these gestures is less spread out: 1.18 (SD = 0.48) for steep angles; 1.12 (SD = 0.38) for rounded angles; and 1.22 (SD = 0.46) for right angles. We would like to note that both groups had a similar difficulty in perception of steep angles, but blind participants perceived rounded angles as slightly easier than steep angles. The increase in perceived difficulty is also proportional to the gesture length: both groups made longer right-angled swipes and both agreed that this is the most difficult of the three angled pairs.

4.2 Absolute features of gestures

4.2.1 Kinematic and bounding box related features

We analyzed length and speed for all valid iterations, except for the two-finger double tap gesture for which speed is not relevant. Although the gesture path length of participants with low vision was 14 % longer than that of blind participants, the respective speed (in pixels per milliseconds) was also 28 % faster. The mean aspect ratio (the proportion of width to height) of the gesture bounding boxes for participants with low vision was 2.03; for blind participants it was 1.73. Regarding the gestures and the screen borders, of the 4763 valid captures 503 had at least one contact point near the display borders (within 3 pixels). Blind participants had a significantly higher percentage of contact points around the border for valid iterations (12.06 %), compared to low-vision participants (7.25 %). In addition, multi-touch iterations had a higher percentage of at least one contact point near the screen borders at 13 % (15.1 % for blind participants), compared to 8.9 % of one-finger gestures (10.5 % for blind participants).

We also analyzed the bounding box center of all the participants' gestures. Given that smartphone displays use screen coordinates in which the origin is at the top left corner, we would like to note that y values increase from top to bottom of the origin (the opposite of

Cartesian coordinates), and x values increase from left to right from the origin (as in Cartesian coordinates). We found no significant difference along the screen's horizontal axis (1080 pixels wide) between blind ($x = 521$, SD = 126 pixels) and low-vision participants ($x = 518$, SD = 88 pixels): Welch unequal variances two sample t-test of t (4024.4) = 0.8866, $p = .375$. Both groups had the mean gestures' center slightly left of the center of the horizontal axis of the screen and the device by 18 to 21 pixels. However, in the vertical axis (1920 pixels wide) we found a significant difference between the two groups: Welch unequal variances two sample t-test of t (3738.4) = 11.754, $p = < .001$. As illustrated in Fig. 5a, the center of the gesture bounding boxes was in a lower position for blind participants ($y = 1068$, SD = 262 pixels) compared to low-vision participants ($y = 987$, SD = 200 pixels). This difference is most noteworthy in the bracket up gesture (Fig. 5b). Both groups' centers were in a lower position with respect to the middle of the screen, located at $y = 960$. However, the middle of the device does not match the middle of the screen because of a 4-millimeter difference (approximately 70 pixels) between the top and bottom margins from the device's border to the screen (c.f. section 3.3). Therefore, offsetting the gesture's center by 35 pixels upward, low-vision participants are above the middle of the device by 8 pixels, and blind participants are below it by 73 pixels.

4.2.2 *Swipe gesture group*

Based on the previous set of features for the swipe gesture group, we found noteworthy (but not significant) differences among the four participant groups. For example, 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 51 %) and faster (by 22 %) than those in the horizontal direction. Concerning multi-finger swipes, we also noted that people blind since birth had the highest aspect ratio (wide bounding boxes) and lowest sharpness (sum of the absolute value of the angles at each point, measured in radians [45]), and people with low vision had the lowest aspect ratio and highest sharpness. However, a Kruskal-Wallis one-way analysis of variance does not confirm significant differences among the four visual impairment sub-groups regarding these two features. Still, sharpness differences are more noteworthy than aspect ratio; like the case of the gesture two-finger swipe right, with $\chi^2_{(3)} = 4.812$, $p > .186$ for aspect ratio, and $\chi^2_{(3)} = 6.1659$, $p > .103$ for sharpness. For all of the swipe gestures, the sharpness of participant sub-groups were: became blind in adulthood, mean = 2.345 rad,

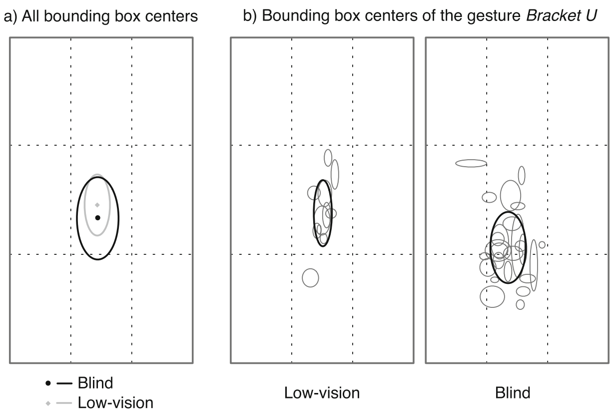


Fig. 5 The bounding *box centers* of blind participants' gestures were lower (**a**), especially for *bracket up* (**b**). *Ellipses' axis lengths* represent the standard deviation of the center

SD = 2.474 rad; blind since adolescence, mean = 0.680 rad, SD = 0.3962 rad; blind since birth, mean = 1.9352 rad, SD = 2.4893 rad; low vision, mean = 1.0389 rad, SD = 1.7460 rad.

4.2.3 Angled swipe group

There are important shape differences in the graphical representation of the participants' mean angled swipes (Fig. 6), especially on steep angle gestures. In this figure, there are also noteworthy differences between people with low vision or who became blind in adulthood, and people blind since an early stage of life. However, despite the high standard deviations on sharpness among participants, we only found participant sub-group significant difference in the swipe chevron up (Table 3).

4.2.4 Taps

We only had one tap gesture in our reference set, two-finger double tap, which was rated by the participants as the second easiest reference gesture. The respective main mean features of each participant tap gesture are depicted in Fig. 7, in which the characteristics of the gesture are represented as follows:

- *Circle:* Each circle represents two consecutive taps for a distinct finger. Because the Android system we used sets a pointer index chronologically for each pointer set, the index of the pointers might not match between taps. Therefore, we grouped the four pointers in two pairs of consecutive pointers based on the closest Euclidean distance between the centers of their respective bounding boxes. The diameter of each circle

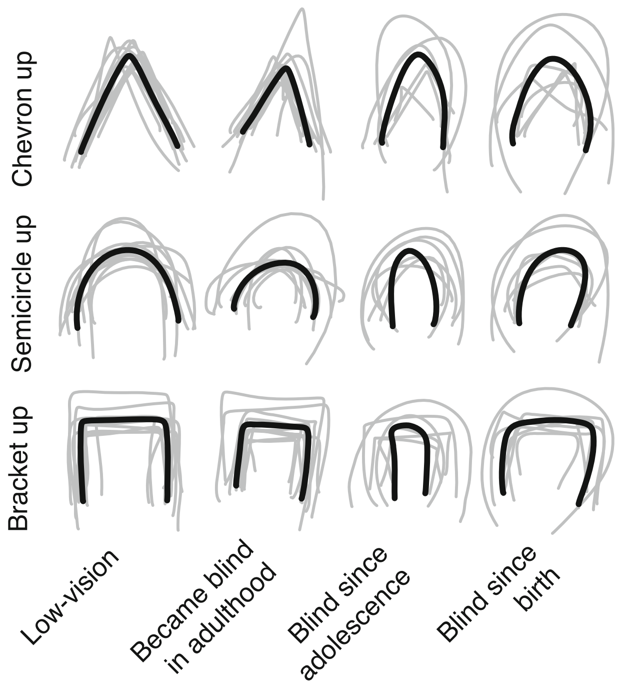


Fig. 6 Participant mean gestures for *vertical angled* swipes by sub-group

Table 3 Kruskal-Wallis one-way analysis of variance by participant sub-group for sharpness in angled swipe gestures (each participant's iterations were aggregated for sub-group and gesture)

Gesture	mean $_{n=36}$ (rad)	sd $_{n=36}$ (rad)	$\chi^2_{(3)}$	p
Chevron up (open at the bottom)	2.8034	2.3654	13.610	< .004
Chevron right (open at the left	4.0718	2.9955	1.8909	> .595
Semicircle up	2.3338	3.1085	2.7577	> .430
Semicircle right	2.4564	2.1508	1.8107	> .612
Bracket up	4.4460	2.8506	3.5576	> .313
Bracket right	4.9567	3.5059	7.0933	> .068

corresponds to this distance and their center is at the middle point between these two bounding boxes. The bigger the circle, the more displacement of the fingers between consecutive taps.

- *Line:* The connecting line between the centers of two circles represents the mean separation, in pixels, of the grouped pointers (i.e., fingers). From the slope of this line, calculated from the circle closer to the origin, we can also calculate the mean angle between the two fingers, relative to the x-axis

The main gesture features' means (Table 4) for all of the participants were: 288.5 ms (SD = 94) for duration, 56857 squared pixels for the bounding box area (SD = 34906) with an aspect ratio of 2.638 (SD = 2.56), 820.3 pixels (SD = 294.4) for the separation line of the two fingers, and an angle of 28.3° (SD = 21.6). We noticed that four participants had a separation line with negative angles. We realized that this is due to the way these participants held the device. When holding the smartphone with their left hand and tapping with their right hand, the separation line angle between the two pointers (usually the index and middle fingers) is typically positive, but it is typically negative when switching hands. We also noticed a difference between people who became blind in adulthood or had low vision, and people blind since birth or adolescence. The former sub-group pair takes longer to make the gesture (23 % more), with a larger area (by 87 %) and a smaller aspect ratio (by 48 %), compared to those blind since an early stage of life. There are no noteworthy differences among participant groups concerning the separation distance between the two fingers.

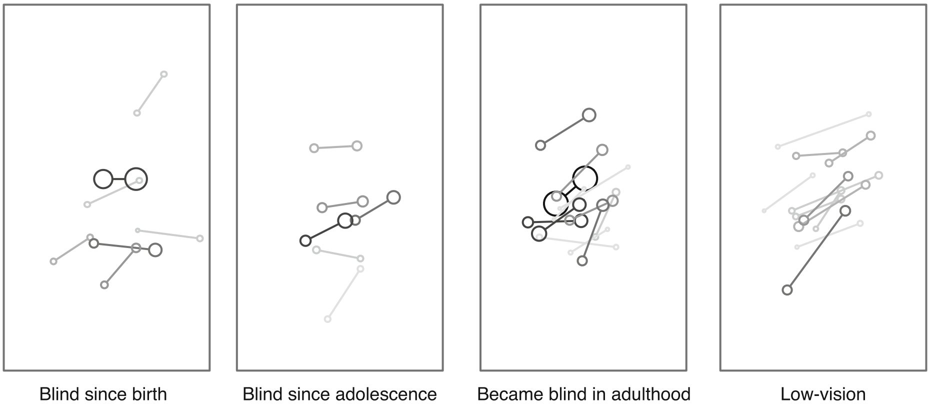


Fig. 7 Screen positioning of *double-tap* gestures by participant sub-group

Table 4 Comparison of features among participant sub-groups for two finger duble-taps

Sub-group	area (px^2)		aspect ratio		duration (ms)	
	mean	*sd*	*mean*	*sd*	*mean*	*sd*
Blind since birth	38981	9778	3.152	2.10	259.2	16.09
Blind since adolescence	34531	8321	4.512	5.58	242.7	75.3
Became blind in adulthood	57098	13418	2.075	0.93	319.9	89.5
Low vision	80147	18865	1.903	0.51	297.8	30.3
All sub-groups	*56857*	*14103*	*2.638*	*2.52*	*288.5*	*63.6*

4.2.5 Direction differences in circle and right-angle swipe gestures

To explore possible differences in gesture performance based on direction, we chose two swipe gestures with similar shapes but opposite directions, swipe right then up and swipe down then left. However, as detailed in Table 5, we did not find significant differences between these gesture types, neither for length, duration, speed, area, nor for aspect ratio. Regarding the circle shape, only a third of the 36 participants performed the gesture counter-clockwise, while the rest of the participants did it clockwise. Of the 25 participants who reported being right-handed, 18 drew the circle clockwise; the 10 participants who reported mixed-handedness were evenly split between the two directions; and the only self-reported left-handed participant performed the gesture clockwise.

4.3 Relative features of gestures

The MAD lab of the University of Washington has released several tools for gesture analysis. We use a couple of these tools to study the relative features of the sampled gestures. Because these tools were used primarily for single-stroke and multi-stroke gestures based on the sets for \$1 and \$p respectively, we only transformed and used the same kind of gestures for these analyses; that is, we do not include multi-touch gestures. The reference template is calculated as the average of all the gestures of the same type and participant group.

Table 5 Comparison of features between gestures swipe right then up and swipe down then left

Swipe feature	down then left		right then up		*t*-test	
	mean	*sd*	*mean*	*sd*	t_{35}	$p >$
Length (pixels)	1448.4	416.4	1528.5	434.8	1.531	.134
Duration (ms)	1069.2	601.9	1132.9	694.5	1.006	.320
Speed (pixels/ms)	1.666	0.713	1.697	0.742	0.512	.611
Area (sq. pixels)	607784	320148	670299	337311	1.546	.130
Aspect ratio	0.7698	0.2777	0.8849	0.3641	1.943	.061

4.3.1 Gesture consistency performance

Participants had the freedom to perform, in the direction and stroke order that they preferred, three of the reference gestures: circle, plus shape, and x shape. In the case of the circle, the starting point and direction (CW or CCW) was left up to the participants, which offered an opportunity for vast gesture variations. In the case of plus and x shapes, both are composed of two strokes that could have four possible starting directions. Thus, for the idealized references of these gestures (i.e., with perfect straight or diagonal lines), there are eight variations of each reference gesture. To study the participants' gesture performance consistency, we used GECKo (GEsture Clustering toolKit), an open source software utility for gesture analysis [3]. First, the tool automatically clusters participants' gesture iterations based on the Euclidian distance of the points, then each gesture is given an agreement rate, from zero (each gesture is unique) to one (all gestures are identical in form) [60]. This rate is calculated based on the gesture's number of clusters and the iterations that each cluster contains. For our analysis, we used the default value for the size of clusters and we did not perform manual correction of the clusters.

The resulting mean consistency among all the participants was relatively low at 0.211. Among blind participants it was 0.235: the circle gesture had the lowest consistency (0.065), followed by the x shape (0.273), and plus shape (0.367). For low-vision participants the mean was higher at 0.321, and the order was the same as blind participants: circle (0.124), x shape (0.24), and plus shape (0.599). For both participant groups the plus shape cluster size was the lowest among the three gestures (i.e., it had fewer variations): low-vision participants had only two variations (for 11 people), while blind participants had four (for 24 people). The circle gesture in both cases had almost one form variation per participant. Although the x shape has similar characteristics to the plus shape, it had 66 % more form variations. Between cardinal directions and inter-cardinal (in diagonal) directions, participants were more consistent with the former kind (Fig. 8).

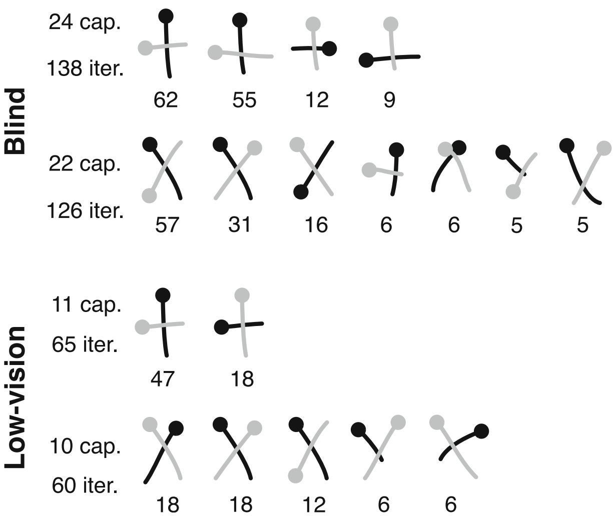


Fig. 8 Cluster variations and size for the valid captures (cap.) and iterations (iter.) of the gestures *plus shape* and *x shape*

4.3.2 Gesture relative accuracy

To improve the descriptive power of gesture beyond absolute features, Vatavu et al. propose a set of twelve new relative accuracy measures for stroke gesture articulation that characterize the geometric, kinematic, and articulation accuracy of single and multi-stroke gestures relative to a gesture task axis [54]. The gesture task axis is a representative way to articulate a stroke gesture used to compute these accuracy measures as local deviations from it. The proposed relative accuracy measures are the following:

- Shape Error (ShE): Average spatial deviation from a reference gesture.
- Shape Variability (ShV): Total spatial deviation from a reference gesture.
- Length Error (LE): Amount of *stretch* relative to a reference gesture.
- Size Error (SzE): Amount of space consumed relative to a reference gesture.
- Bending Error (BE): Average *turn* relative to a reference gesture.
- Bending Variability (BV): Total *turn* relative to a reference gesture.
- Time Error (TE): Average temporal deviation from a reference gesture.
- Time Variability (TV): Total temporal deviation from a reference gesture.
- Speed Error (VE): Average deviation in speed compared to a reference gesture.
- Speed Variability (VV): Total deviation in speed compared to a reference gesture.
- Stroke Count Error (SkE): Difference in number of strokes compared to a reference gesture.
- Stroke Ordering Error (SkOE): How similar the stroke ordering is compared to a reference gesture.

To compute these accuracy measures we used the open-source tool GREAT (Gesture Relative Accuracy Toolkit), released by the authors along with the paper describing their work. We computed in a user-independent manner the above relative measures for single-touch single-stroke (using the chronological order of input) and multi-stroke gestures (using the point-cloud matching procedure of $P). We did it separately for both blind and low-vision participants. We selected the centroid of the gesture iterations as the task axis of reference. We would like to note that since we automatically marked as incorrect gestures that had a different number of strokes compared to the reference gesture, the SkE is not relevant. In the following paragraph we report on the relative measures that revealed the most differences in gesture execution among the participants.

In general, blind participants had less deviation from the reference gesture compared to low-vision participants. For instance, the results indicate that the average deviation in speed (VE) compared to a reference gesture of blind participants is less than that of low-vision participants by 19 %, both in single-stroke and multi-stroke gestures, having an almost equal standard deviation. Given that low-vision participants performed gestures 28 % faster than blind participants, the variability is also higher [54]. However, there were a few measures in which the deviation from the task axis in blind participants was higher than low-vision participants (Fig. 9), particularly on the BE of angled swipes (by 21 %). In particular, the similarity of multi-stroke gestures for blind participants was SkOE = 5644 (SD = 6114), and for participants with low vision it was SkOE = 11066 (SD = 7394). These results are coherent with the difficulties that blind participants have with multi-stroke gestures and non-rounded angles.

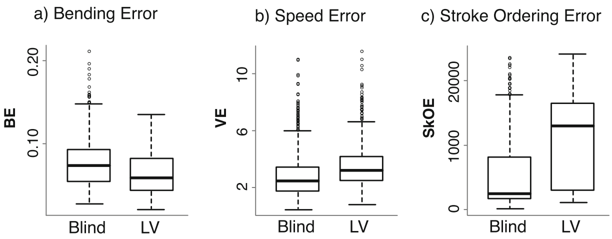


Fig. 9 Relative geometric and kinematic measures with important differences between blind and low-vision (LV) participants

5 Discussion

5.1 Overall results

In general, participants rated gestures with multiple fingers and strokes or greater length as more difficult. These results are consistent with previous studies, both for sighted [41] and visually impaired people [20]. We think that the overall low rating was due primarily to minimum feedback on their gesture performance accuracy, as we only let the participants know if they had used (usually unknowingly) the incorrect number of pointers, number of strokes or wrong gesture direction. However, participants 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 immediately. 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 on smartphones (e.g., speech recognition), or due to the low familiarity with touchscreen devices of some of the participants. Besides, we also think this result is not surprising because many people who became blind at an early age did not learn print characters or symbols [20].

In addition, blind people had difficulty with gestures that have steep or right angles. They performed these gestures in a more rounded manner than did participants with low vision. In effect, both groups prefer rounded gestures over otherwise angled gestures. Our results also suggest noteworthy differences regarding angled gestures among the four different sub-groups of visually impaired people. However, so far we do not conclude that these differences are significant, with the exception of the chevron up gesture. On the other hand, we found significant differences between people with low vision and blindness. For instance, blind participants performed the gestures at a lower speed, and they positioned their gestures lower on the screen compared to low-vision people.

5.2 Capture issues

As stated in the section 4, some of the participants had issues with certain gestures and they did not have all their captures or iterations marked as valid. 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 with at most three participants was scheduled to last around 75 min), and to avoid the frustration of the participants after a few repetitions, in the latter cases we decided to proceed to the following gesture type. The lack of tactile edges on the device's display was one of the prevalent difficulties, especially for blind participants. If the participants performed a gesture outside the boundaries of the display, our Android application would incorrectly consider the gesture as finished. This was most notorious in the case of gestures with several changes in direction (e.g., to-and-fro swipes group).

We had anticipated the occurrence of this issue, but we decided not to use a screen overlay [19] so we could better understand how some gestures are performed around the boundaries of the screen in real-life situations. In many cases, we noticed the participants' fingernails (not the fingertips) would make contact with the screen, so the gesture was not registered as intended. This problem was more prevalent in women than men, mainly because of long fingernails. Furthermore, we think this explains the higher percentage of women who had their captures marked as invalid compared to men. Above all, 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 [35]. For example, some participants would invert the directions of the gesture while capturing it. For instance, they would make a swipe up and down in one iteration, and a swipe down and up in the next one. Indeed, one of the participants stated that the difference between a person blind since birth and one who became blind in adulthood in terms of smartphone use performance may be blurred depending on personal experience, interest, and personal aptitudes.

5.3 Study limitations

We realize that the mean age of our participants, 48 years (SD = 15.8), implies an underrepresentation of younger visually impaired people. Although we have gathered and analyzed valuable data in this study, this issue is further exacerbated by the significant differences related to age regarding touch-based interaction [2, 13], even when comparing blindfolded participants to blind participants [40]. Moreover, given the positive effect of practice in gesture consistency [3], 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. Furthermore, younger generations grow 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 [49]. We hope that our future research efforts will be able to overcome these issues, working with more comprehensive groups of participants. In addition, the difficulty perceiving the gestures is likely to be skewed by a certain degree because of the increasing order we used for the capture sessions. People are more likely to judge a better option (e.g., less difficult) more positively after a less favorable option, and to judge more negatively a less favorable option (e.g., more difficult) after a better option [29]. Although we were aware of this potential issue, we preferred to maintain an increasing order of difficulty for all participants to minimize eventual frustrations during the sessions.

5.4 Gesture suggestions

Based on the results on the gesture preferences and features in this study, we offer a few recommendations for choosing and designing gestures that are easier for visually impaired people to perform.

- *Avoid multi-touch gestures*. Smartphone screens vary in size, but they are relatively small compared to desktop computers or tablets. Thus, multi-touch gestures are more likely to go outside the boundaries of the screen especially if the display is not physically delimited. Multi-touch taps are generally an exception, as there is no movement of the fingers across the screen. Nonetheless, try to keep a low finger count.
- *Prefer single-stroke gestures*. The sequence of multi-stroke gestures must be performed within a brief period of time, which could be detrimental to the accuracy of the subsequent strokes. In addition, with each additional stroke, gestures performed by blind users are more likely to deviate from the template, as blind people have different spatial awareness compared to sighted people. Again, tap gestures are generally an exception because the form of the gesture is not relevant.
- *Favor short gestures*. Long gestures are perceived as more difficult and they are more prone to deviate from the gesture template. In addition, long gestures are more likely to go outside the boundaries of the smartphone display.
- *Assign cardinal directions to gestures when possible*. Visually impaired participants produce gestures with cardinal directions, across the sides of the screen, more consistently. Moreover, inter-cardinal directions along the screen diagonals are not only more prone to deviate from the gesture template, but they are usually longer and thus more difficult.
- *Prefer rounded angles for more complex gestures*. Blind users may have difficulty executing gestures with right or steep angles, especially across the narrower axis of the screen. Rounded gestures are perceived as less difficult, they are easier to perform, and visually impaired people prefer them to other angled gestures.

6 Conclusions

In this paper, we have presented an extended analysis of gesture performance and preferences regarding touch-based smartphones, among people blind from different stages of life or with low vision. We also provided an overview of the underlying mobile gestural interaction, including accessibility aspects, as well as how gestures' features influence the recognition techniques to be used. We have also presented a novel approach and system to wirelessly capturing touch-based gestures from several users at the same time on mobile devices, using web technologies that are cross-platform compatible. Overall, visually impaired people generally prefer simple gestures: one finger, one stroke, preferably in one direction or with round angles otherwise. We also found significant differences between blind people and those with low vision. Although differences regarding the age of onset of blindness are noteworthy, we did not find them to be significant (except for the chevron up gesture). Finally, we proposed suggestions for the selection and design of usable and accessible touch-gestures on handheld mobile devices based on our results.

We would like to note that more research needs to be done involving younger visually impaired people, especially because they are more likely to already use touchscreen mobile devices. Unfortunately, in our case the number of young participants was limited, given the difficulties of recruiting visually impaired people. We hope that future initiatives will be able to surmount this limitation. Regarding future research, we think that it is worth comparing visually impaired people's performance of gestures across different sizes of touch screens. We hypothesize that the size might have a significant influence on what is considered to be a more usable gesture (e.g., more space available, the possibility of using both hands at the same time). Finally, we hope that our approach, methodology, and results, as well as the discussion of the issues we encountered during the study, will be useful to other researchers working to advance the accessibility and usability of mobile interfaces.

References

1. Albinsson P-A, Zhai S (2003) High precision touch screen interaction. In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, p 105–112
2. Anthony L, Brown Q, Nias J, Tate B, Mohan S (2012) Interaction and recognition challenges in interpreting children's touch and gesture input on mobile devices. Paper presented at the Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces, Cambridge, Massachusetts, USA
3. Anthony L, Vatavu R-D, Wobbrock JO (2013) Understanding the consistency of users' pen and finger stroke gesture articulation. Paper presented at the Proceedings of Graphics Interface 2013, Regina, Sascatchewan, Canada
4. Anthony L, Wobbrock JO (2010) A lightweight multistroke recognizer for user interface prototypes. In: Proceedings of Graphics Interface 2010. Canadian Information Processing Society, p 245–252
5. Anthony L, Wobbrock JO (2012) $ N-protractor: a fast and accurate multistroke recognizer. In: Proceedings of Graphics Interface 2012. Canadian Information Processing Society, p 117–120
6. Azenkot S, Lee NB (2013) Exploring the Use of Speech Input by Blind People on Mobile Devices. Paper presented at the 15th International ACM SIGACCESS Conference on Computers and Accessibility, New York, NY, USA
7. Bayus BL, Jain S, Rao AG (1997) Too little, too early: introduction timing and new product performance in the personal digital assistant industry. J Mark Res 34(1):50–63
8. Benyon D (1993) Accommodating individual differences through an adaptive user interface. Hum Factors Inform Technol 10:149
9. Bragdon A, Nelson E, Li Y, Hinckley K (2011) Experimental analysis of touch-screen gesture designs in mobile environments. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, p 403–412
10. Buzzi MC, Buzzi M, Leporini B, Trujillo A (2014) Designing a text entry multimodal keypad for blind users of touchscreen mobile phones. In: Proceedings of the 16th international ACM SIGACCESS conference on computers & accessibility. ACM, p 131–136
11. Buzzi MC, Buzzi M, Leporini B, Trujillo A (2015) Exploring visually impaired people's gesture preferences for smartphones. In: Proceedings of the 11th Biannual Conference on Italian SIGCHI Chapter. ACM, p 94–101
12. Charland A, Leroux B (2011) Mobile application development: web vs. native. Commun ACM 54(5):49–53. doi:10.1145/1941487.1941504
13. Findlater L, Froehlich JE, Fattal K, Wobbrock JO, Dastyar T (2013) Age-related differences in performance with touchscreens compared to traditional mouse input. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France
14. Goldberg D, Richardson C (1993) Touch-typing with a stylus. In: Proceedings of the INTERACT′93 and CHI′93 conference on Human factors in computing systems. ACM, p 80–87
15. Guerreiro T, Lagoá P, Nicolau H, Santana P, Jorge J (2008) Mobile text-entry models for people with disabilities. Paper presented at the 15th European conference on cognitive ergonomics
16. Heo S, Gu J, Lee G (2014) Expanding touch input vocabulary by using consecutive distant taps. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, p 2597–2606
17. Kammer D, Wojdziak J, Keck M, Groh R, Taranko S (2010) Towards a formalization of multi-touch gestures. In: ACM International Conference on Interactive Tabletops and Surfaces. ACM, p 49–58
18. Kane SK, Bigham JP, Wobbrock JO (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, p 73–80
19. Kane SK, Morris MR, Wobbrock JO (2013) Touchplates: low-cost tactile overlays for visually impaired touch screen users. Paper presented at the 15th International ACM SIGACCESS
20. Kane SK, Wobbrock JO, Ladner RE (2011) Usable gestures for blind people: understanding preference and performance. Paper presented at the SIGCHI Conference
21. Kendall MG, Smith BB (1939) The problem of m rankings. p 275–287. doi:10.1214/aoms/1177732186
22. Kristensson P-O, Zhai S (2004) SHARK 2: a large vocabulary shorthand writing system for pen-based computers. In: Proceedings of the 17th annual ACM symposium on user interface software and technology. ACM, p 43–52

23. Lettner F, Holzmann C (2012) Heat maps as a usability tool for multi-touch interaction in mobile applications. In: Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia. ACM, p 49
24. Li Y (2010) Protractor: a fast and accurate gesture recognizer. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, p 2169–2172
25. Long Jr AC, Landay JA, Rowe LA, Michiels J (2000) Visual similarity of pen gestures. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems. ACM, p 360–367
26. Luthra V, Ghosh S (2015) Understanding, evaluating and analyzing touch screen gestures for visually impaired users in mobile environment. In: Universal Access in Human-Computer Interaction. Access to Interaction. Springer, p 25–36
27. MacKenzie IS, Zhang SX (1997) The immediate usability of Graffiti. In: Graphics Interface. p 129–137
28. McGookin D, Brewster S, Jiang W (2008) Investigating touchscreen accessibility for people with visual impairments. Paper presented at the 5th Nordic conference on Human-computer interaction
29. Moore DA (1999) Order effects in preference judgments: evidence for context dependence in the generation of preferences. Organ Behav Hum Decis Process 78(2):146–165
30. Morris MR, Danielescu A, Drucker S, Fisher D, Lee B, Schraefel MC, Wobbrock JO (2014) Reducing legacy bias in gesture elicitation studies. Interactions 21(3):40–45. doi:10.1145/2591689
31. Morris J, Mueller J (2014) Blind and deaf consumer preferences for android and iOS smartphones. In: Langdon PM, Lazar J, Heylighen A, Dong H (eds) Inclusive Designing. Springer International Publishing, p 69–79. doi:10.1007/978-3-319-05095-9_7
32. Myers CS, Rabiner LR (1981) A comparative study of several dynamic time-warping algorithms for connected-word recognition. Bell Syst Tech J 60(7):1389–1409
33. Oh U, Findlater L (2013) The challenges and potential of end-user gesture customization. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, p 1129–1138
34. Oh U, Kane SK, Findlater L (2013) Follow That sound: using sonification and corrective verbal feedback to teach touchscreen gestures. Paper presented at the 15th International ACM SIGACCESS, New York, NY, USA
35. Oliveira J, Guerreiro T, Nicolau H, Jorge J, Gonçalves D (2011) Blind people and mobile Touch-based Text-entry: acknowledging the need for different flavors. Paper presented at the 13th International ACM SIGACCESS Conference on Computers and Accessibility, New York, NY, USA
36. Oviatt S, Cohen P, Wu L, Duncan L, Suhm B, Bers J, Holzman T, Winograd T, Landay J, Larson J (2000) Designing the user interface for multimodal speech and pen-based gesture applications: state-of-the-art systems and future research directions. Hum Comput Interact 15(4):263–322
37. Pittman JA (1991) Recognizing handwritten text. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, p 271–275
38. Plamondon R, Srihari SN (2000) Online and off-line handwriting recognition: a comprehensive survey. IEEE Trans Pattern Anal Mach Intell 22(1):63–84
39. Plimmer B, Crossan A, Brewster SA, Blagojevic R (2008) Multimodal collaborative handwriting training for visually-impaired people. Paper presented at the SIGCHI Conference
40. Postma A, Zuidhoek S, Noordzij ML, Kappers AM (2007) Differences between early blind, late blind and blindfolded sighted people in haptic spatial configuration learning and resulting memory traces. Perception 36(8):1253–1265
41. Rekik Y, Vatavu R-D, Grisoni L (2014) Understanding users' perceived difficulty of multi-touch gesture articulation. In: Proceedings of the 16th International Conference on Multimodal Interaction. ACM, p 232–239
42. Rivera J, van der Meulen R (2014) Gartner says sales of smartphones grew 20 percent in third quarter of 2014. Gartner, Egham
43. Rodrigues A, Montague K, Nicolau H, Guerreiro T (2015) Getting smartphones to TalkBack: understanding the smartphone adoption process of blind users. Proceedings of ASSETS′15
44. Romano M, Bellucci A, Aedo I (2015) Understanding touch and motion gestures for blind people on mobile devices. In: Human-Computer Interaction–INTERACT 2015. Springer, p 38–46
45. Rubine D (1991) Specifying gestures by example. Paper presented at the proceedings of the 18th annual conference on computer graphics and interactive techniques
46. Ruiz J, Li Y, Lank E (2011) User-defined motion gestures for mobile interaction. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada
47. Sandnes F, Tan T, Johansen A, Sulic E, Vesterhus E, Iversen E (2012) Making touch-based kiosks accessible to blind users through simple gestures. Univ Access Inf Soc 11(4):421–431. doi:10.1007/s10209-011-0258-4
48. Schmidt M, Weber G (2009) Multitouch haptic interaction. In: Universal access in human-computer interaction. Intelligent and ubiquitous interaction environments, vol 5615. Lecture notes in computer science. Springer, Berlin Heidelberg, pp 574–582. doi:10.1007/978-3-642-02710-9_64
49. Sears A, Hanson V (2011) Representing users in accessibility research. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada

50. Sezgin TM, Davis R (2005) HMM-based efficient sketch recognition. In: Proceedings of the 10th international conference on intelligent user interfaces. ACM, p 281–283
51. Spano LD, Cisternino A, Paternò F (2012) A compositional model for gesture definition. In: Human-Centered Software Engineering. Springer, p 34–52
52. Vatavu R-D, Casiez G, Grisoni L (2013) Small, medium, or large?: estimating the user-perceived scale of stroke gestures. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France
53. Vatavu R-D, Anthony L, Wobbrock JO (2012) Gestures as point clouds: a $ P recognizer for user interface prototypes. In: Proceedings of the 14th ACM international conference on multimodal interaction. ACM, p 273–280
54. Vatavu R-D, Anthony L, Wobbrock JO (2013) Relative accuracy measures for stroke gestures. Paper presented at the Proceedings of the 15th ACM on international conference on multimodal interaction, Sydney, Australia
55. Vatavu R-D, Anthony L, Wobbrock JO (2014) Gesture Heatmaps: understanding gesture performance with colorful visualizations. Paper presented at the Proceedings of the 16th International Conference on Multimodal Interaction, Istanbul, Turkey
56. Vidal S, Lefebvre G (2010) Gesture based interaction for visually-impaired people. In: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries. ACM, p 809–812
57. Walker G (2012) A review of technologies for sensing contact location on the surface of a display. J Soc Inf Disp 20(8):413–440
58. Westerman W, Elias JG, Hedge A (2001) Multi-touch: A new tactile 2-d gesture interface for human-computer interaction. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol 6. SAGE Publications, p 632–636
59. Willems D, Niels R, van Gerven M, Vuurpijl L (2009) Iconic and multi-stroke gesture recognition. Pattern Recogn 42(12):3303–3312
60. Wobbrock JO, Aung HH, Rothrock B, Myers BA (2005) Maximizing the guessability of symbolic input. In: CHI′05 extended abstracts on Human Factors in Computing Systems. ACM, p 1869–1872
61. Wobbrock JO, Wilson AD, Li Y (2007) Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. In: Proceedings of the 20th annual ACM symposium on User interface software and technology. ACM, p 159–168
62. Xu S (2015) Improving accessibility design on touchscreens. In: Universal Access in Human-Computer Interaction. Access to Interaction. Springer, p 161–173
63. Zhong Y, Raman T, Burkhardt C, Biadsy F, Bigham JP (2014) JustSpeak: enabling universal voice control on Android. In: Proceedings of the 11th Web for All Conference. ACM, p 36

Maria Claudia Buzzi, MD National Council Research (CNR), Institute of Informatics and Telematics (IIT). Research focus on Web, eHealth, eGovernment and user interface design. She investigates techniques and methods to make the user interfaces accessible and usable to users with special needs. Coordination of technological and research projects. She is member of Program Committees of many International Conferences and is also involved in the management of the Italian W3C Office. http://www.iit.cnr.it/claudia.buzzi

Marina Buzzi, MD National Council Research (CNR), at Institute of Informatics and Telematics (IIT). Head of the Web Application for the Future Internet Lab at IIT. Research activities on Web usability and IoT, eHealth. Coordination of technological and research projects. She follows activities of the W3C and IETF groups. Professional member of ACM and SIGACCESS since 2004. Co-author of one hundred international publications. http://www.iit.cnr.it/marina.buzzi

Barbara Leporini, PhD Researcher at the CNR of Pisa (ISTI). She investigates techniques and methods to make the user interfaces - e.g. for Web and mobile applications, publishing, museum contents, educational and e-learning systems - accessible and usable to users with special needs. Beyond research, Barbara has been teaching computer sciences classes. She has also provided technical support for accessibility and usability at various levels such as in boards and groups working on different problems concerning visually-impaired people. http://giove.isti.cnr.it/Users/Barbara/it/index.html

Amaury Trujillo PhD student of Computer Science, University of Pisa. He obtained a Telematics B.Eng. in Mexico and an Information Systems M.Eng. in France. He is currently a PhD student in Computer Science at the University of Pisa, and does research at the Institute of Informatics and Telematics (IIT), of the Italian National Council Research (CNR). His current research interests focus on accessible human-computer interaction and eHealth technology applications. http://www.iit.cnr.it/en/amaury.trujillolarios