User Gesture Elicitation of Common Smartphone Tasks for Hand Proximate User Interfaces

Shariff A. M. Faleel
University of Manitoba
mohommas@myumanitoba.ca

Michael Gammon
University of Manitoba
michael.gammon@umanitoba.ca

Yumiko Sakamoto
University of Manitoba
umsakamo@umanitoba.ca

Carlo Menon
Simon Fraser University
cmenon@sfu.ca

Pourang Irani
University of Manitoba
irani@cs.umanitoba.ca

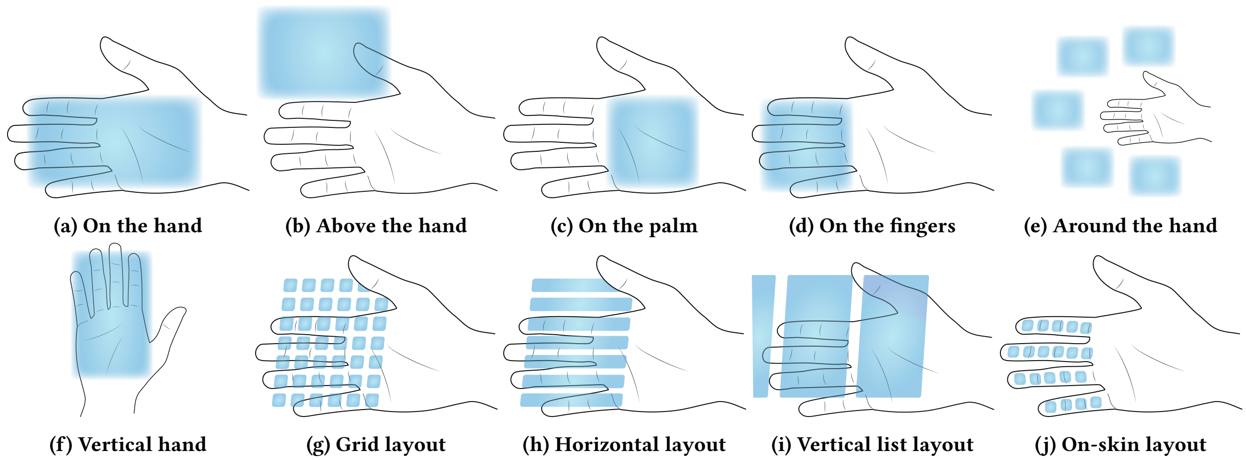


Figure 1: Proposed layouts from Study 1. Blue sections on the figure represent virtual interfaces. The figures (a)-(e) represent proposed general layouts. Figure (f) shows the vertical alignment of the hand. Figures (g) - (h) represent different ways to organize elements. (j) is a unique case (on-finger/grid layout) where the hand anatomy is taken into consideration.

ABSTRACT

The ubiquity of smartphone interactions along with the advancements made in mixed reality applications and gesture recognition present an intriguing space for novel interaction techniques using the hand as an interface. This paper explores the idea of using hand proximate user interfaces (UI), i.e. interactions with and display of interface elements on and around the hand. We conducted two user studies to gain a better understanding of the design space for such interactions. The first study identifies the possible ways in which various elements can be displayed on and around the hand in the context of common smartphone applications. We conduct a second study to build a gesture set for interactions with elements displayed on and around the hand. We contribute an analysis of the data and observations collected from the two studies, resulting in a layout set and a gesture set for interactions with hand proximate UIs.

CCS CONCEPTS

• **Human-centered computing** → **Gestural input**; **User studies**; **Mixed / augmented reality**; *Mobile devices.*

KEYWORDS

Mixed-reality interactions, One-handed interaction, Gestural input, User-defined gestures, Elicitation study

ACM Reference Format:
Shariff A. M. Faleel, Michael Gammon, Yumiko Sakamoto, Carlo Menon, and Pourang Irani. 2020. User Gesture Elicitation of Common Smartphone Tasks for Hand Proximate User Interfaces. In *11th Augmented Human International Conference (AH '20), May 27–29, 2020, Winnipeg, MB, Canada.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3396339.3396363

AH '20, May 27–29, 2020, Winnipeg, MB, Canada

ACM ISBN 978-1-4503-7728-7/20/05…$15.00
https://doi.org/10.1145/3396339.3396363

1 INTRODUCTION

Since personal smart-devices have become ubiquitous [4] over the past decade, many have wondered what the next step is for mobile productivity and entertainment. More recently, head-worn displays (HWDs) have shown potential as people's main mobile computing device. HWDs offer a much larger interaction space than traditional smart devices while facilitating more cohesive interactions with the user's surroundings (e.g., spatial mapping on the Microsoft Hololens)[11]. In the context of the interactions, one of the main issues is that the visual interface is suspended in mid-air which is uncomfortable/unnatural to interact with, both physically and socially.

To deal with this challenge, we propose to replace the physical mobile device with a virtual interface anchored on and around the user's hand, augmented through the use of a HWD. We define these User Interfaces as Hand Proximate UIs. Since smartphone interaction is already well understood in general, following its model will allow us to make use of the device's current functionality while expanding upon it. Presenting a mobile interface on and around a user's hand through hand tracking allows us to reap the benefits of both the smartphone (e.g., comfort, familiarity, social acceptability) and the HWD (e.g., larger design space for displays and interactions, better contextualization).

Since our hands are already the primary means of interaction with devices and the world, the familiarity and dexterity already exists for most users, along with higher levels of social acceptance [17, 37, 41], making it a natural place to anchor a display. While many previous studies explored using hands as an input method with the visualization displayed elsewhere, having the actual display anchored to the hand remains largely unexplored.

We focus on gaining insight for designing mixed reality (MR) hand proximate UIs via elicitation studies [43]. We explore two aspects of such interfaces; (a) displaying elements, and (b) interacting with the displayed elements. Based on the fact that previous studies found strong preferences for single-handed interaction over two-handed counterparts [7, 28, 37], we focus on single-handed interactions. Two studies were conducted: Study 1 derived the layouts for interfaces on and around the hand, while Study 2 explored the interaction techniques for the layouts developed in Study 1. Results from Study 1 showed participants' strong preferences for the interface to be displayed on and above the hand (Figure 1). Thus, these two layouts were explored further to elicit preferred gestures from the participants in Study 2. In total, five different layouts and 462 gestures were generated. The contributions of this paper are as follows: (a) a set of layouts for a virtual interface on and around a user's hand; (b) a set of gestures to consider when designing interfaces for MR where the interface is positioned on and around the hand; (c) insights into the contributing factors to optimal gestures in this context.

2 RELATED WORK

The literature exploring using hands as an interface primarily looks at hands as an input device. However, with improvements in MR technology, using the hand as an interface to display information is a possibility. In this section, we provide a review of the literature exploring using hands as an interface with MR application.

2.1 Gestural Interactions Using the Hand

Gestures and interaction techniques using hands have been proposed in various contexts with varying use of the hands (e.g., on-skin interactions [16, 41], broader gestures [7, 37], two-handed gestures [5, 14, 28], microgestures [6, 8], and thumb-to-finger gestures[19, 34, 36]). This work looks at single handed interactions, specifically, microgestures. Previous studies show that users prefer single-handed interactions over two handed interactions [7, 28, 37]. Furthermore, the common reasons for the exploration of single handed interactions in many studies were; the flexibility, ease of use, and social acceptability [19, 36]. This also coincides with the observations made when developing gestures for different technologies (hand-held mobile devices [21], surface computing [26, 43], immersive multimedia [30] and smart TV [10]), where the users prefer single-handed interaction whenever possible.

Several previous studies focus on identifying the functional workspace of microgestures. Studies conducted by Huang et al. [19], Tsai et al. [36], and Dewitz et al. [8] show that finger-to-finger interactions, specially the thumb-to-finger interactions, are considered more comfortable with single-handed microgestures compared to interacting with the palm. These studies propose using hand as a functional input space as opposed to treating it only as a gestural input modality. In identifying the gestural primitives for microgestures, studies identify touch/tap and swipe as the primary primitives [6, 34].

Notable in this body of literature is the emphasis on tactile interactions. Tactile interactions are necessary for eye-free interactions[9, 19, 44] or to disambiguate regular hand movement from a gesture. When an interface is displayed in a MR setting and hands are used for direct input, tactile interactions can be less important. Gustafson et al. [14] observed that visual cues, more than any other cues (including tactile), played a more important role in the users interactions with the hands when they can see the hands.

The strides made in hand pose estimation using deep learning techniques [25, 35] allow for more accurate detection of hand poses, and in turn hand interactions. This is ideal for HWDs, as no additional instrumentation will be required, and the hand interactions can be detected only using the cameras on the HWD. Yet, in MR applications, interacting with virtual elements using only hands has been a challenge [1].

2.2 User Elicitation Studies

Elicitation studies[43] have been widely used in different contexts to obtain user generated gestures during the design process. Gesture sets generated with this approach have been shown to be more user-friendly [26] and memorable [27] compared to gestures generated by exports. As this study focuses on single-handed interaction in MR applications, studies more relevant to this work are in elicitation studies for hand gestures, particularly single-handed microgestures [5, 6, 19], and elicitation studies for MR interactions [28]. This study also follows a similar design to gain insight into the interaction space when using the hand as an interface to display and interact.

2.3 Hand Proximate UIs in Head Worn Displays

Lee and Hui [23] provide a detailed survey on the literature for interaction methods for smart glasses. They identify a set of interaction goals for interaction techniques for smart glasses. They note that none of the interaction techniques studied attempt to address all identified interaction goals. The goals they propose can be further divided into two categories: (a) Interaction with 2D elements (e.g., menus), (b) interaction with 3D elements (e.g., objects placed in the environment). It is also notable that none of the interaction techniques address all goals for the 2D category. To the best of our knowledge, this still remains to be the case. In addition to our paradigm extending the smartphone interaction, it can also be perceived as addressing the goals for interactions with 2D elements. The body of work discussed in Section 2.1 and related work on text input using hands gestures[39, 40, 42, 44] provides a foundation for such an interface.

The primary form of interaction with 2D elements (such as menus) discussed in the literature are either hand gestures[15] or pointing [12, 20, 22, 29]. Compared to using an external device for pointing and selecting elements, using the head-movement or pointing with bare-hands is prone to fatigue [3, 18] and also shown to be less efficient compared to traditional point and click [13, 32]. Taking inspiration from the smartphone interaction model, we believe using hand proximate UIs to display and interact with 2D elements can alleviate this limitation. OmniTouch [16] proposes a system that is closely related to our proposal. It is a shoulder-mounted projector and depth sensor, where the interface is projected on the skin and interactions are on-skin touch interactions. PalmType [39] is an interface on the hand for typing. To define the area on the hand corresponding to each key, they use a design study which was then validated with a user study. This shows the need to consider the related nature of the interface and interaction technique for such an interface [23]. Digitouch [42] follows a similar rationale to ours, where they use the hand interactions for input in smart glasses, where they develop a glove to detect gestures while the interface is displayed on the smart glasses.

3 STUDY METHODOLOGY

We derived insight into hand proximate UIs by conducting two studies. Study 1 generated a set of layouts for displaying different elements. The two most commonly proposed layouts were then selected to be used in Study 2, which involved eliciting gestures for interacting with each layout. The remainder of this section details the procedure and results of each study. Note that the term "layout" is used to refer to all the different configurations of how various elements are placed or arranged.

3.1 Study 1: Deriving the Layouts

3.1.1 Selection of Tasks. The referents used in this study have their basis on previous works [2, 5, 31, 33]. We focused on a set of tasks that were common in simple mobile applications. Hence, we excluded activities such as editing and transforming [28]. Further, text entry was also excluded in this study due to the complex nature of the task [39, 40, 42, 44], which might interfere with the central goal of this study (i.e., the development of basic layouts for Hand Proximate User Interface). Researchers found that dichotomous referents (e.g., zoom in and zoom out) elicit a pair of gestures which are "reversible" in nature [28, 43]. Thus, such referents were combined into one referent (e.g., zoom in/out) in our study. The selected tasks for Study 1 are listed in Table 1.

3.1.2 Participants. For Study 1, participants with a human-computer interaction (HCI) background were recruited (N = 12; F = 4, M = 8, their age ranged between 24 and 33; M = 27.5, SD = 2.68). All of them were familiar with the use of HWDs. Participants with HCI background were recruited due to their expertise with the UI design process and their experience with designing UIs for various technologies (such as HWDs, smart watches, gestural interfaces, etc.).

3.1.3 Procedure. First, participants were greeted and asked to turn off any smart devices to avoid distractions. They were then asked to complete a general demographic questionnaire. Subsequently, the participants were asked to imagine designing a technology that allows them to display information on and around their hand in a MR setting. They were then asked to illustrate their design ideas for the five applications: Map, Browser, Media player, home screen (for opening a new application), and app switching screen (for switching to an open application) (Table 1); onto a black and white outline of the human hand printed on a paper while considering the respective referents as interaction scenarios. The home screen and app switching screen together will be referred to as *general applications* in the remainder of this paper.

3.1.4 Results. The descriptions for the different layouts were coded by two coders (Cohen's $kappa$ = .86), and any disagreement between them were resolved through discussion. The descriptions were coded along three dimensions: 1) the position of the interface relative to the hand, 2) if the participants preferred holding the hand vertically vs. horizontally, and 3) the layout of the elements being displayed. Note that the third dimension is only applicable to the general applications. The percentage of proposals for each category along each dimension can be seen in Figure 2. The layouts described could be categorized into five broader categories: (a) *on the hand* (Figure 1a) (b) *above the hand* (Figure 1b) (c) *on the palm* (Figure 1c) (d) *on the fingers* (Figure 1d) (e) *around the hand* (Figure 1e). Even though both *on the palm* and *on the fingers* can fall under the *on the hand* category, they were distinguished as some participants (Figure 2) clearly restricted the display to only one region of the hand. 18% of the total proposed layouts were oriented such that the hand can be held vertically (Figure 1f), as some participants found holding the hand vertically to resemble a smartphone screen and thus more comfortable. For the map, browser, and media player applications, the interfaces described were similar to that of the interfaces commonly seen on personal computers or smartphones. A larger variation was observed with the general applications. The third dimension was used during coding to capture this variation, which was primarily in terms of how the elements will be organized. As such, to visualize different elements with interactions in mind, five different layouts were proposed: 1) Arranging the elements in a grid (Figure 1g), 2) Using a horizontal list (Figure 1h), 3) Using a vertical list (Figure 1i), 4) Laying out the elements as they are on the skin, particularly such that they can be reached by the thumb (Figure 1j), 5) Displaying different elements around the hand (Figure 1e). Some participants preferred using gestures for some applications instead of interacting with an interface element; 1 participant for home screen and 4 participants for app switching screen (i.e., the gaps in Figure 2). Thus, with the layouts ranked, Study 2 used the top layouts as stimuli for each of the general applications.

Table 1: The referents used in Study 2 and the applications related to them with the agreement rate (AR) for each referent, and the consensus gestures for each referent. The percentage columns show the percentage of participants who produced the consensus gestures of the respective referent for a given condition (on-hand, off-hand). Refer to Figure 4 for descriptions of the gestures listed under *consensus gesture.* Agreement rates are highlighted based on the categorization Vatavu and Wobbrock [38]: blue - very high agreement, green - high agreement, yellow - medium agreement, red - low agreement.

Application	Referent	AR(on-hand)	AR(off-hand)	Consensus gesture	%(on-hand)	%(off-hand)
Map	Pan (left/right/up/down)	0.371	0.615	Swiping	60%	79%
	Zooming(in/out)	0.343	0.725	Pinch	60%	86%
	Rotating(clockwise/anti-clockwise)	0.257	0.253	Hand flick	46%	50%
Browser	Scrolling (up/down)	0.867	0.725	Swiping	93%	86%
	Selecting/clicking elements	0.867	0.418	Tapping	93%	64%
	Navigate pages (prev./next)	0.210	0.209	Swiping	33%	36%
				Tapping	33%	
Media player	Play/pause/stop	0.352	0.352	Tapping	53%	50%
	fast-forward/rewind	0.066	0.077	Tapping	20%	21%
	Navigate media (prev./next)	0.229	0.242	Tapping	60%	43%
	Volume control (increase/decrease)	0.286	0.231	Swipe along fingers	53%	43%
Home screen (General application)	Going to home screen	0.295	0.244	Full hand gesture	40%	36%
	Navigating a grid of elements	0.371	0.253	Tapping	60%	50%
	Selecting from grid of elements	0.752	0.857	Tapping	87%	93%
App switching screen (General application)	Going to the app switching screen	0.257	0.242	Full hand gesture	47%	43%
	Navigating a horizontal list of elements	0.190	0.242	Swiping	40%	50%
	Selecting from horizontal list of elements	0.552	0.615	Tapping	73%	79%

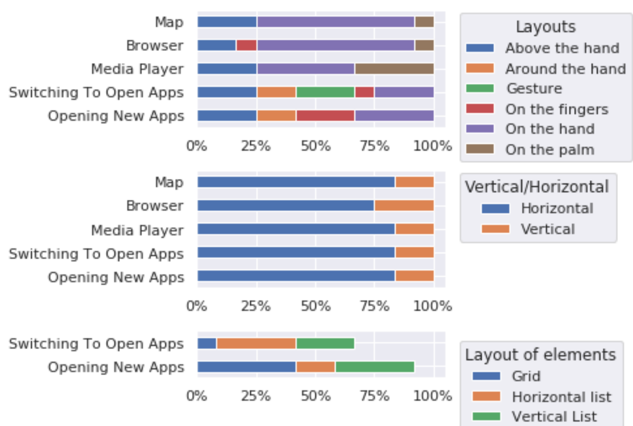


Figure 2: Breakdown of the proposed layouts along the three dimensions for each application. The y-axis presents the different applications considered. The x-axis presents the percentage breakdown for each category along a dimension.

3.2 Study 2: Gesture elicitation study

3.2.1 Stimuli for the study. Two visual interfaces were selected from the first study as the stimuli for Study 2 (the top two proposed layouts: on the hand; 46%, and above the hand; 23%). Both layouts required the dominant hand being held out horizontally in front of the user with the palm normal to the line of sight. For the layout of elements, the majority for each application was selected (grid layout for the home screen; 41%, horizontal list layout for app switching screen; 33%). The first visual interface was a horizontally oriented rectangle with the interface superimposed on the palm and fingers, referred to as "on-hand" approach (see left column of Figure 3). Akin to this, the second visual interface was floating up above the palm and fingers, occupying the same space as the thumb when the hand was in a relaxed position. This was referred to as "off-hand" approach (see right column of Figure 3). The stimuli selected were based on current mobile and personal computing platforms; the familiarity was expected to allow participants to imagine the interactions and provide more coherent proposals. Adobe After Effects was used to mock up the five different application demos as video clips for each interface position.

3.2.2 Participants. Participants were recruited from a local university for the study (N = 15; F = 5, M = 10, their age ranged between 18 and 27; M = 24.27, SD = 3.24). Participants used their smartphone between 1 and 15 hours per day (M = 4.93, SD = 3.47). Only one participant was left-handed. Ishihara colour blindness tests indicated that none of our participants had red-green colour blindness.

3.2.3 Procedure. First, the participants were greeted and asked to turn off any smart devices to avoid distractions. The structure of the study was explained to them. We followed the explanation by going over and signing the consent form. A questionnaire with general demographic information and mobile device usage habits was completed as well. Subsequently, the participants were shown the first video clip (based on the their condition, either on-hand or off-hand). After observing the visual interface, they were asked to describe how they would perform the referents described in Table 1 using only the hand displaying the anchored interface. The participants were allowed to go through the video clip as many times as necessary for them to come up with a gesture. They were

Figure 3: Images captured from the stimuli used in Study 2. The left column shows the interfaces displayed on the hand (on-hand). The right column shows the interfaces displayed above the hand (off-hand). The rows correspond to the different applications considered: 1) Map 2) Browser 3) Media player 4) Home screen (grid layout) 5) App switching screen (horizontal list layout).

asked to treat the technology as a "magic box"[1], that is, assume the technology can detect any gesture produced.

Following the gesture proposals for each referent/application, the participants were asked if they would consider making any alterations to the proposed gestures if they were to be used in public. Once the participants had proposed a full set of gestures for a given condition (on-hand or off-hand), they were then asked to fill out a questionnaire regarding the interfaces of that condition, with feedback on how cool, comfortable, and convenient using the technology would be, as well as how likely they were to use the technology or recommend it to someone. The same process was repeated for the second video clip (based on the their condition, either on-hand or off-hand). Finally, Ishihara colorblindness test was administered[2]: Due to the visual nature of the stimuli, we considered color blindness as a potential factor affecting our results. Thus, we used Ishihara Color Blindness Test to control for color blindness as a factor.

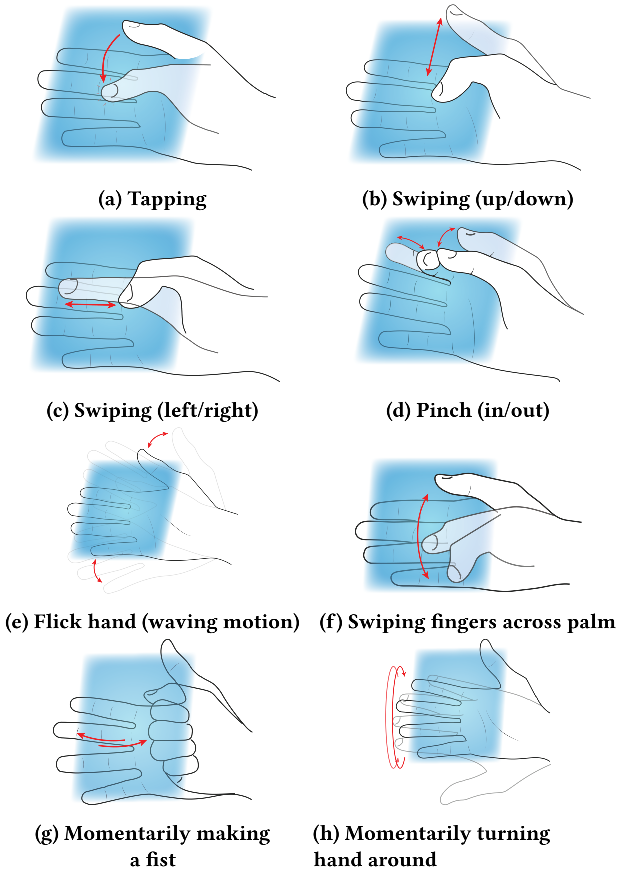
3.2.4 Results. The taxonomy used to classify the gestures is given in Table 2, and is based on the taxonomy proposed by Wobbrock et al. [43]. The *nature* dimension used in the previous studies [7, 28, 37, 43] is not considered here, as it does not directly relate to single-handed gestures. The *form* dimension describes the interaction method. *Single-finger static* gestures are gestures that involve only a single finger and do not involve a dynamic movement of the finger (e.g., tap, double tap). *Single-finger dynamic* gestures involve a single finger in a dynamic motion (e.g., swipe). *Multiple fingers* would use two or more of the fingers for a gesture, such as finger flexion of multiple fingers, whereas *hand* gestures would use the whole hand (e.g., making a fist, waving hand). The definition of the *flow* dimension follows the definition of Wobbrock et al. [43], a gesture is considered *discrete* if the event corresponding to a gesture occurs after the gesture has been completed (e.g., tapping/clicking); a gesture is considered *continuous* if the corresponding event is ongoing as the gesture is being performed (e.g., swipe or press and hold to increase volume). The *binding* dimension describes whether the gesture involves interacting with a visual element on the interface (e.g., click on visible button) or not (e.g., swipe to change screens). The data was coded by two coders (Cohen's *kappa* = .79) and the conflicts between the coders were resolved through discussion.

Table 2: Taxonomy used for the classification of gestures; based on Wobbrock et al. [43]

Dimension	Categories	Descriptions
Form	Single-finger static	Uses only one finger and doesn't involve a dynamic movement of the finger, such as a swipe.
	Single-finger dynamic	Uses only one finger and involves a dynamic movement of the finger, such as a swipe.
	Multiple fingers	Uses multiple fingers for a gesture.
	Hand	Uses the whole hand for a gesture.
Flow	Discrete	Response occurs after the user acts.
	Continuous	Response occurs while the user acts.
Binding	Visual element dependent	Gesture interacts with a visual element.
	Visual element independent	Gesture does not interact with a visual element.

A total of 462 gestures were elicited in our study. The *agreement rate* (or AR) and the classification based on AR, proposed by Vatavu and Wobbrock [38] was used for the analysis. Table 1 shows the AR for the referents for both on and off-hand. 4 referents for on-hand and 5 referents for off-hand had very high agreement ($AR > 0.5$; highlighted in red in table 1), 4 referents for on-hand and 2 referents for off-hand had high agreement ($0.5 \geq AR > 0.3$; highlighted in yellow in table 1) and 7 referents from on-hand and 8 referents from off-hand had medium agreement ($0.3 \geq AR > 0.1$; highlighted in green in table 1). *Fast-forward/rewind* had a low agreement rate for both approaches ($0.1 \leq AR$; highlighted in blue in table 1). The biggest difference between the gestures for on-hand and off-hand are seen for zooming (*difference* = 0.38) and selection (*difference* = 0.45). For zooming, a large majority of participants selected "pinch" gesture when the interface was displayed above the hand (off-hand - 86%). In comparison to off-hand, gestures proposed for on-hand had more variability, although the majority of people still proposed

[1]A generalized phrase for the phrase ("magic brick") used by Ruiz et al.[31]
[2]https://www.color-blindness.com/ishihara-38-plates-cvd-test/



(a) Tapping (b) Swiping (up/down) (c) Swiping (left/right) (d) Pinch (in/out) (e) Flick hand (waving motion) (f) Swiping fingers across palm (g) Momentarily making a fist (h) Momentarily turning hand around

Figure 4: The consensus gesture set based on Study 2.

the same gesture (on-hand - 60%). For selection, the participants preferred directly tapping the elements when the interface was on the hand (on-hand - 93%). The same gesture was preferred for off-hand as well, but only at 64%.

Based on previous work [24, 28, 43] we derive a consensus gesture set based on Study 2. All referents had a similar consensus of gestures for both on-hand and off-hand (Table 1) except for the *navigating pages* referent, where for the on-hand, tapping was also part of the consensus gesture set for the referent. The derived set of gestures are presented in Figure 4 and the referents to which they are mapped are listed in Table 1. The hand gestures were put into two categories: making a fist, and flicking the complete hand (along any axis). Since the consensus gesture set for both on-hand and off-hand cases were similar, for brevity, Figure 4 outlines the consensus gestures with the on-hand condition.

Comfort, Convenience, Coolness. The participants' subjective experiences were explored with three questions in the second questionnaire. Based on the results of Kolomogorov-Smirnov tests, non-parametric alternatives (Wilcoxon Signed Rank tests) were conducted for this part of the analyses and explored differences between medians. Participants perceived both on and off-hand as equally cool ($Z = -.98$, $p = .33$, $r = .28$), convenient ($Z = -.86$, $p = .39$, $r = .25$), and comfortable ($Z = -.16$, $p = .87$, $r = .05$). Next, participants' interest in using the technology in the future was explored, but there was no significant effect across the approaches (i.e., on-hand vs. off-hand, $Z = -.36$, $p = .72$, $r = .10$). Note that we have to interpret the results with caution since our sample size was small.

Interest in Future Use. Finally, multiple regression analyses were conducted to explore whether the levels of comfort, convenience, and coolness could predict the participants' interest in using the technology in the future. Regarding the on-hand approach, the results indicated that the model accounted for approximately 68 percent of variance ($R^2_{adj} = .68$) in predicting the level of interest in using the technology in the future at a significant level; $F(2, 11) = 10.81$, $p = .001$. Note that due to smaller sample size, adjusted R squares are reported here. Among the predictor variables, perceived convenience was the only significant contributor to the model ($Beta = .80$, $p = .002$) while comfort ($Beta = -.03$, $p = .86$) and coolness ($Beta = .21$, $p = .43$) were not. A similar pattern emerged for the off-hand approach. Once again, the model predicted the participants' interest in using the off-hand approach in the future at significant level: The model accounted for approximately 68 percent of the variance ($R^2_{adj} = .68$) in the participants' interest level; $F(3, 11) = 11.06$, $p = .001$. Again, while perceived convenience was a significant contributor to the model ($Beta = .66$, $p = .049$), comfort ($Beta = .19$, $p = .53$) and coolness ($Beta = .19$, $p = .15$) were not. Thus, perceived usefulness predicted the participants' interest in using the technology.

4 DISCUSSION

Legacy Bias: Using the Smartphone Interaction Model. As anticipated, some potential influence of legacy bias was recognized especially when participants were generating gestures for many of the referents - they commented on the interface being similar to a smartphone, allowing them to use similar gestures for the hand proximate UI: This legacy bias could be one of the contributing factors for higher agreement rates seen in Table 1. A similar sentiment was reported with regards to social acceptability. Only two participants suggested changes (from using the hand itself to using gestures with fingers only, for the map application) when inquired about using the gestures in public. Other participants were comfortable with using the originally generated gestures in public, some emphasizing that it is very similar to using a smartphone. Some participants also commented on how it was not possible to transfer certain gestures directly from the smartphone for some referents (e.g., zooming) due to the single-handed requirement: The fact that they still were able to propose gestures for these referents implies that this interaction paradigm has a larger functional space to be explored, in comparison to the space provided by smartphones.

Interface Layout Design. As discussed in Section 3.2.4, the participants did not have a strong preference for the position of the interface, and proposed gestures were similar across both layout approaches. However, based on participants' comments, there are a few points to be considered. First, participants preferred to directly interact with the element using their thumb when placing elements that can be clicked/tapped/selected. Observations made during Study 2 corresponds to the functional workspace of the thumb discussed by Huang et al. [19] and Dewitz et al. [8]. Interestingly the preference for direct input was observed even when the interface was displayed off the hand (Figure 3, right column), though the participants commented on the reach of the thumb being limited. For elements beyond their reach, they preferred using scrolling/zooming to bring an element within the reach of the thumb to use direct input. In addition to the thumb to finger interaction space, it is worth further exploring the complete functional workspace of the thumb when virtual interactions are also taken into consideration. Another parallel question is about the importance of tactile interactions when the interface is overlaid on the hand. When proposing interaction methods with displayed virtual interfaces, most of the participants were rather focused on directly interacting with the virtual element, and tactile feedback was often not considered. For example, some commented that not being able to see the fingers through the display discouraged them from interacting with those fingers such as with on-hand (Figure 3, left column). This can be seen with the *zooming* referent. When the interface was above the hand, many proposed the pinch gesture, but when it was on the hand, the preference for pinch dropped and preferred gestures were similar to contracting/expanding the fingers or double/triple tap. Some of those who did recommend the pinch in/out gestures for the on-hand approach attempted to treat the virtual interface as a surface on which to perform the gesture.

During both studies, some participants preferred the vertical alignment of the hand as opposed to holding the hand horizontally. The participants who preferred this alignment described it being more comfortable and familiar, and the aspect ratio of the interface would be closer to that of the smartphone. One participant commented that it might be more comfortable if the interface was not horizontally or vertically aligned, but rather aligned in an intermediate position. This can be further extended to consider if users would prefer the interface being oriented and positioned relative to the head as opposed to the interface being positioned relative to the hand, and the orientation be relative to the point-of-view. While only the top two ranking layouts from the first study were chosen to be used in the second study, in future studies it would be beneficial to explore the other proposed layouts as well.

Gestures for interacting with interfaces displayed on and around the hand. Based on the consensus gestures (Table 1, Figure 4), tap (or touch) and swipe gestures, primarily using the thumb, were the most common. For applications that mostly require binary input (such as the media player) participants prefer assigning meaning to different fingers or finger segments and interacting with them. Alternatively, for universal applications (home screen, app switching screen) participants preferred having unique gestures, where most proposed using the complete hand, such as making a fist or flipping the whole hand once. Three participants preferred having an icon on the interface to interact with to bring up these applications, citing social acceptability as a cause for not using whole hand gestures. Following the observations of the interface layout design, participants preferred to use the thumb to directly interact with elements that are displayed on the interface. When the interface design requires laying out a grid of elements or a list to interact with, this can be taken into consideration. The observations made in these studies identify the preliminary design space for hand proximate UIs; with implementations of these interfaces more studies need to be conducted to gain better insight. For example, how the reach of the thumb will impact the use of the off-hand approach cannot be fully understood based on the conducted studies. Future studies with a larger sample size and larger age range will be beneficial.

5 CONCLUSION

In this work we introduced an alternative interaction paradigm for mixed reality settings. We contributed to the literature on hand proximate UIs by exploring elements displayed on and around the hand, and interacting with them using the hand as an anchor. We explored this interaction space by conducting two studies. Study 1 was conducted with participants with a background in Human-Computer Interaction to inform the different possible layouts for a hand-based interface. Study 2, a user-elicitation study, was conducted where the participants generated gestures to be used for interacting with the selected interface layouts proposed during Study 1. Based on the results of these studies, a set of UI layouts and consensus gestures are presented. The results show that this interaction paradigm, while providing a familiar, socially acceptable interaction modality for mixed reality applications, also extends the smartphone interaction space. Our findings can serve as guidelines for designing interfaces in this interaction paradigm while also informing future research directions.

ACKNOWLEDGMENTS

We acknowledge financial support from an NSERC Strategic Project Grant to Menon, and CRC funding to Irani.

REFERENCES

[1] Ammar Ahmad, Cyrille Migniot, and Albert Dipanda. 2019. Hand Pose Estimation and Tracking in Real and Virtual Interaction:a Review. *Image and Vision Computing* 89, nil (2019), 35–49. https://doi.org/10.1016/j.imavis.2019.06.003

[2] Jason Alexander, Teng Han, William Judd, Pourang Irani, and Sriram Subramanian. 2012. Putting Your Best Foot Forward: Investigating Real-World Mappings for Foot-Based Gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1229–1238. https://doi.org/10.1145/2207676.2208575

[3] Ilyasse Belkacem, Isabelle Pecci, and Benoît Martin. 2019. Pointing task on smart glasses: Comparison of four interaction techniques. *CoRR* abs/1905.05810 (2019). arXiv:1905.05810 http://arxiv.org/abs/1905.05810

[4] Anabela Berenguer, Jorge Goncalves, Simo Hosio, Denzil Ferreira, Theodoros Anagnostopoulos, and Vassilis Kostakos. 2017. Are Smartphones Ubiquitous?: an In-Depth Survey of Smartphone Adoption By Seniors. *IEEE Consumer Electronics Magazine* 6, 1 (2017), 104–110. https://doi.org/10.1109/mce.2016.2614524

[5] Idil Bostan, Oundefineduz Turan Buruk, Mert Canat, Mustafa Ozan Tezcan, Celalettin Yurdakul, Tilbe Göksun, and Oundefineduzhan Özcan. 2017. Hands as a Controller: User Preferences for Hand Specific On-Skin Gestures. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*. ACM, New York, NY, USA, 1123–1134. https://doi.org/10.1145/3064663.3064766

[6] Edwin Chan, Teddy Seyed, Wolfgang Stuerzlinger, Xing-Dong Yang, and Frank Maurer. 2016. User Elicitation on Single-Hand Microgestures. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 3403–3414. https://doi.org/10.1145/2858036.2858589

[7] Zhen Chen, Xiaochi Ma, Zeya Peng, Ying Zhou, Mengge Yao, Zheng Ma, Ci Wang, Zaifeng Gao, and Mowei Shen. 2017. User-Defined Gestures for Gestural Interaction: Extending From Hands To Other Body Parts. *International Journal of Human-Computer Interaction* 34, 3 (2017), 238–250. https://doi.org/10.1080/10447318.2017.1342943

[8] Bastian Dewitz, Frank Steinicke, and Christian Geiger. 2019. Functional Workspace for One-Handed Tap and Swipe Microgestures. In *Mensch und Computer 2019 - Workshopband*. Gesellschaft für Informatik e.V., Bonn. https://doi.org/10.18420/muc2019-ws-440

[9] Niloofar Dezfuli, Mohammadreza Khalilbeigi, Jochen Huber, Florian Müller, and Max Mühlhäuser. 2012. PalmRC: Imaginary Palm-Based Remote Control for Eyes-Free Television Interaction. In *Proceedings of the 10th European Conference*

on Interactive TV and Video (EuroITV '12). ACM, New York, NY, USA, 27–34. https://doi.org/10.1145/2325616.2325623

[10] Haiwei Dong, Ali Danesh, Nadia Figueroa, and Abdulmotaleb El Saddik. 2015. An Elicitation Study on Gesture Preferences and Memorability Toward a Practical Hand-Gesture Vocabulary for Smart Televisions. *IEEE Access* 3, nil (2015), 543–555. https://doi.org/10.1109/access.2015.2432679

[11] Niklas Elmqvist and Pourang Irani. 2013. Ubiquitous Analytics: Interacting With Big Data Anywhere, Anytime. *Computer* 46, 4 (2013), 86–89. https://doi.org/10.1109/mc.2013.147

[12] Barrett Ens, Ahmad Byagowi, Teng Han, Juan David Hincapié-Ramos, and Pourang Irani. 2016. Combining Ring Input with Hand Tracking for Precise, Natural Interaction with Spatial Analytic Interfaces. In *Proceedings of the 2016 Symposium on Spatial User Interaction (SUI '16)*. ACM, New York, NY, USA, 99–102. https://doi.org/10.1145/2983310.2985757

[13] Jéssica Franco and Diogo Cabral. 2019. Augmented Object Selection through Smart Glasses. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia (MUM '19)*. ACM, New York, NY, USA, Article Article 47, 5 pages. https://doi.org/10.1145/3365610.3368416

[14] Sean G. Gustafson, Bernhard Rabe, and Patrick M. Baudisch. 2013. Understanding Palm-Based Imaginary Interfaces: The Role of Visual and Tactile Cues When Browsing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 889–898. https://doi.org/10.1145/2470654.2466114

[15] Taejin Ha, Steven Feiner, and Woontack Woo. 2014. WeARHand: Head-worn, RGB-D camera-based, bare-hand user interface with visually enhanced depth perception. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 219–228. https://doi.org/10.1109/ISMAR.2014.6948431

[16] Chris Harrison, Hrvoje Benko, and Andrew D. Wilson. 2011. OmniTouch: Wearable Multitouch Interaction Everywhere. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, New York, NY, USA, 441–450. https://doi.org/10.1145/2047196.2047255

[17] Chris Harrison and Haakon Faste. 2014. Implications of location and touch for on-body projected interfaces. In *Proceedings of the 2014 conference on Designing interactive systems - DIS '14*. nil. https://doi.org/10.1145/2598510.2598587

[18] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-Air Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1063–1072. https://doi.org/10.1145/2556288.2557130

[19] Da-Yuan Huang, Liwei Chan, Shuo Yang, Fan Wang, Rong-Hao Liang, De-Nian Yang, Yi-Ping Hung, and Bing-Yu Chen. 2016. DigitSpace: Designing Thumb-to-Fingers Touch Interfaces for One-Handed and Eyes-Free Interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1526–1537. https://doi.org/10.1145/2858036.2858483

[20] Zhanpeng Huang, Weikai Li, and Pan Hui. 2015. Ubii: Towards Seamless Interaction between Digital and Physical Worlds. In *Proceedings of the 23rd ACM International Conference on Multimedia (MM '15)*. ACM, New York, NY, USA, 341–350. https://doi.org/10.1145/2733373.2806266

[21] Amy K. Karlson, Benjamin B. Bederson, and Jose L. Contreras-Vidal. 2006. *Understanding Single-Handed Mobile Device Interaction.* Technical Report. University of Maryland. https://www.cs.umd.edu/hcil/trs/2006-02/2006-02.pdf

[22] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article Paper 81, 14 pages. https://doi.org/10.1145/3173574.3173655

[23] Lik-Hang Lee and Pan Hui. 2018. Interaction Methods for Smart Glasses: a Survey. *IEEE Access* 6, nil (2018), 28712–28732. https://doi.org/10.1109/access.2018.2831081

[24] Nathan Magrofuoco, Jorge-Luis Perez-Medina, Paolo Roselli, Jean Vanderdonckt, and Santiago Villarreal. 2019. Eliciting Contact-Based and Contactless Gestures With Radar-Based Sensors. *IEEE Access* 7, nil (2019), 176982–176997. https://doi.org/10.1109/access.2019.2951349

[25] Jameel Malik, Ahmed Elhayek, and Didier Stricker. 2019. Whsp-Net: a Weakly-Supervised Approach for 3d Hand Shape and Pose Recovery From a Single Depth Image. *Sensors* 19, 17 (2019), 3784. https://doi.org/10.3390/s19173784

[26] Meredith Ringel Morris, Jacob O. Wobbrock, and Andy Wilson. 2010. Understanding Users' Preferences for Surface Gestures. In *GI '10 Proceedings of Graphics Interface 2010*. Canadian Information Processing Society, 261–268. https://www.microsoft.com/en-us/research/publication/understanding-users-preferences-surface-gestures/

[27] Miguel A. Nacenta, Yemliha Kamber, Yizhou Qiang, and Per Ola Kristensson. 2013. Memorability of Pre-Designed and User-Defined Gesture Sets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1099–1108. https://doi.org/10.1145/2470654.2466142

[28] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-Defined Gestures for Augmented Reality. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. ACM, New York, NY, USA, 955–960. https://doi.org/10.1145/2468356.2468527

[29] Hyocheol Ro, Jung-Hyun Byun, Yoon Jung Park, Nam Kyu Lee, and Tack-Don Han. 2019. Ar Pointer: Advanced Ray-Casting Interface Using Laser Pointer Metaphor for Object Manipulation in 3d Augmented Reality Environment. *Applied Sciences* 9, 15 (2019), 3078. https://doi.org/10.3390/app9153078

[30] Gustavo Alberto Rovelo Ruiz, Davy Vanacken, Kris Luyten, Francisco Abad, and Emilio Camahort. 2014. Multi-Viewer Gesture-Based Interaction for Omni-Directional Video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 4077–4086. https://doi.org/10.1145/2556288.2557113

[31] Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-Defined Motion Gestures for Mobile Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 197–206. https://doi.org/10.1145/1978942.1978971

[32] Lawrence Sambrooks and Brett Wilkinson. 2013. Comparison of Gestural, Touch, and Mouse Interaction with Fitts' Law. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration (OzCHI '13)*. ACM, New York, NY, USA, 119–122. https://doi.org/10.1145/2541016.2541066

[33] Shaikh Shawon Arefin Shimon, Sarah Morrison-Smith, Noah John, Ghazal Fahimi, and Jaime Ruiz. 2015. Exploring User-Defined Back-Of-Device Gestures for Mobile Devices. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15)*. ACM, New York, NY, USA, 227–232. https://doi.org/10.1145/2785830.2785890

[34] Mohamed Soliman, Franziska Mueller, Lena Hegemann, Joan Sol Roo, Christian Theobalt, and Jürgen Steimle. 2018. FingerInput: Capturing Expressive Single-Hand Thumb-to-Finger Microgestures. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces (ISS '18)*. ACM, New York, NY, USA, 177–187. https://doi.org/10.1145/3279778.3279799

[35] Bugra Tekin, Federica Bogo, and Marc Pollefeys. 2019. H+O: Unified Egocentric Recognition of 3D Hand-Object Poses and Interactions. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 4511–4520.

[36] Hsin-Ruey Tsai, Te-Yen Wu, Da-Yuan Huang, Min-Chieh Hsiu, Jui-Chun Hsiao, Yi-Ping Hung, Mike Y. Chen, and Bing-Yu Chen. 2017. SegTouch: Enhancing Touch Input While Providing Touch Gestures on Screens Using Thumb-To-Index-Finger Gestures. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 2164–2171. https://doi.org/10.1145/3027063.3053109

[37] Ying-Chao Tung, Chun-Yen Hsu, Han-Yu Wang, Silvia Chyou, Jhe-Wei Lin, Pei-Jung Wu, Andries Valstar, and Mike Y. Chen. 2015. User-Defined Game Input for Smart Glasses in Public Space. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3327–3336. https://doi.org/10.1145/2702123.2702214

[38] Radu-Daniel Vatavu and Jacob O. Wobbrock. 2015. Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1325–1334. https://doi.org/10.1145/2702123.2702223

[39] Cheng-Yao Wang, Wei-Chen Chu, Po-Tsung Chiu, Min-Chieh Hsiu, Yih-Harn Chiang, and Mike Y. Chen. 2015. PalmType: Using Palms as Keyboards for Smart Glasses. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15)*. ACM, New York, NY, USA, 153–160. https://doi.org/10.1145/2785830.2785886

[40] Cheng-Yao Wang, Min-Chieh Hsiu, Po-Tsung Chiu, Chiao-Hui Chang, Liwei Chan, Bing-Yu Chen, and Mike Y. Chen. 2015. PalmGesture: Using Palms as Gesture Interfaces for Eyes-Free Input. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15)*. ACM, New York, NY, USA, 217–226. https://doi.org/10.1145/2785830.2785885

[41] Martin Weigel, Vikram Mehta, and Jürgen Steimle. 2014. More than Touch: Understanding How People Use Skin as an Input Surface for Mobile Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 179–188. https://doi.org/10.1145/2556288.2557239

[42] Eric Whitmire, Mohit Jain, Divye Jain, Greg Nelson, Ravi Karkar, Shwetak Patel, and Mayank Goel. 2017. Digitouch. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–21. https://doi.org/10.1145/3130978

[43] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-Defined Gestures for Surface Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1083–1092. https://doi.org/10.1145/1518701.1518866

[44] Zheer Xu, Pui Chung Wong, Jun Gong, Te-Yen Wu, Aditya Shekhar Nittala, Xiaojun Bi, Jürgen Steimle, Hongbo Fu, Kening Zhu, and Xing-Dong Yang. 2019. TipText: Eyes-Free Text Entry on a Fingertip Keyboard. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. ACM, New York, NY, USA, 883–899. https://doi.org/10.1145/3332165.3347865