GestureCalc: An Eyes-Free Calculator for Touch Screens

Bindita Chaudhuri[1]*, Leah Perlmutter[1]*, Justin Petelka[2], Philip Garrison[1]
James Fogarty[1], Jacob O. Wobbrock[2], Richard E. Ladner[1]
[1]Paul G. Allen School of Computer Science and Engineering, [2]The Information School
University of Washington, Seattle, WA 98195, U.S.A.
[1]{bindita, lrperlmu, philipmg, jfogarty, ladner}@cs.washington.edu, [2]{jpetelka, wobbrock}@uw.edu

ABSTRACT
A digital calculator is one of the most frequently used touch screen applications. However, keypad-based character input in existing calculator applications requires precise, targeted key presses, which is time consuming and error-prone for screen readers users. We introduce *GestureCalc*, a digital calculator that uses target-free gestures for arithmetic tasks. It allows eyes-free input in the form of taps and directional swipes with one to three fingers, guided by minimal audio feedback, to enter digits and operations. We conducted a mixed methods longitudinal study with eight screen reader users and found that they entered characters with *GestureCalc* 40.5% faster on average than with a typical touch screen calculator. Participants made more mistakes but also corrected more errors with *GestureCalc*, resulting in 52.2% fewer erroneous calculations than the baseline. Over the three sessions in the study participants were able to learn the *GestureCalc*gestures well to efficiently perform short calculations. From our interviews after the second session, participants recognized the effort in learning a new gesture set, yet reported confidence in their ability to become fluent in practice.

Author Keywords
Eyes-free entry; gesture inputs; digital calculator; touch screen; mobile devices.

INTRODUCTION
The digital calculator is a very common application that people use on touch screen devices. For most people this application is easy to use; they visually locate targets in the form of soft buttons and tap them to get a desired result. Buttons stand for digits (0-9), the decimal point and operations (e.g. subtraction, multiplication, backspace, equals etc.). For people who use screen readers (VoiceOver for iOS or TalkBack for Android), finding a button using one-finger exploration and then activating it can be time consuming. The purpose of this paper is to explore a design for a calculator that uses target-less gestures only, eliminating the need to explore for a button when using a screen reader. We call this *GestureCalc*, an eyes-free gesture-based calculator for touch screens. Eyes-free interaction have been explored by several researchers [5, 30, 34, 18, 4, 12]. The authors in [24] have also uncovered the differences in the types of gestures sighted and blind people find intuitive. Researchers have also proposed specific improvements to the accessibility of digital calculators. These eyes-free solutions include gesture-based numeric input [36], minimal audio feedback for numeric input [3], and Braille-based gesture calculators [1]. Our goal is to combine the advantages and address the disadvantages of existing solutions to create a single gesture-only calculator application usable for screen reader users, typically visually impaired users.

To create *GestureCalc* we modified DigiTaps[1.8] [3] for digits and defined new metaphor-based gestures for basic calculator operations. The overall goal is to improve the accessibility of state-of-the-art digital calculator applications by a) designing gesture codes based on conceptual metaphors and b) requiring one or two simple gestures for each digit or operation. We evaluated our prototype in a longitudinal study with eight

(a) *GestureCalc* (b) Baseline

Figure 1: (a) Performing a three-finger tap on *GestureCalc* and (b) Typing "5" on *ClassicCalc*, the baseline. Using our novel eyes-free app with target-free rich gestures, participants entered calculations 40.5% faster and performed 52.2% fewer erroneous calculations than with the baseline.

*The first two authors contributed equally to this work.

ASSETS '19, October 28–30, 2019, Pittsburg, PA, USA
ACM xxx-x-xxxx-xxxx-x/xx/xx. . . $15.00
DOI: http://dx.doi.org/xx.xxxx/xxxxxxx.xxxxxxx

participants over three sessions. During each session we evaluated: 1) the rate of error corrections in the input, 2) the speed of entering the gestures, and 3) the memorability and intuitiveness of the gestures.

Our main contributions are:

- We designed a novel eyes-free target-less digital calculator application that uses a minimal number of accessible gestures to enter digits and operations. Our code is available online[1] and we also plan to release the app on the App store.
- We proposed intuitive gestures based on conceptual metaphors that are both memorable and easily learnable for visually impaired people.
- We conducted a mixed methods study where participants perform real mathematical expressions in order to evaluate how our application might be adopted.

RELATED WORK

Gestures and Metaphors

Modern understanding of metaphors go beyond substitution and comparison in language [42, 16]. In cognitive linguistics, Lakoff and Johnson's theory of conceptual metaphors [25] states that we understand new knowledge by "mapping from known domains to unknown domains" through recurring structures within our cognitive processes (image schemas) or deeper conceptualizations based in our physical, bodily experience. Though these conceptualizations are influenced by our sociocultural environment and unique physical experiences, many schemas are grounded in experiences that transcend culture, such as forward motion, occlusion, containment, sensation, and perception. [8, 28].

Metaphors have recently gained popularity in the HCI community. Hurtienne and Blessing [22] found schema-aligned interface elements to be more intuitive to users than interface elements that reversed the established schema. Loeffler et al. isolated schemas in user interview transcripts, and mapped these to their designed interface [27]. Hiniker et al. found that users were more efficient in working with metaphorical visualizations that aligned with documented image schemas [20]. Finally, Kane et al. found symbol-based gesture languages to be less intuitive than metaphor-based languages for blind people [24]. Given the potential of metaphors to improve usability for blind people, we designed *GestureCalc* to use metaphoric input instead of symbolic input. Gestures with directional swipes are grouped according to whether they increase (e.g. multiplication/addition) or decrease (e.g. division/subtraction) the value and oriented according to the metaphor "More is up" (which is why addition is an upward swipe, etc.).

Eyes-free Entry

Several researchers have developed eyes-free text entry systems for touch screen devices [5, 30, 34, 4, 12, 41]. Touch screen operating systems generally have default screen readers, such as Apple's VoiceOver for iOS [2] or Android's TalkBack [15], which use interaction techniques introduced in SlideRule [23]. Screen reader based text entry guides people through audio feedback to search for targets with one or more fingers on the screen and then perform a second gesture such as a double-tap to select that character to input, making the process cumbersome. Bonner et al. developed No-Look-Notes [9], a soft keyboard dividing character input into a first split-tap to select from 8 segments of a pie menu and a second split-tap to select a single character from that segment. This makes character selection easier and faster than with a QWERTY soft keyboard with screen reader, where buttons are very small. However, this still requires searching and targeting, which results in slow character input rates. Speech input is a possible alternative for eyes-free entry, but it has several limitations: (a) it cannot be used in quiet environments, (b) the input is prone to error especially in noisy environments, and (c) it is not a feasible solution for speech-impaired users.

Eyes-free entry techniques using Braille have also been used to improve touch screen accessibility for blind people [34, 30, 6]. Alnfiai et al. proposed BrailleTap [1], a calculator application that uses taps in the form of Braille patterns for numeric input and other gestures for calculator operations. However, Braille-based inputs have several drawbacks. First, these techniques require multiple gestures per character input (up to six), which leads to a low input speed. Second, Braille-based inputs require knowledge of Braille dot patterns. While Braille was developed for people with visual impairments, a report from the National Federation of the Blind states that only 10% of legally blind people can read Braille [33]. This makes Braille legible to only a small percentage of blind people, but we wanted our app to be accessible to a wider population.

Eyes-free numeric input methods to enter digits (like *Tap2Count* [18] and *Digitaps* [3]) and operations (like *Tapulator* [36]) also exist in the literature. *Tap2Count* allows users to touch an interactive screen with one to ten fingers to enter digits 0-9. This requires considerable physical effort and cannot be easily scaled to small touch screen devices like mobile phones. *DigiTaps* uses an eyes-free gesture language based on a prefix-free coding scheme for numeric input with haptic or audio feedback on touch screen devices. However, both *Tap2Count* and *DigiTaps* do not include operations, which increase the challenges of designing usable gestures and implementing a working system. *Tapulator* extends *Digitaps* by adding gestures for operations, but the gestures are symbolic, based on the printed structures of the operators, rather than metaphorical, which should hinder learnability and memorability for visually impaired users.

Existing Products

The idea of using gestures for calculations dates back to when a touch screen calculator was implemented on a digital watch (Casio AT-550) [11]. Recently, MyScript created a calculator [32] that takes handwritten characters on a touch screen as input for mathematical calculations, but it is not usable by blind users. A handful of gesture based calculators are also available online, for example, Sumzy for iPhone [7], Swipe Calculator for Android [31] and Rechner Calculator [35]. All these calculators have a 3×3 keypad for the digits, and only the operations are performed using tap and swipe gestures on or above the keypad area on the screen. To the best of our

[1] https://github.com/bindita/GestureCalculator

knowledge, our calculator is the first application which is free of any buttons or targeted gestures.

DESIGN

We build on *DigiTaps*[1.8] to include metaphor-based gestures for calculator operations to ensure intuitiveness of our gesture language. Our application also improves character entry speed compared to a classic calculator application by (a) allowing users to interact with any part of the screen (i.e. target-less interaction), (b) requiring a maximum of two gestures for every input, and (c) avoiding complex gestures by using only swipes and taps with up to three fingers.

Gestures

Common gestures for interacting with touch screen devices include tap, swipe, pinch, shake, rotate, etc. In our design, we only use taps and swipes since they have been found to be easiest to perform and accessible to blind users [24]. We also use a variation of the tap gesture called long tap, where users press and hold a finger against the touch screen for a short duration. Our swipe gestures are directional: up, down, left, and right. All swipes, taps, and long taps can be performed with one, two, or three fingers simultaneously. We restrict each gesture to involve a maximum of three fingers since interaction with the 'pinky' finger is difficult [36]. Figure 2 shows the symbolic representations of the different gestures as possible inputs in our design. The gestures can be performed anywhere on the screen.

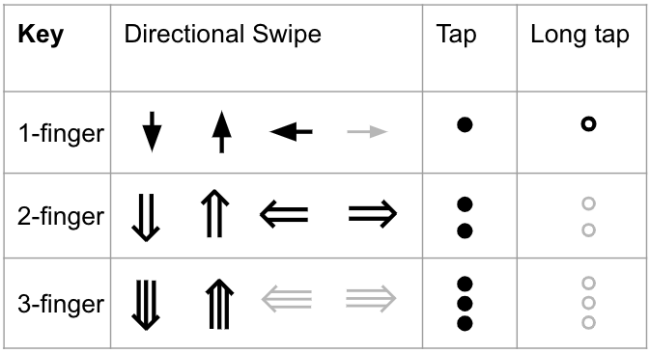

Figure 2: Symbolic representations of a superset of our gestures, with a visual shortcut for each gesture. Those used in *GestureCalc* are marked in black and the unused ones are marked in grey.

Character Codes

We define the term 'characters' as all the digits and operations that *GestureCalc* accepts as input. Each character is encoded by a combination of one or two gestures. Our character codes are prefix-free, meaning that no character's code is the prefix of another code. This property is important for unambiguous parsing of the input. In addition, our design aims to balance two properties: a) the codes should be semantically meaningful so they are easy to remember, and b) each code should use a minimal number of gestures for fast entry.

Digits

We designed 10 different codes to represent the digits 0 - 9 similar to *DigiTaps*[1.8]. Digit 0 is represented by a one-finger downward swipe, digit 1 by a one-finger tap and digit 2 by a two-finger tap. The digits 3, 4, 5 and 6 (the 3 block) can be represented as (3 + 0), (3 + 1), (3 + 2) and (3 + 3) respectively, hence they are encoded by a three-finger tap followed by a one-finger downward swipe, a one-finger tap, two-finger tap and three-finger tap respectively. The delimiter of 0 at the end of digit 3 ensures prefix-free property.

In *Digitaps*[1.8] [3], the digits 7, 8 and 9 are expressed as (10 - 3), (10 - 2) and (10 - 1) respectively. This is inconsistent with our additive scheme of designing the codes. For this reason, we updated the codes for 7, 8, and 9 to be semantically similar to 4, 5, and 6, using addition rather than subtraction. We denote digits 6, 7, 8 and 9 (the 6 block) as (6 + 0), (6 + 1), (6 + 2) and (6 + 3) respectively and represent the prefix 6 for these digits using a one-finger upward swipe. Hence, 6, 7, 8 and 9 are represented by a one-finger upward swipe followed by a one-finger downward swipe one-finger tap, two-finger tap or three-finger tap respectively. Note that the digit 6 has two different representations. Figure 3 shows a visual representation of the codes for all the digits using visual shortcuts introduced in Figure 2.

Our coding scheme uses 1.7 gestures on average per digit, which is fewer than 1.8 gestures for *Digitaps*[1.8] and 2.5 gestures for *BrailleTaps*. The trade-off as compared to *Digitaps*[1.8] is that we use directional swipes. *GestureCalc* digit codes are therefore slightly more complex, but our pilot study offered evidence that they are still easily learnable. This suggests that the increased benefit for fast character entry may offset the cost of additional complexity.

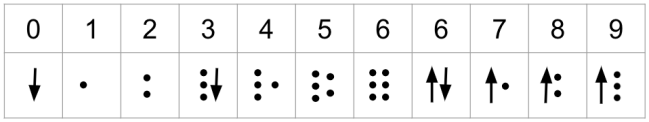

Figure 3: Codes for entering digits.

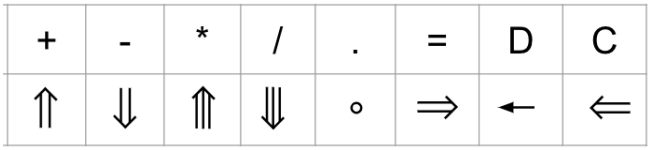

Figure 4: Codes for entering operations.

Operations

For our operations, we oriented our directional swipes according to the conceptual metaphor "more is higher" [25]. For instance, addition increases the value of a number, so we chose to represent addition with a two-finger upward swipe. Subtraction, on the contrary, decreases a number's value and is hence represented by a two-finger downward swipe. It is important to note that our gesture for subtraction can be used as either an operator between two operands or to negate a single

operand. Multiplication implicitly means multiple additions, hence it is represented by a three-finger upward swipe. Similarly, division being multiple subtractions (and the inverse of multiplication) is represented by a three-finger downward swipe. Finally, the '.' or decimal operation is a point and is represented with a long tap. We provide haptic feedback to the user from our application to indicate when the tap is long enough to be recognized as a long tap gesture.

The '=' or equals operation metaphorically moves the expression forward by generating a result. Hence, it is represented by a two-finger horizontal swipe from left to right (abbreviated as two-finger right swipe). Incidentally, this also resembles the shape of the equals symbol. When the user enters '=', the application displays and speaks the computation's result and automatically clears the input for the next computation. The 'D' or delete operation deletes one character at a time and speaks the deleted character. In left-to-right writing systems such as Braille [44] or written English, backspace conventionally deletes a character to the left of the cursor when editing electronically. Hence, we represent this with a one-finger left swipe. The 'C' or clear operation deletes all characters in the input (equivalent to multiple deletions), hence it is represented by a two-finger left swipe. Fig. 4 shows the symbolic representations of the codes for entering operations.

Formative Pilot Study

We conducted a pilot study to evaluate the memorability and intuitiveness of our gesture codes and improve the usability and functionality of our prototype application. We recruited four participants (one of whom was blind) and conducted two 30-minute sessions, each having two tasks. During task 1, participants were asked to verbally describe the gesture code for each character in a random order. The error rate (percentage of wrong answers) in recalling gesture codes decreased to 0 in session 2 itself for all participants, suggesting that our gesture language is easily memorable. During task 2, the participants were required to enter a series of 15 random expressions of varying length into the app. All participants achieved a 3% or less error rate in every session, suggesting that the gesture set and the app have a good level of usability. The rate of character entry increased by 24% from session 1 to session 2, indicating that the speed improves with more practice.

We observed that it was difficult for participants to remember and enter the entire prompted expression. They often asked for repetitions, which affected character entry speed. To avoid this, we reduced the overall length of expressions for the main study and read the expressions in parts. During post-pilot interviews, participants mentioned that grouping the digits and operators to share common types of gesture(s) helped. Participants had difficulty remembering the digit 6 was two consecutive three-finger taps whereas 6 in denoting digits 7, 8, 9 is a one-finger upward swipe. We mitigated this confusion by overloading the code for 6 to include both $3+3$ and $6+0$. Finally, our blind participant suggested reading deleted characters aloud.

Additional Features

Similar to *Digitaps*, *GestureCalc* provides audio feedback (i.e. speaks the entered character) after every digit or operator is entered. *Digitaps* also used different types of haptic feedback for different gestures as an alternative to audio feedback and found that haptic feedback results in faster input. However, haptic feedback was found to have a higher error rate than audio feedback. Error recovery in long calculations is more costly than in simple number entry, since it requires users to start over the expression from the beginning. Additionally, providing multiple types of haptic feedback for our diverse gesture set may confuse users. Our implementation currently supports mathematical calculations involving a maximum of two operands at a time. The readback feature enables users to shake the device to trigger audio feedback, reading aloud the expression entered till now. This feature is designed to be helpful for feedback while entering long expressions.

EVALUATION METHOD

We conducted an IRB-approved study to evaluate the design and implementation of our application. We describe the study methods and outcomes in the following subsections.

Participants

We recruited 8 participants for this study, 6 male and 2 female, ranging in age from 23 to 58. Our inclusion criterion was answering "yes" to the question "When you use a touch screen, do you typically use a screen reader?" One participant used an Android device as their primary touch screen device, all others used Apple devices. All participants except one reported using a touch screen every day. Most participants said they use a calculator at least once a week at home/office/classroom/public places for different activities such as calculating tips, unit conversions, and personal budgeting. Their most common calculations (e.g. addition, multiplication, taking average, calculating percentages) do not require a scientific calculator. Table 1 describes individual participants' mobile device and calculator use in detail. We compensated participants with $80 over three sessions and reimbursement for travel expenses.

Apparatus and Conditions

GestureCalc is developed for iOS using Xcode platform on Mac and *Swift* programming language. For the study, we installed it on an iPhone 7 with touch screen dimensions of 5.44×2.64 inches. We allowed both portrait and landscape modes for using the application.

Most applications in prior research require some target-dependent input, hence we compared *GestureCalc* to the default iOS calculator, which we call *ClassicCalc*. To accurately measure performance and add detailed logging, we recreated the default iOS calculator with same button locations and sizes, same audio labels etc. However, we removed the '%' button and added the functionality of the '+/-' button to the '-' button to be consistent with our *GestureCalc* implementation. In *ClassicCalc*, participants type digits by using the standard eyes-free typing mode of iOS, which involves seeking with audio guidance from VoiceOver, then performing a double-tap or split-tap to activate the selected key. The only target-free calculator close to our work that we are aware of is BrailleTap [1], but we did not compare *GestureCalc* with it because we did not require Braille literacy for study participants.

PID	Age	Gender	Primary touch screen device	Primary mobile input method	primary mobile output method	Mobile device use frequency	primary calculator	calculator frequency
P1	27	M	iPhone	braille keyboard	braille display	daily	Voice assistants	a few times a week
P2	39	M	Key One Blackberry	phone's tactile keyboard	TalkBack	daily	Windows 10 calculator	at least a couple times a month
P3	42	F	iPhone	braille screen input	VoiceOver	daily	Windows calculator	every day at work
P4	46	M	iPad	virtual keyboard w/ magnification, color inversion	magnification, color inversion, VoiceOver	daily	iOS calculator with hand lens	once or twice a week
P5	23	F	iPhone	braille screen input	VoiceOver	daily	iOS calculator	several times a week
P6	58	M	iPhone	touch typing	VoiceOver	daily	Siri, Windows calculator, iOS calculator	daily
P7	27	M	iPhone	braille screen input	VoiceOver	daily	Python console	weekly or every few weeks
P8	46	M	iPhone	touch typing	VoiceOver	weekly	Windows calculator	several times a week

Table 1: Participant demographics and summary of mobile device and calculator use. PID denotes participant ID.

Procedure

We conducted three 1-hour sessions of study with each participant, each pair of consecutive sessions separated by at least 4 hours but not more than 57 hours for a given participant, as in [29]. Each participant used both *GestureCalc* and *ClassicCalc* in every session, counterbalanced so that half of participants used *GestureCalc* first in their first and third sessions, while the rest used *GestureCalc* first in their second session.

In the first session, participants went through a learning period followed by a testing period for each app. The *GestureCalc* learning period started with a facilitator describing each gesture and giving the participants a chance to perform the gesture once. The participants were then asked to type "practice sequences" consisting of the gestures that had just been learned. Practice sequences included "012345689.", "CD", and "+-*/=". Each participant was asked to type each practice sequence three times. The *ClassicCalc* learning period started with a facilitator describing the spatial layout of the classic calculator. The participant was asked to type the same practice sequences as with *GestureCalc*. During the learning period for their first app, each participant was asked to set their preferred VoiceOver speed so that their performance during testing would not be affected due to an uncomfortably fast or slow VoiceOver speed. The same speed was then used for all their testing periods throughout the three sessions.

The testing period for each app consisted of a series of trials. In each trial, the participant was given an arithmetic expression or computation to enter into the calculator. An example of a trial is given below:

Desired input: 72 + 58 =
Transcribed input: 73D2 /D+ 59 =
Final input: 72 + 59 =

Each expression was 4 to 6 characters long and had one of the following two forms:

a) <operand> <operator> <operand> <equals>
b) <operand> <clear>

Each entity (enclosed within <>) was prompted separately from a laptop, allowing the participant to enter the entity on the mobile device before the next entity is prompted from the laptop. This helped the participants to remember the prompts easily while entering them.

The testing period started with 5 unrecorded warm-up trials. The warm-up trials were followed by three blocks of 10 recorded trials. For the recorded trials, participants were requested to "type as quickly and accurately as you can". The expressions in the trials were generated randomly, but we ensured same frequencies across digits and across operators within each block. Each participant was given the same set of expressions in the same order during session 1. During the recorded trials, we recorded a time stamp at the beginning and end of each prompt, gesture (for *GestureCalc*), button press (for *ClassicCalc*) and audio feedback. To calculate the total time for a trial, we subtracted the time taken to prompt the expressions so as to only count the amount of time taken for character entry.

The second session consisted of a testing period for each app followed by an interview. The testing period was conducted exactly as in session 1, except with a different set of expressions for the recorded trials. During the warm-up trials, participants had a chance to re-familiarize themselves with the app and ask for clarifications if needed. Interview methodology is described in the following subsection.

In the third session, there was a testing period followed by a NASA Task Load Index (TLX) [17] for each calculator. The recorded trials used a different set of expressions from session 1 or session 2. For the TLX, the facilitator asked the participants six questions to rate the workload of the application after using each calculator, based on their use of that application in session 3 only.

Interview Methodology

Interviews were semi-structured, organized around five research questions: (a) what are our participants' text input techniques? (b) what are our participants' current calculator needs? (c) what are the trade-offs between gesture calculator and other calculators? (d) what improvements can we make to gesture calculator? and (e) what would it take for gesture calculator to be adopted? Interviews were conducted by one author in English and lasted about 30 minutes.

The interview transcripts and the interviewer's notes were then coded and organized via thematic analysis [10]. The interviews were analyzed by the interviewer and a second author. First, they both independently coded one interview and then discussed all the discrepancies to come to a shared understanding of the initial codes. Each remaining interview was coded by a single author. Our six initial codes (not to be confused with *GestureCalc*'s character codes) were: one for each of the five research questions, plus one based on our literature review, "metaphor." Five inductive codes were added during coding (e.g. "guesswork", a term first mentioned by one of our participants). Codes were applied to arbitrary selections of text.

Study Design and Analysis

Our experiment utilized a $2 \times 3 \times 3 \times 10$ within-subjects design with the following factors and levels:

- *Technique*: Classic Calculator, GestureCalc
- *Session*: 1-3
- *Block*: 1-3
- *Trial*: 1-10

Factors of particular interest were *Technique* and *Session* as we were interested in how the two techniques compared, and how their performance evolved over the 3 sessions. Within each session, each of the 8 participants used both techniques in a series of 3 blocks of 10 trials each, with short breaks in between. Thus, our study data consisted of $8 \times 2 \times 3 \times 3 \times 10 = 1440$ trials in all. For each trial, we computed characters per second (CPS), uncorrected error rate (UER), a binary value indicating whether there were any errors in the final calculation, and the corrected error rate (CER). (Definitions for UER and CER were taken from established text entry research [39].)

For our statistical analyses, we used a linear mixed model analysis of variance for CPS [14, 26, 43]. For our analysis of UER and CER, which do not conform to the assumptions of ANOVA, we used the nonparametric aligned rank transform procedure [19, 37, 38, 46]. In all of these analyses, *Technique* and *Session* were modeled as fixed effects. *Block* was modeled as a random effect nested within *Session* and *Trial* was modeled as a random effect nested within *Block* and *Session*. *Participant* was also modeled as a random effect to account for repeated measures. Any effect of VoiceOver speed was considered to be part of the random effect of *Participant*.

RESULTS

In this section, we present the results of our within-subjects experiment, looking at the effects of *Technique* (GestureCalc/ClassicCalc) and *Session* on characters per second (CPS), uncorrected error rate (UER), number of erroneous calculations (NEC) and corrected error rate (CER).

Characters Per Second

We examined speed or rate of character entry using characters per second (CPS). The average CPS for the *ClassicCalc* was 0.536 (SD=0.150) whereas the average CPS for *GestureCalc* was 0.753 (SD=0.240), a 40.5% speed-up. An omnibus test showed there were significant main effects of *Technique* ($F_{1,1340} = 751.34$, $p < .0001$) and *Session* ($F_{2,6} = 18.24$, $p < .005$) on characters per second. There was also a significant *Technique* $\times$ *Session* interaction ($F_{2,1340} = 27.45$, $p < .0001$). Figure 5 shows the characters per second for *GestureCalc* and *ClassicCalc* over each session, averaged over all participants. The graph shows that *GestureCalc* has higher input speed compared to *ClassicCalc* in all the three sessions.

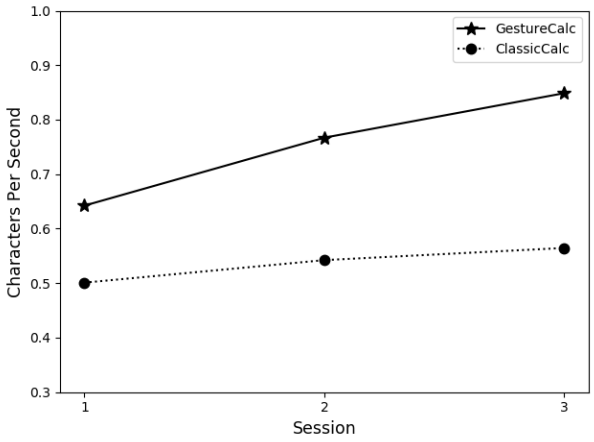

Figure 5: Characters Per Second by *Session* $\times$ *Technique*.

Post hoc pairwise comparisons corrected with Holm's sequential Bonferroni procedure [21] reveal that all pairwise comparisons in Figure 5 are significantly different except for sessions 1 vs. 2 (t_{274} = -1.69, *n.s.*) and 2 vs. 3 (t_{274} = -0.91, *n.s.*) for *ClassicCalc*. *ClassicCalc* did, however, improve from session 1 to 3 (t_{274}) = -2.59, $p < .05$). *GestureCalc*, on the other hand, improved significantly between all sessions: 1 vs. 2 (t_{274} = -5.09, p < .0001), 2 vs. 3 (t_{274} = -3.32, $p < .01$), and 1 vs. 3 (t_{274} = -8.41, $p < .0001$). This indicates that the participants entered characters at a faster rate with more practice with *GestureCalc*, whereas the performance for *ClassicCalc* saturated even with more practice.

Uncorrected Error Rate

Uncorrected error rate (UER) refers to the rate of incorrect characters remaining in the final input. Low UER indicates whether participants could reliably and accurately use each calculator application (i.e. receive the correct output). The average UER for *ClassicCalc* is 2.29% (SD=7.84%) whereas the average UER for *GestureCalc* is 0.92% (SD=4.40%), a 59.8% reduction. An omnibus test showed there was a significant main effect of *Technique* ($F_{1,1340} = 11.06$, $p < .001$) on UER. However, we did not find a main effect of *Session*

($F_{2,6}$ = 2.67, *n.s.*), nor did we find a significant *Technique* × *Session* interaction ($F_{2,1340}$ = 1.55, *n.s.*). This indicates that UER remained almost similar for both methods over all three sessions.

Number of Erroneous Calculations

An 'erroneous calculation' can be defined as any trial that has uncorrected errors in the final input, because such errors would result in incorrect calculations for the user. *ClassicCalc* had 69 (9.58%) erroneous trials whereas *GestureCalc* had only 33 (4.58%) erroneous trials, 52.2% fewer than *ClassicCalc*. A Fisher's exact test [13] shows a significant difference in these error proportions in favor of *GestureCalc* ($p < .001$). A second analysis using logistic regression in a generalized linear mixed model [40] shows that *Technique* had a significant effect on the likelihood of having an erroneous trial ($\chi^2_{(1,N=1440)} = 14.25$, $p < .001$) . There was no *Session* main effect ($\chi^2_{(1,N=1440)} = 3.09$, *n.s.*) or *Technique* × *Session* interaction ($\chi^2_{(2,N=1440)} = 2.81$, *n.s.*) , corroborating the aforementioned analyses of uncorrected error rate. Figure 6 shows that participants entered more erroneous calculations with *ClassicCalc* than with *GestureCalc* in all the 3 sessions.

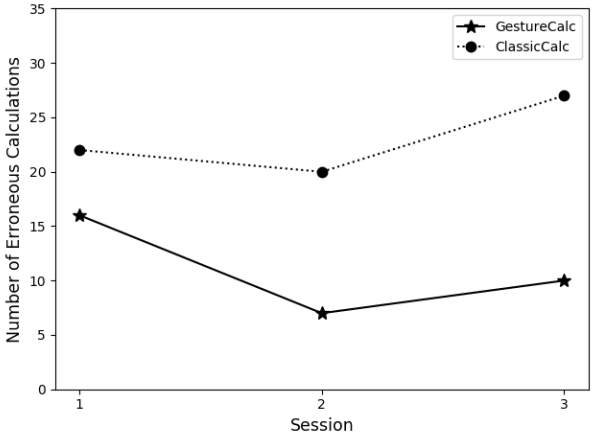

Figure 6: Number of Erroneous Calculations by *Session* × *Technique*.

Corrected Error Rate

Corrected error rate (CER) refers to the rate of incorrect characters in the transcribed input but later corrected in the final input. Therefore, corrected errors do not adversely affect calculator calculations, but they do take time and attention to fix (i.e. using the delete or clear operators). The average CER for *ClassicCalc* is 2.23% (SD=8.00%) whereas the average CER for *GestureCalc* is 5.31% (SD=9.65%), a 138.1% increase. An omnibus test showed there was a significant main effect of *Technique* ($F_{1,1340}$ = 27.17, $p < .0001$) on CER. An omnibus test also showed a main effect of *Session* on CER ($F_{2,6}$ = 12.67, $p < .01$). Finally, we found a significant *Technique* × *Session* interaction ($F_{2,1340}$ = 11.44, $p < .0001$). The CER values decreased after the first session, but increased after the second session.

Post hoc pairwise comparisons conducted with Wilcoxon signed-rank tests [45] revealed that *GestureCalc*'s CER was lower in session 2 than in session 1 ($p = .080$). By contrast, *ClassicCalc* did not show significant changes in CER over sessions. Within session 1, the two techniques were only marginally different ($p = .107$). By sessions 2 and 3, the two techniques were significantly different ($p < .05$).

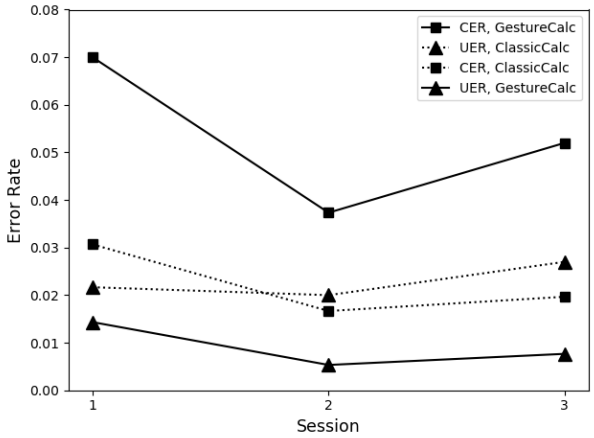

Figure 7: Uncorrected and Corrected Error Rates by *Session* × *Technique*.

PID	GC_{CPS} - CC_{CPS}	CC_{TER} - GC_{TER}	CC_{NEC} - GC_{NEC}
P1	54.1%	-0.013	2
P2	61.4%	-0.028	0
P3	30.3%	-0.031	-3
P4	12.8%	0.034	9
P5	34.8%	-0.067	-5
P6	32.5%	-0.049	1
P7	67.3%	0.045	22
P8	39.3%	-0.029	10

Table 2: Metrics by participant (larger values imply better performance of *GestureCalc* (GC) over *ClassicCalc* (CC). Here CPS is characters per second expressed as a percentage over CPS for CC, TER (Total Error Rate) = UER + CER and NEC is number of erroneous calculations.

Figure 7 shows the UER and CER values using *GestureCalc* and *ClassicCalc* averaged over all participants. We see that participants made more errors while entering expressions using *GestureCalc* compared to using *ClassicCalc*, but were also able to correct the errors more frequently so that the final inputs using *GestureCalc* were more accurate than using *ClassicCalc*. Table 2 shows the performance of individual participants based on our metrics, averaged over all trials in all the 3 sessions. It is important to note that every participant except one had more than 30% faster rate of character entry (CPS) with *GestureCalc* compared to *ClassicCalc* with negligible difference in the total error rate (sum of UER and CER). All but 2 participants also entered fewer erroneous calculations with *GestureCalc* compared to *ClassicCalc*. Overall, for our

participants, *GestureCalc* was more efficient to use and has a better overall performance compared to *ClassicCalc*.

NASA Task Load Index

NASA Task Load Index (TLX) asks participants to rate the workload of a task on six different scales: mental, physical, temporal, performance, effort, and frustration [17]. We used the Wilcoxon signed-rank test (referred to as WST) to determine the statistical significance of differences.

The mental demand scale prompted participants with the question, "How mentally demanding was the task?" and ranged from low (1) to high (20). The average mental demand for *ClassicCalc* was 5.88 (SD=4.64) and for *GestureCalc* was 7.50 (SD=4.47). This difference was not statistically significant according to WST (Z=-0.63, *n.s.*). The higher mental demand of the *GestureCalc* was due to the learning curve, since participants were familiar with *ClassicCalc* but had to familiarize themselves with *GestureCalc*.

The physical demand scale prompted participants with the question, "How physically demanding was the task?" and ranged from low (1) to high (20). The average physical demand for *ClassicCalc* was 3.13 (SD=2.03) and for *GestureCalc* was 3.25 (SD=3.81). This difference was not statistically significant according to WST (Z=0.64, *n.s.*).

The temporal demand scale prompted participants with the question, "How hurried or rushed was the pace of the task?" and ranged from low (1) to high (20). The average temporal demand for *ClassicCalc* was 5.13 (SD=5.03) and for *GestureCalc* was 3.63 (SD=3.11). This difference was not statistically significant according to WST (Z=0.00, *n.s.*). However, the lower demand for *GestureCalc* conforms to our understanding that our gesture set was easy to memorize and enter.

The performance scale prompted participants with the question, "How successful were you in accomplishing what you were asked to do?" and ranged from perfect (1) to failure (20). The average performance rating for *ClassicCalc* was 4.63 (SD=4.41) and for *GestureCalc* was 5.50 (SD=3.34). This difference was not statistically significant according to WST (Z=-0.70, *n.s.*). It is important to note that this rating is in line with our observation of higher CER but lower UER of *GestureCalc* compared to *ClassicCalc*. Participants rated their performance more towards failure because they were aware of the mistakes they made during character entry.

The effort scale prompted participants with the question, "How hard did you have to work to accomplish your level of performance?" and ranged from low (1) to high (20). The average effort rating for *ClassicCalc* was 7.00 (SD=4.72) and for *GestureCalc* was 6.38 (SD=4.41). This difference was not statistically significant according to WST (Z=0.07, *n.s.*) but participants agreed that *GestureCalc* requires less effort than *ClassicCalc*.

INTERVIEW RESULTS

In this section we summarize results from interviews with individual participants after they completed the second session.

Calculator Use

Our participants regularly do a wide variety of calculations, calculations involving money being more frequent. For personal finances, participants compute tips, monthly expenses, and expenses in the grocery store. At work, participants compute bills and financial estimates. One participant primarily uses calculator for unit conversion. P5 highlighted the importance of accessible calculators when she told us, "*Right now, I'm studying to take the pre-test to get into math classes. I haven't been able to take the math test yet because I don't have a decent calculator.*" Indeed, participants often preferred to use a laptop or desktop computer for doing calculations because digits can be typed directly with keyboard buttons, but perhaps such a computer is not allowed in P5's math test.

Device Issues

Physicality of devices and fingers played a role in participants' ability to perform gestures accurately. Our phone was too narrow for some participants to make a three-finger tap in portrait mode (the default), so they switched to landscape. P2 (among others) noted it was important to know where the edge of the screen was in order to perform the gestures accurately, saying, "*You had to keep your fingers in the middle of the screen, no matter if you were doing it in landscape or portrait.*" P8 described his own phone case, which has "*a raised rim around both sides, so your fingers don't actually get off the touch area of the screen.*" P2 also pointed out that sweaty fingers, residue on the screen, or a moving environment such as a bus, could affect gesture performance.

Causes of Errors and Confusion

Participants' feedback on the relative difficulty of different gestures varied widely across participants, often conflicting. One common thread, however, was that participants found operators relatively intuitive (P1, P6, P7). In fact, one participant accidentally tried swiping to delete while using the classic calculator. The different blocks in the gesture set design for digits caused confusion. Despite disagreement on the relative difficulty of the 3 block and the 6 block, P6 identified context-switching between the blocks as an important hurdle, saying, "*making that transition like from a 5 to a 7, that's a little challenging.*"

Memorability and Mental Demand

The mental demand of remembering the codes was the primary aspect (that they spoke about) of participants' experience using *GestureCalc*, and it was a source of error and confusion. Each participant's first session with *GestureCalc* started with the facilitator teaching them to use it, so they all had similar, structured introduction to the codes. As P8 put it, using *GestureCalc* "*takes a little rearranging of your way of thinking.*" Through practice and repetition, however, participants felt they could learn the codes.

Like *GestureCalc*, learning Braille also requires memorization. P4 explained, "*It's not a hard learning curve . . . I started to get a little anxious at first, because I thought, 'Oh, no, here we go with Braille again.' But once I just let that thought go . . . No, I wouldn't hesitate to tell anybody to try this.*" For him, learning our codes was not as challenging as learning

Braille. P7, on the other hand, suspected that people may not bother learning the codes, saying, "*I wouldn't call this a really steep learning curve, [but] there's a learning curve. If I'm downloading a calculator, I want to be able to just start using it.*" Helping people learn the codes will be an important hurdle in achieving broad impact of *GestureCalc*.

Feature Suggestions

Indeed, most of the participants identified that new users would need a tutorial or manual to learn the gesture set. It could either be an interactive tutorial, "*similar to [the learning period] that we did in the first session*" (P7), a text-based "*quick-reference*" (P8), or both. We heard a wide variety of suggestions for improving *GestureCalc*, including (a) additional mathematical operations to support, (b) using the iOS VoiceOver instead of our custom self-voicing (to avoid having to turn off VoiceOver and to share VoiceOver's settings), and (c) ways to navigate, edit, and read back the input or repeat the output. To our surprise, the existing input readback feature (activated by shaking the device) was rarely used.

Some participants suggested redesigning the gesture set to be completely different. For example, the digit codes could be closely based on the spatial location of each digit key on a phone keypad, incorporating upward swipes for keys on the top row, and downward swipes for keys on the bottom row (P3), or numeric gestures could have the user "*trace a number onto the screen*" (P8). P1 participant suggested keeping the operator gestures, but replacing digit gestures with Braille input or a virtual keypad. He also suggested tracing a square on the screen for squaring a number. As a mnemonic device, mapping the squaring operation to the shape of a square seems promising (at least for English-speaking users), though previous work argues that a shape gesture may be challenging to recognize accurately while supporting blind users [24].

Helpfulness of Metaphors

In *GestureCalc*, the gesture codes are based on particular metaphors (e.g. "A then B" for gestures within digits means "$A+B$", "up swipes mean increasing" etc.), which we referenced in the training to help participants remember the codes. P6 mentioned that these metaphors were useful for him. He explained, "*The fact that [the codes] are in a way that makes sense—up, more, larger—down, smaller—and those connect well with plus, minus, times, divide. They have projected outcomes that you can relate to easily that makes it user friendly.*" P8 drew attention to the fact that all the metaphors we chose are not universal. Explaining his difficulty with the gestures for 3, 5 and 6, he referred to the uniqueness of his cognition: "*It may have a lot to do with the way I coordinate, my own thought processes, and the way my brain works.*" This reminds us that different gestures, different codes, and different calculators should work better for different people.

Comparisons between Calculator Apps

Comparing *GestureCalc* to *ClassicCalc*, participants found it helpful that with *GestureCalc* they can perform the gesture anywhere on the screen, but they recognized that in return this requires memorizing the gestures. Though participants were generally positive about *GestureCalc*, one (P8) felt that he was slower with *GestureCalc* because it took "*more movement*" to perform the gestures, and another (P2) felt that he made more mistakes with the *GestureCalc*. P5 pointed out the main improvement of *GestureCalc* over other phone calculators when she remarked, "*gestures are good; they take the guesswork out.*" The guesswork involved in using a screen reader is to guess where to move your finger on the screen to find what you want. In the words of P1, *GestureCalc* would improve on this since, "*you wouldn't have to hunt around the screen.*"

Some participants identified a similarity between *GestureCalc* and Braille input systems, that both encode digits and operators symbolically, rather than spatially. As P5 put it, "*For me, Braille screen input sped up my typing. ... I think that Gesture Calculator would be the exact same thing. Once people got familiar with how to use it, I think it would speed up calculations.*" Participants who use Braille inputs appreciated *GestureCalc* for this similarity to Braille inputs.

DISCUSSION

In this section we discuss the limitations of our study and issues that arose from the study and interviews.

Limitations

Due to difficulty in recruiting blind participants, we conducted our study with only 8 participants and over only 3 sessions. The limited sample pool and size of our study provides good evidence, but are not fully generalizable nor conclusive. In addition, the counterbalancing of the order of app usage for each participant was not ideal because of the odd number of sessions. Nonetheless, we still observe considerable uniformity in relative performance of individual participants on the two calculators in terms of the metrics we used (see Table 2).

Our *ClassicCalc* app supported standard typing but not touch typing[2], to be consistent across all participants regardless of the typing mode they are used to. While standard typing is thought to be the more conventional mode of text entry, most of our participants expressed a preference for touch typing, and some said it was faster for them than standard typing. P6 acknowledged a speed-accuracy tradeoff, noting that touch typing is more error prone. Those who preferred touch typing, however, were also familiar with standard typing, and appeared quite fluent in the technique.

Scalability: Adding Operations

Though *GestureCalc* supports the most common basic mathematical operations, some participants suggested adding scientific operations. If we extended *GestureCalc* we would follow participant suggestions to increase the memorization requirements as little as possible, leaning toward a modal approach that does not add unique gestures for each new symbol but instead uses one or more new prefix gestures (selected from the unused gestures in Figure 2) to enable reuse of existing

[2]VoiceOver supports two modes of typing, standard typing and touch typing. In standard typing, the user seeks for a key by sliding a finger over the screen, and VoiceOver speaks the name of each key as it is touched. The user performs a double tap or split tap anywhere on the screen to activate last key spoken. In touch typing, the user slides a finger across the screen to seek as VoiceOver reads they key under the finger, and lifting the finger activates the last key touched.

gestures to represent new operators and symbols. Another possibility is introducing an escape character to enter a mode where the user can swipe through a set of less common operators and symbols; a repetitive swiping design is a common approach to exploring and selecting from lists with a screen reader. These approaches would leave the existing gesture set as-is, following the design principle of making common interactions easy, and uncommon interactions possible.

Error Rates and Speedup

A few participants (P2, P3, P4) observed that the gesture recognition was very sensitive, often capturing stray accidental touches and interpreting them as gestures. Simple gesture codes such as "1" (one-finger tap) can be typed quickly and easily, but could also easily be generated by accidental touches. In contrast, *ClassicCalc* is robust to stray touches because keys are activated by double tap or split tap, which is much less likely to occur accidentally, but also makes *ClassicCalc* slower. This speed-accuracy trade-off may have contributed to the higher error rate with *GestureCalc*. Two-gesture digits may have also raised the error rate with *GestureCalc*. For example, consider the situation where a user tries to type "7": a one-finger upward swipe followed by a one-finger tap. If the swipe is incorrectly recognized as a tap and the user continues typing, they end up typing "11", generating two errors that both subsequently need to be deleted.

For *ClassicCalc*, the amount of time spent typing can be broken down into think time (deciding what number to type), seek time (finding the key), and gesture time (double or split tap). Because of its target-free design, *GestureCalc* eliminated the need to seek for keys, thereby reducing the app's verbosity. Participants appreciated the low verbosity of *GestureCalc*. P8 said "*The amount of speech [in GestureCalc] was right, it wasn't overly wordy... one of the problems I have with a lot of talking devices is they just go on and on and on.*" The low verbosity of *GestureCalc* might have made it easier for participants to recognize errors made while typing, contributing to the low uncorrected error rate of *GestureCalc*. Furthermore, the elimination of seek time (one participant expressed frustration at having to seek for the delete key with *ClassicCalc*) and intuitiveness of the delete gesture may have made encouraged error correction with *GestureCalc*, further reducing uncorrected error rates. Though the rich gestures of *GestureCalc* may take longer than the simple double tap or split tap of *ClassicCalc*, we believe that much of the *GestureCalc* speedup can be attributed to elimination of seek time.

Design and Implementation

Out of all the entries for the digit 6, 57% were in the form 3+3 and 43% were in the form 6+0, indicating that the option of having two possible representations made the participants free to choose whichever they remembered on the fly. This validates our design decision to add the alternate gesture code for 6. Participants found three-finger gestures difficult to perform compared to one- or two-finger gestures. This was often because it was difficult for them to synchronize three differently sized fingers to touch the screen at the same time or to fit three fingers within the limited screen width. Two out of 8 participants preferred landscape mode compared to portrait mode because of more space to fit three fingers. However, while landscape mode helped with three-finger taps and horizontal swipes, portrait mode was preferred for vertical swipes.

Impact

Potential use cases for *GestureCalc* are similar to use cases for mobile phone calculators for sighted people: for people without PCs, when in public and/or noisy areas (P7), calculating tips (P8), shopping (P6), or at their computer when they have too many windows open (P6). Although it is true that any mobile phone calculator app has limitations (e.g., less powerful than a desktop computer for calculations), *GestureCalc* supports the most common use cases. Furthermore, our contribution extends beyond the app to include our exploration of target-free metaphorical gestures for future applications.

GestureCalc employs what we call a *rich gesture* set, meaning it involves target-free gestures, where the information is contained in the gesture itself, rather than simple targeted gestures where the information is contained in target location. Though *GestureCalc* may be a small application, in the bigger picture it demonstrates the utility of rich metaphorical gestures. Target-free rich gestures can provide as much information as targeted simple gestures, such as the location-specific taps that a sighted user use for a touch screen calculator, yet can be more accessible to a greater population of users. Apple's iOS takes advantage of this by employing rich gestures to facilitate use of various apps. As a consequence, it is hard to implement rich gesture apps that are compatible with VoiceOver, since some of the most desirable rich gestures activate VoiceOver features. One potential design direction is for gesture-based apps to register their gestures with VoiceOver, thereby disabling those gestures for VoiceOver while the app is active.

CONCLUSION AND FUTURE WORK

We designed a basic digital calculator application that allows eyes-free target-free input of digits and operations on touch screen devices. Our input is based on simple gestures like taps and swipes with one, two or three fingers and the coding scheme is based on conceptual metaphors to improve the intuitiveness and memorability. We conducted a study with blind participants to determine the usability, learnability and memorability of our gesture set design and implementation. Participants entered characters at a significantly faster rate and made less errors in the final calculations compared to the Classic Calculator. This indicates the efficiency of our application, which is validated by detailed study and analysis.

In the future, we plan to compare the performance of our application with a wider variety of state-of-the-art calculator applications. Finally, our current application has room for improvement: some gestures from the set of possible gestures and screen-edge gestures [24] are not encoded, parentheses are not represented, and our application only accepts two operands. To further improve the accessibility of ubiquitous touch screen calculators, we will further develop our calculator application to include a more robust set of operations.

ACKNOWLEDGMENTS

We would like to thank our participants for the pilot study and the main study for devoting their time and effort in our study.

REFERENCES

1. Mrim Alnfiai and Srinivas Sampalli. 2017. BrailleTap: Developing a Calculator Based on Braille Using Tap Gestures. In *Universal Access in Human–Computer Interaction. Designing Novel Interactions*. Springer International Publishing, Cham, 213–223.
2. Apple, Inc. 2018. iPhone Accessibility. (2018). https://www.apple.com/accessibility/iphone/vision/
3. Shiri Azenkot, Cynthia L. Bennett, and Richard E. Ladner. 2013. DigiTaps: Eyes-free Number Entry on Touchscreens with Minimal Audio Feedback. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. 85–90.
4. Shiri Azenkot, Richard E Ladner, and Jacob O Wobbrock. 2011. Smartphone haptic feedback for nonvisual wayfinding. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 281–282.
5. Shiri Azenkot, Kyle Rector, Richard Ladner, and Jacob Wobbrock. 2012a. PassChords: secure multi-touch authentication for blind people. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 159–166.
6. Shiri Azenkot, Jacob O. Wobbrock, Sanjana Prasain, and Richard E. Ladner. 2012b. Input Finger Detection for Nonvisual Touch Screen Text Entry in Perkinput. In *Proceedings of Graphics Interface 2012 (GI '12)*. 121–129.
7. BIHA Studio. 2013. Sumzy for iPhone. (2013). https://www.youtube.com/watch?v=pHCoEIk_cB4
8. Frank Boers. 1999. When a bodily source domain becomes prominent: The joy of counting metaphors in the socio-economic domain. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4* (1999), 47–56.
9. Matthew N Bonner, Jeremy T Brudvik, Gregory D Abowd, and W Keith Edwards. 2010. No-look notes: accessible eyes-free multi-touch text entry. In *International Conference on Pervasive Computing*. Springer, 409–426.
10. Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
11. Bill Buxton. 1984. Touch Screen Calculator Watch. (1984). https://www.youtube.com/watch?time_continue=1&v=UhVAsqhfhqU
12. James Clawson, Kent Lyons, Thad Starner, and Edward Clarkson. 2005. The impacts of limited visual feedback on mobile text entry for the twiddler and mini-qwerty keyboards. In *Wearable Computers, 2005. Proceedings. Ninth IEEE International Symposium on*. IEEE, 170–177.
13. Ronald A Fisher. 1922. On the interpretation of χ 2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society* 85, 1 (1922), 87–94.
14. Brigitte N Frederick. 1999. Fixed-, Random-, and Mixed-Effects ANOVA Models: A User-Friendly Guide for Increasing the Generalizability of ANOVA Results. (1999).
15. Google, LLC. 2018. Android Accessibility Suite. (2018). https://play.google.com/store/apps/details?id=com.google.android.marvin.talkback
16. Anne Hamilton. 2000. Metaphor in theory and practice: the influence of metaphors on expectations. *ACM Journal of Computer Documentation (JCD)* 24, 4 (2000), 237–253.
17. Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139 – 183.
18. Tobias Hesselmann, Wilko Heuten, and Susanne Boll. 2011. Tap2Count: numerical input for interactive tabletops. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. ACM, 256–257.
19. James J Higgins and S Tashtoush. 1994. An aligned rank transform test for interaction. *Nonlinear World* 1, 2 (1994), 201–211.
20. Alexis Hiniker, Sungsoo Hong, Yea-Seul Kim, Nan-Chen Chen, Jevin D West, and Cecilia Aragon. 2017. Toward the operationalization of visual metaphor. *Journal of the Association for Information Science and Technology* 68, 10 (2017), 2338–2349.
21. Sture Holm. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* (1979), 65–70.
22. Jörn Hurtienne and Luciënne Blessing. 2007. Design for Intuitive Use-Testing image schema theory for user interface design. In *ICED*, Vol. 7. Citeseer, 1–12.
23. Shaun K Kane, Jeffrey P Bigham, and Jacob O Wobbrock. 2008. Slide rule: making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 73–80.
24. Shaun K. Kane, Jacob O. Wobbrock, and Richard E. Ladner. 2011. Usable Gestures for Blind People: Understanding Preference and Performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. 413–422.
25. George Lakoff and Mark Johnson. 1980. The metaphorical structure of the human conceptual system. *Cognitive science* 4, 2 (1980), 195–208.
26. RC Littell, PR Henry, and CB Ammerman. 1998. Statistical analysis of repeated measures data using SAS procedures. *Journal of animal science* 76, 4 (1998), 1216–1231.

27. Diana Loeffler, Anne Hess, Andreas Maier, Joern Hurtienne, and Hartmut Schmitt. 2013. Developing intuitive user interfaces by integrating users' mental models into requirements engineering. In *Proceedings of the 27th International BCS Human Computer Interaction Conference*. British Computer Society, 15.
28. Diana Löffler, Klara Lindner, and Jörn Hurtienne. 2014. Mixing languages': image schema inspired designs for rural Africa. In *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 1999–2004.
29. I. Scott MacKenzie and Shawn X. Zhang. 1999. The Design and Evaluation of a High-performance Soft Keyboard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. 25–31.
30. Sergio Mascetti, Cristian Bernareggi, and Matteo Belotti. 2011. TypeInBraille: a braille-based typing application for touchscreen devices. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 295–296.
31. Mehmed Mert and Ergün Kayis. 2013. Swipe Calculator for Android. (2013). `http://www.swipecalculator.com/`
32. MyScript. 2013. MyScript Calculator 2. (2013). `https://www.myscript.com/calculator/`
33. National Federation of the Blind. 2009. The Braille Literacy Crisis in America. (2009).
34. João Oliveira, Tiago Guerreiro, Hugo Nicolau, Joaquim Jorge, and Daniel Gonçalves. 2011. BrailleType: unleashing braille over touch screen mobile phones. In *IFIP Conference on Human-Computer Interaction*. Springer, 100–107.
35. Rechner. 2012. Rechner Calculator. (2012). `http://rechner-app.com/`
36. Vaspol Ruamviboonsuk, Shiri Azenkot, and Richard E. Ladner. 2012. Tapulator: A Non-visual Calculator Using Natural Prefix-free Codes. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '12)*. 221–222.
37. KC Salter and RF Fawcett. 1985. A robust and powerful rank test of treatment effects in balanced incomplete block designs. *Communications in Statistics-Simulation and Computation* 14, 4 (1985), 807–828.
38. KC Salter and RF Fawcett. 1993. The ART test of interaction: a robust and powerful rank test of interaction in factorial models. *Communications in Statistics-Simulation and Computation* 22, 1 (1993), 137–153.
39. R William Soukoreff and I Scott MacKenzie. 2003. Metrics for text entry research: an evaluation of MSD and KSPC, and a new unified error metric. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 113–120.
40. Robert Stiratelli, Nan Laird, and James H Ware. 1984. Random-effects models for serial observations with binary response. *Biometrics* (1984), 961–971.
41. Hussain Tinwala and I Scott MacKenzie. 2009. Eyes-free text entry on a touchscreen phone. In *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)*. IEEE, 83–88.
42. Amy Vidali. 2010. Seeing what we know: Disability and theories of metaphor. *Journal of Literary & Cultural Disability Studies* 4, 1 (2010), 33–54.
43. Brady T West. 2009. Analyzing longitudinal data with the linear mixed models procedure in SPSS. *Evaluation & the health professions* 32, 3 (2009), 207–228.
44. Wikipedia. 2018. Braille. https://en.wikipedia.org/wiki/Braille. (2018).
45. Frank Wilcoxon. 1945. Individual comparisons by ranking methods. *Biometrics bulletin* 1, 6 (1945), 80–83.
46. Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 143–146.