Interaction and Recognition Challenges in Interpreting Children's Touch and Gesture Input on Mobile Devices

Lisa Anthony[1], Quincy Brown[2], Jaye Nias[2], Berthel Tate[2], Shreya Mohan[1]

[1]UMBC
Information Systems
1000 Hilltop Circle
Baltimore MD 21250 USA
lanthony@umbc.edu

[2]Bowie State University
Computer Science Department
14000 Jericho Park Road
Bowie MD 20715 USA
qbrown@bowiestate.edu

ABSTRACT

As mobile devices like the iPad and iPhone become increasingly commonplace, touchscreen interactions are quickly overtaking other interaction methods in terms of frequency and experience for many users. However, most of these devices have been designed for the general, typical user. Trends indicate that children are using these devices (either their parents' or their own) for entertainment or learning activities. Previous work has found key differences in how children use touch and surface gesture interaction modalities *vs.* adults. In this paper, we specifically examine the impact of these differences in terms of automatically and reliably understanding *what kids meant to do*. We present a study of children and adults performing touch and surface gesture interaction tasks on mobile devices. We identify challenges related to (*a*) intentional and unintentional touches outside of onscreen targets and (*b*) recognition of drawn gestures, that both indicate a need to design tailored interaction for children to accommodate and overcome these challenges.

Author Keywords

Touch interaction; surface gesture interaction; child computer interaction; mobile devices; touchscreens; Android; $N; gesture recognition.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

As mobile devices like the iPad and iPhone become increasingly more commonplace, touchscreen interactions are overtaking other interaction methods such as keyboard and mouse in terms of frequency and experience for many users. Furthermore, educators and researchers are identifying mobile devices as a powerful platform to provide personalized learning environments, games and entertainment for kids [6,20,26,27]. However, most of these devices are mass-market, consumer-oriented products that have been designed for the general, typical user. Trends indicate that children are using these devices (either their parents' or their own) for entertainment or learning activities (Figure 1) [27]. Children have smaller fingers, weaker arms, less fine motor control, less manual dexterity, and (typically) less experience with technology than adults. These factors, among potentially others, contribute to key differences in how children can, expect to, and do use touch and gesture interaction and impact their success with mobile devices. So far, this impact has not been examined thoroughly for design implications.

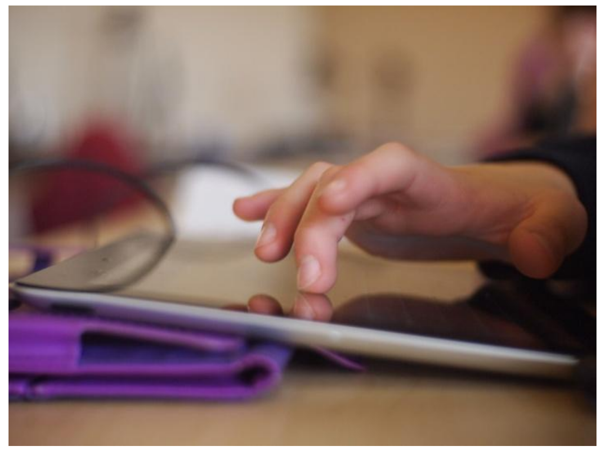


Figure 1. A child using an iPad.
Image courtesy of Lexie Flickinger.
http://www.flickr.com/photos/56155476@N08/6660076003/

Previously we have found key differences in how children use touch and surface gesture interaction modalities when compared to adults [4]. That work was a small study focusing on characterizing differences and theorizing about their impact, *e.g.*, in terms of missed touch events and bad recognition. In this paper, we expand on that work with a larger study of children and adults performing similar touch and surface gesture interaction tasks on mobile devices, in which the goal is to identify interaction and recognition challenges in *interpreting users' intended input*, especially for children. Given a touch event, can we be reasonably certain of which target the child intended to hit? Given a drawn gesture, can we confidently recognize and interpret

ITS'12, November 11–14, 2012, Cambridge, Massachusetts, USA.

the desired command? Similar issues may occur for adults, but in our study, we find much higher proportions of challenging input from kids.

We specifically find issues related to (*a*) intentional and unintentional touches outside of onscreen targets and (*b*) recognition of drawn gestures, both of which indicate a need to design tailored interaction for children to accommodate and overcome these challenges. We analyze patterns of touches for onscreen targets, and we present recognition accuracy results of a modern user interface gesture recognizer called $N [1,2] on the gestures generated by children and adults. Understanding these challenges, and designing interactions and implementing algorithms to overcome them, is critical to providing children the most successful user experience possible on mobile devices.

RELATED WORK

Touch and Surface Gesture Interactions

Prior work has examined the usability of touch and surface gesture interaction for *adults* on a variety of platforms, *e.g.*, PDAs and mobile devices [31], Tablet PCs and tablet computers [28,29], tabletop and surface displays [10,23,25], but *children* have been largely overlooked. In studies that did include children, analysis grouped children and adults together [25], focusing on general factors rather than on *differences* between children and adults that could be used to design better interactions. Some research has investigated only children, impeding the comparison of adults' and kids' interaction patterns [10,23]. More recently, researchers are specifically acknowledging and investigating differences between children and adults as users of such devices [4,11], but more work is needed in this area.

Children's Pointing Interactions

Prior studies have been conducted comparing differences in how children and adults acquire targets using mice or other pointing devices by studying how well Fitts' law (developed for adults) applies to children [7,12,16]. Typically, performance on pointing tasks increases for older children. Prior work has also found consistent evidence that children have difficulty with drag-and-drop interactions [14,15], implying that simple pointing is preferable. Pointing-based studies must be updated and extended to include modern touchscreen interactions with finger-based touch and surface gesture input.

Children's Gestures and Mobile Interactions

Stylus-based handwriting interaction has been extensively studied with children with the goal of comparing to analogous investigations with adults (*e.g.*, [22]). So far, work that looks at direct touch and gesture interaction for children has examined this question primarily on tabletop surfaces [10,11,23], and we do not yet know whether those findings will also apply to smaller-screen mobile devices.

Recent past work has examined children's direct touch and gesture interactions on mobile applications [4,5], and found that children experience different challenges in each interaction mode (*e.g.*, touch *vs.* gesture). For example, swipe gestures, which must be executed in one smooth stroke, are difficult for children, who tend to lift their fingers during the stroke [5]. Also, children have trouble accurately touching small onscreen targets, requiring larger interactors onscreen (*e.g.*, widgets and icons) and reducing the amount of onscreen information that could be displayed [4,5]. These findings form a good foundation but must be deepened to allow us to design interactions and technology that will be successful for the range of children's abilities.

Gesture Recognition Approaches and Performance

Once we can characterize the gesture input we can expect to see from children for various tasks, we can develop recognizers that are tailored to kids' input patterns. Current common gesture recognition approaches for user interfaces such as $N perform at reasonably high accuracy levels (*e.g.*, 95 to 99%) [1,2], but testing of such recognizers has used gestures from adults only, not children. Prior published benchmark datasets for gestures and handwriting recognition have only included samples from adults, not children [1,8,13,30]. No systematic testing of gesture recognition for children yet exists. These recognition benchmarks must be extended to represent the broader population of potential users of such technology.

APPROACH

To prepare for this work, a survey of mobile applications in the Android Marketplace was conducted. The surveyed applications were evaluated to identify the interaction modes employed in mobile games. Of the 23 surveyed game apps, 21 utilized touch as the primary mode of user interaction. Six applications employed surface gestures as a second mode of interaction, and only three employed tilting as an interaction mode. When examining children's education apps specifically, we find that tracing or handwriting practice activities using letters and numbers are commonly used, for example, Jaloby's AlphaCount[1], gdiplus' I Write Words[2], or Critical Matter, Inc.'s ABC Letter Tracing[3]. Additionally, prior work has found that users prefer gesture short cuts for commands that are character-based [18]. Given these results, we have chosen to focus our current investigations on touch and surface gesture interactions. We have developed two controlled tasks that resemble real-world touch and gesture interaction activities such as tapping, sketching and writing [4]. These abstracted tasks simplify the interaction to basic elements and eliminate any possible effects of application context.

PRELIMINARY STUDY

This work extends a preliminary study we conducted [4]. In that study, 8 children (aged 7 to 11 years) and 6 adults

---

[1] http://itunes.apple.com/us/app/alphacount/id359046783?mt=8
[2] http://itunes.apple.com/us/app/iwritewords-handwriting-game/id307025309?mt=8
[3] http://itunes.apple.com/us/app/abc-letter-tracing-free-writing/id416326981?mt=8

(aged 18+ years) participated. This work seeks to expand on that investigation by considering a wider range of users. We briefly review that study and its findings, pointing toward where this study replicates *vs.* improves on the previous work. In the preliminary study, two tasks were used: (1) a touch target acquisition task; and (2) a surface gesture interaction task. This study uses similar tasks with a few significant changes, which we now describe.

Preliminary Study: Touch Target Acquisition Task

This task involved users touching targets drawn onscreen [4]. The preliminary study used 43 square targets of 4 different sizes: extra-small (20 pixels), small (40 pixels), medium (60 pixels), and large (100 pixels), in 13 different interface positions [4]. In this study, we use the same four target sizes but define them based on absolute physical size rather than pixels, to account for varying screen resolutions and pixel densities (see Study Design section). In the preliminary study, participants tapped the target on the screen; the software automatically advanced to the next target upon registering a tap event, even if the tap was not within the target [4]. In this study, we require the user to touch the target correctly before moving on, recording every touch event generated per target.

Preliminary Study: Gesture-Based Interaction Task

This task involved users drawing samples of eight different gestures, *e.g.*, letters or shapes [4]. In this study, we expand the set to 20 gestures, allowing us to examine more types of gestures such as numbers and symbols. In the preliminary study, no visual feedback was given to the users as they drew gestures [4]; in this study, we show a trace of the user's finger path on the screen to more closely mimic real-world touch and gesture sketching and drawing activities. Lastly, in the preliminary study, users only drew one example per gesture type, limiting the types of gesture recognition testing that could be performed [4]. In this study, we collect six samples per gesture type to enable exploration of within-user recognition and consistency.

Preliminary Study: Results

In general, the preliminary study found several significant differences between children and adults in how they completed the two tasks. For example, children missed targets 50% more often than adults, by a greater distance, and had more trouble with smaller targets than did adults [4]. The children in the preliminary study also made significantly smaller touch points but exerted significantly less pressure than did adults [4]. Also, recognition of children's gestures in the preliminary study was more challenging for modern recognizers than recognition of adults' gestures, likely due to the inconsistencies in how children made gestures (*e.g.*, generating more individual strokes, writing with a slower input speed) [4]. In addition to these findings, touches that were located in the vicinity of the previous target, *holdovers*, were also observed. These results point to several promising avenues for exploration of tailored interactions for children, but the small sample size and limited task set restrict the scope of the findings. Therefore, we have extended this preliminary study and addressed these limitations in order to yield broader results.

STUDY DESIGN

The study consisted of two phases: (1) a touch task designed to gather data on touch target acquisition; and (2) a gesture task designed to gather data on gesture-based interaction. The target acquisition task was intended to investigate interactions similar to those required when performing tasks such as tapping in a game or pressing an interface widget (*e.g.*, checkbox, menu item). The gesture task was designed to examine users' surface gestures entered with visual gesture feedback. All participants did both tasks in the same session, with breaks between tasks.

Participants

A total of 30 participants (16 children, 14 adults) participated in the study. The ages of the children varied from 7 to 16 years (mean: 11.5 yrs, stdev: 2.8 yrs), and all of the adults were over the age of 18 (mean: 22.5 yrs, min: 18 yrs, max: 29 yrs, stdev: 3.6 yrs). Of the 14 adults, 6 were female, and of the 16 children, 8 were female. The large majority of our participants were right-handed (24 of 30); three considered themselves ambidextrous, and three were left-handed.

In terms of self-reported familiarity with touchscreen devices, the adults rated themselves "average" (5 of 14, or 36%) or "expert" (9 of 14, or 64%), whereas 2 of the 16 (13%) children rated themselves as "beginners," 5 of 16 (31%) rated themselves "average," and 9 of 16 (56%) rated themselves "expert." For the children, these self-ratings were not correlated with age (r=-.08, *n.s.*, N=16). Table 1 shows the percentage of participants (adults and children) that owned and used various touchscreen devices. These demographics reflect the increased pervasiveness of touchscreen devices such as smartphones and iPads. We did not ask about tabletops and surfaces in our study because they are still very uncommon in the home environment.

		Mobile Phone	Tablet	MP3 Player	Tablet PC
Adults	**Own it**	79%	21%	7%	7%
	Use it daily	79%	29%	7%	0%
Children	**Own it**	50%	19%	44%	6%
	Family owns it	75%	56%	31%	6%
	Use it daily	63%	19%	38%	0%

Table 1. Distribution of ownership and usage of touchscreen devices among our participants. Note that, in our sample, no children under age 10 owned their own touchscreen devices.

Equipment

Both phases of the experiment were performed on Samsung Google Nexus S smartphones running the Android 4.0.4 operating system; the phones measured 4.88 x 2.48 x 0.43

inches (123.9 x 63 x 10.9 mm) with a 4-inch (101.6-mm) screen, measured diagonally. The display resolution was 480 x 800 pixels, with a pixel density of approximately 233 pixels per inch (ppi). The study interfaces and software were designed in Eclipse using the Android SDK. Note that this particular Android device does support multitouch interaction, but our apps were not designed to respond to it.

Phase 1: Touch Interactions

The goal of the touch task was to extend the previous study's findings regarding differences between the touch patterns of adults and children for specific targets of varied sizes and locations. The goal was to learn whether it was possible to reliably correlate users' touches to the desired targets, *e.g.* figure out what target they meant to touch, and to compare these patterns for kids and adults.

Design

This phase of the study used 104 targets of 4 different sizes: very small, small, medium, and large, in 13 different interface positions. Because a user's finger remains a constant size no matter what device he or she is using, we aimed for cross-platform uniformity in the absolute physical sizes of the targets: (a) large: half-inch squares (12.7 mm); (b) medium: three-eighths-inch squares (9.5 mm); (c) small: quarter-inch squares (6.35 mm); and (d) very small: one-eighth-inch squares (3.175 mm). The device-specific pixel size of each target was computed based on the device-reported ppi. We hypothesized that size of the target would be inversely related to difficulty: smaller targets were expected to be harder than larger targets. We used data from the pre-study mobile game app survey to determine the appropriate target sizes for the study. Additionally, Android recommends 7 to 10 mm. Thus, we also selected targets slightly smaller and larger than what we found in practice and in the guidelines in order to probe the boundaries of acceptable target sizes.

The 13 possible target locations were defined by dividing the display into a 3 x 5 grid. Of the 15 areas, 13 were selected using locations frequently identified in the surveyed apps, *e.g.*, along edges, in corners, in the center of screen. Half the targets were drawn with edge padding, causing them to appear slightly inset from the edge of the screen, whereas the other half were drawn exactly aligned with the edge of the screen. Edge padding was a constant value of 10 pixels. Thus, we had 4 sizes x 13 locations x 2 edge padding conditions = 104 possible targets. Every combination was represented once in the study stimuli. The order of targets was designed to evenly represent all possible transitions between target positions and sizes, and no two consecutive targets had the same size or position.

The participant's task for each discrete target trial was to tap the target drawn on the screen. All tasks were performed sitting and using the hand that the participant preferred. The participants were told to perform the tasks as naturally as they could and were allowed to hold the phone or rest it on the table as they preferred. Each touch on the screen was recorded. If a registered touch was not located within the bounds of the onscreen target, nothing happened and the participant had to try again. The software would not advance to the next target unless the user was able to register a touch within the current target. Figure 2(a) shows an example target (small). All participants were shown the targets in the same order. In our study, only one person (an adult) gave up on a target and needed help to move on.

Measures

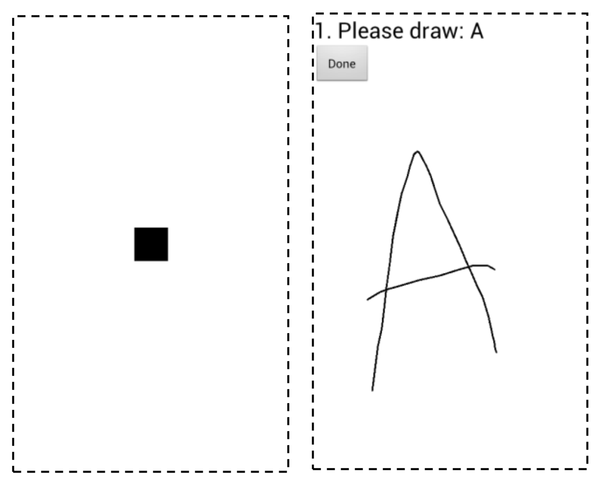
Application logs recorded the *x*-coordinate, *y*-coordinate, time, touch pressure, and touch size of each down event, *i.e.*, touch, by the participants. Additionally, other measures were calculated from these data, including (1) whether the touch occurred within the boundaries of the target, (2) the distance a touch occurred from the center point of the target (in both pixels and inches), (3) the time delay from the time of the appearance of the target to the time the user successfully touched it, and (4) the attempt number of this touch for the current target.

Phase 2: Gesture Interactions

The goal of the gesture interaction phase was to explore whether there were differences between adults and children creating free-form gestures that may create challenges in recognizing and interpreting these gestures.

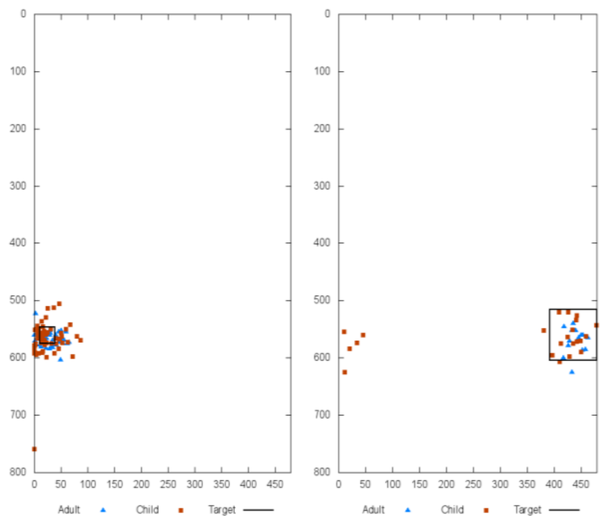
Design

In this phase of the study, participants were shown an interface screen with text indicating which gesture to make and a "Done" button. A screenshot of this interface for one gesture is shown in Figure 2(b). Users were asked to use their finger to draw gestures on the device screen and press the "Done" button when finished. The complete gesture set (20 in all) included letters (A, E, K, Q, and X), numbers (2, 4, 5, 7, and 8), symbols (line, plus, arch, arrowhead, and checkmark), and geometric shapes (circle, square/rectangle, triangle, diamond, and heart). These gestures were chosen



(a) **(b)**

Figure 2. Example of the two task interfaces. (a) Touch target task: the user was required to touch the target with his or her finger. (b) Gesture data capture interface: the prompt at the top of the screen tells the participant what symbol to draw.



(a) **(b)**

Figure 3. Holdovers from target (a) to target (b).

based on a survey of existing gesture datasets for adults [2,9,13,32], and gesture / shape generation literature from educational psychology [3]. Participants were given a paper sheet showing what each gesture should look like, in case they were not familiar with every symbol by name, but were asked to make the gesture however they normally would and not to focus on copying the gesture sample. Participants entered one example of each gesture type one after another, and repeated this five times, yielding a total of six examples of each gesture type. As participants drew each gesture, a trace appeared under their finger of the gesture, but they were not able to edit their gestures. This decision was made to prevent users from becoming overly concerned with making perfect gestures.

Measures

To aid in the understanding of users' stroke patterns, the application logs recorded the *x*-coordinate, *y*-coordinate, time, touch pressure, and touch size for each down, move, and up event as the user made the gesture. All strokes made before hitting "Done" were counted as part of one gesture and stored together. After the session, other measures were calculated from these data, such as (1) the size of the bounding box containing each gesture (height, width, area), (2) the length of the gesture, (3) the duration in milliseconds that it took the user to enter the gesture, (4) the speed at which the gesture was entered (length over time), (5) the number of points sampled during the gesture, and (6) the number of strokes made. We also ran the gestures collected in this task through a gesture recognizer, discussed in more detail later, and computed recognition accuracy for kids *vs.* adults, as well as identifying commonly confused gestures.

DATA ANALYSIS AND RESULTS

We have analyzed both tasks to understand challenges that may occur in understanding what kids meant to do in touch and gesture interaction tasks, and compare them to adults.

Phase 1: Touch Interactions

Log data from two participants were lost for the target task (1 child, 1 adult), so the analysis in this section includes only the other 28 participants (15 children, 13 adults).

Because we initially intended to examine time to acquire targets, we removed specific targets from analysis if participants took a break in the task before completing that target, or if the task software had to be re-started, inflating the time taken to touch the target. The first target for every person was counted as practice and excluded from analysis. Thus, after removing 43 attempts for these reasons, we had 4650 attempts across 28 participants and 103 targets.

Users of touchscreen devices commonly register unintentional touches due to a slip of the finger, contacting the surface with other parts of the hand, or other reasons. We coded the touch locations based on the target location, and occurrences on the very edges of the device occurred only 8 times in our dataset (all done by children). We retained them in our analysis because we expect similar behaviors to occur during use of mobile devices.

Touch Target Acquisition: Holdovers

In this task, we observed a category of touches that were located within the vicinity of the *previous* target, which we call *holdovers*. These touches occurred because the targets automatically advanced when a successful touch was registered; users often hit the screen a few more times before noting the target had changed. This behavior occurred most often on the very small targets (81% of the time), as users had the most difficulty with these targets on a variety of measures. Interestingly, children exhibited this behavior far more often than adults (96% of all holdover occurrences were children), indicating a greater challenge in discerning a touch event's intended target for kids than for adults. Figure 3 shows an example of several holdovers occurring after a "very small" target (a) advances to a "medium" target (b). As users of mobile devices ourselves, we have noted similar behavior occurs in real tasks: touchscreen devices may be slow to register a touch and so we touch again intentionally in the same location, but the interface has already moved on, causing unintentional effects. Note that our prior study [4] also found that holdovers occurred at a similar rate and contributed to a large proportion of the missed targets, since in that task the target advanced on any touch. We discuss qualitative implications of this behavior in the Discussion section.

Touch Target Acquisition: Misses

Not counting the holdovers, children's touches were not within the target on the first attempt 23.1% of the time (*misses*), whereas adults only missed 16.9% of the time. The per-user average proportion of targets missed on the first try separated by target size is shown in Figure 4. A repeated measures ANOVA was performed on the per-user proportion of first-attempt misses with a within-subjects factor of *target size* (large, medium, small, very small) and a between-subjects factor of *participant type* (adult or

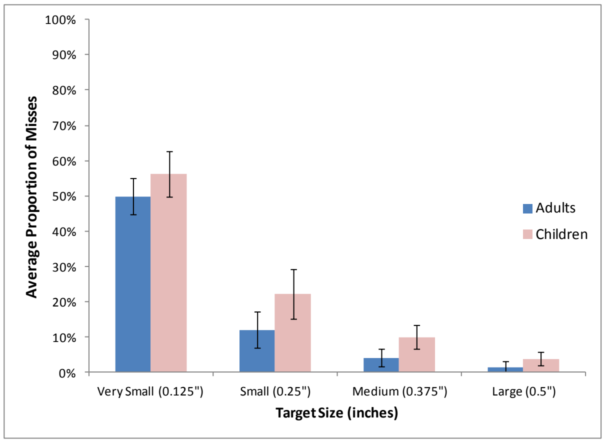


Figure 4. Average proportion of misses by target size for adults and children over all targets. Error bars show 95% confidence interval.

child). Multivariate tests indicate no significant interaction between *participant type* and *target size* ($F_{3,24}$=1.73, *n.s.*), but there are significant main effects of *target size* ($F_{3,24}$=209.6, p<0.0001) and *participant type* ($F_{1,26}$=6.57, p<0.05). Both children and adults experienced the most difficulty with the "very small" targets, and the difference in proportion of mistaken touches between adults and children was smallest with the "large" targets (though the interaction is not significant). Figure 5 shows a representative target for each size and all the touches that occurred within or outside the target for both children and adults. In these typical examples, kids make many more misses than do adults, indicating another challenge for discerning desired targets in kids' interactions.

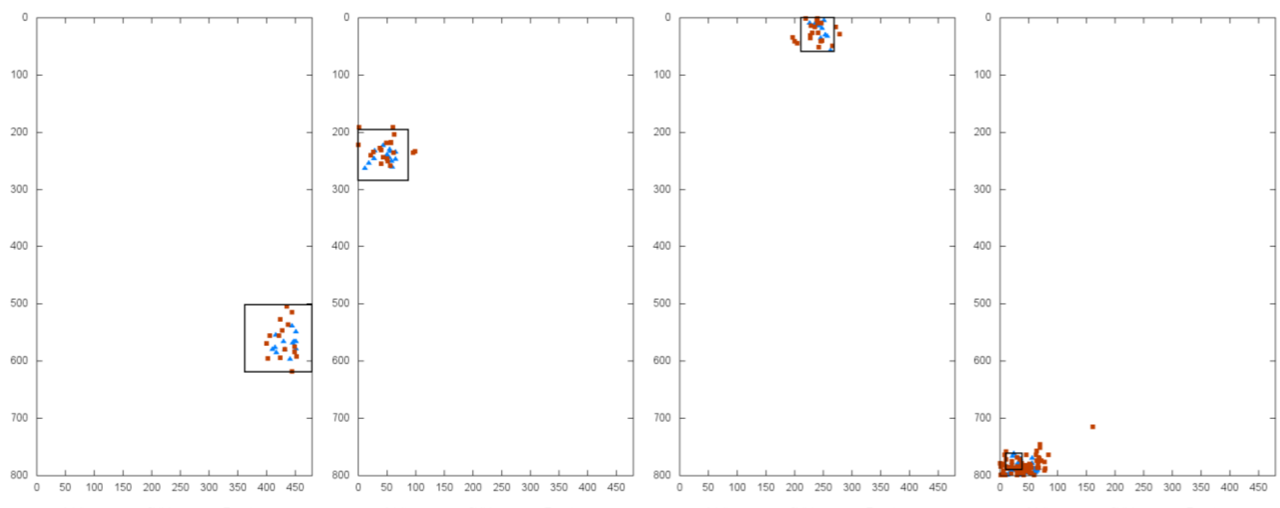
The presence or lack of edge-padding on targets also contributed to the proportion of misses per target. In the presence of edge padding, the proportion of first-attempt misses nearly doubled (children: 30.2% and adults: 22.0%) compared to no edge padding (children: 17.8% and adults: 11.9%). A repeated measures ANOVA was performed on the per-user proportion of first-attempt misses with a within-subjects factor of presence of *edge padding* (yes or no) and a between-subjects factor of *participant type* (adult or child). Again, no significant interaction was found between *participant type* and *edge padding* by multivariate tests ($F_{1,26}$=0.542, *n.s.*), but significant main effects of both *edge padding* ($F_{1,26}$=52.54, p<0.0001) and (consistent with prior tests) *participant type* ($F_{1,26}$=7.05, p<0.05) were found. Note that nearly all (99%) of the misses on edge-padded targets occurred within the edge padding "gutter" itself (the space between the target and the edge of the screen). Kids and adults both had difficulty with edge-padded targets, and the prevalence of this problem indicates we must account for this challenge during touch interaction.

Touch Features

The Android operating system can sense the size and pressure exerted during a touch event. In unpaired samples two-tailed t-tests on the per-person average touch point size and pressure, we found that children make significantly smaller touch points (t(27)=2.67, p<0.05) and exert significantly less touch pressure (t(27)=2.37, p<0.05) than adults do, replicating our earlier study's findings [4]. An implication of this finding is that apps that use touch size or pressure to change pen color or thickness (such as using Adobe Photoshop with a Wacom pressure-sensitive pen) may need to be calibrated separately for kids. As long as the touch events are registered by the hardware, these results do not point to a technical challenge for interpreting input.

Phase 2: Gesture Interactions

In total we collected 3600 gesture samples from 30 people over 6 samples of 20 gestures each. The first round of gestures was considered practice, leaving 3000 gesture



(a) (b) (c) (d)

Figure 5. Examples of miss patterns in our target touch task: (a) a "large" target with no misses; (b) a "medium" target with a few misses from children; (c) a "small" target with several misses, mostly from children; and (b) a "very small" target with many misses from both children and adults, many of which appear in the gutter between the target and the screen's edge.

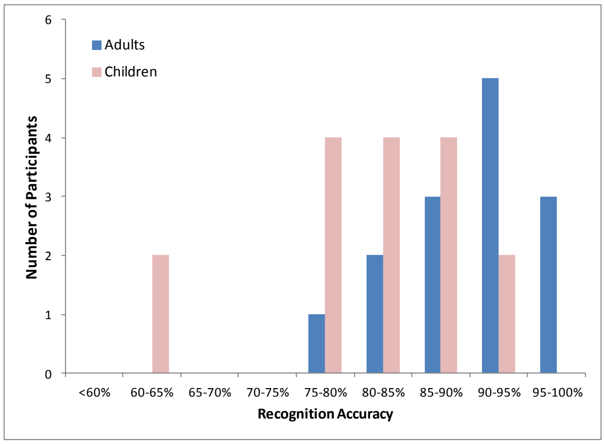


Figure 6. Histogram showing recognition accuracy distribution for gestures drawn by children and by adults.

samples for our analysis. We noted several gesture anomalies that were artifacts of the task, such as extra touches on the canvas near the "Done" button, that we could reasonably infer were not part of the user's intended gesture sample (something a real app could take into consideration through simple heuristics). In total we removed 5.8% of the samples due to such issues, leaving 2825 gestures for our analysis.

Gesture Recognition: Accuracy

Before we can act on gesture input, we must first interpret it. To determine how recognition of gestures performed by children might be different than of gestures performed by adults, we conducted recognition tests of the gestures collected in this study using the popular $N-Protractor recognizer [2]. We performed *user-dependent* testing, meaning the recognizer was trained and tested on data from the same user one at a time; final recognition accuracy rates are averaged across all children or all adults.

On average, children's gestures were recognized more poorly (mean: 81%, min: 61%, max: 92%, stdev: 9%) than adults' (mean: 90%, min: 76%, max: 98%, stdev: 6%), and this difference is significant by an unpaired samples two-tailed t-test on per-user recognition accuracy ($t(26.6)=3.18$, $p<0.01$). This result supplements the finding in our previous study [4] that *user-independent* recognition testing (*i.e.*, training the recognizer on different users than it is tested on) also shows significantly worse performance on children's gestures than on adults' gestures [4]. In this case, the effect size is smaller because user-dependent testing is in general more accurate. However, in both cases, interpretation of kids' intended gestures is more difficult.

Figure 6 shows a histogram of the number of children and adults whose gestures were recognized in bins of 5% intervals: children tend to be less accurately recognized. We theorized that recognition accuracy might be correlated with a participant's actual age. Indeed, there was a strong positive correlation between age (in years) and average recognition accuracy ($r=.59$, $p<0.01$, $N=30$).

Gesture Recognition: Confusions

When recognition of a drawn gesture is not correct, the recognizer typically returns some other gesture (rather than "[no result]"). How can we be sure the recognized gesture is what the user meant to draw? Confidence scores are used in recognition approaches to enable some sense of "degree of certainty" that the recognition result matches the intention [21]. $N [1,2] does not return a confidence score, but we can look at the confusable gestures to understand any patterns that may occur. Gestures that are commonly confused for each other should reduce our confidence that recognition results returning these gestures are correct.

Table 2 shows the top five most highly confused pairs of gestures for children and adults. Because recognition for children is lower, there are more instances of the confused pairs than for adults. It is interesting to note that in the top five, only A-to-K is common between children and adults.

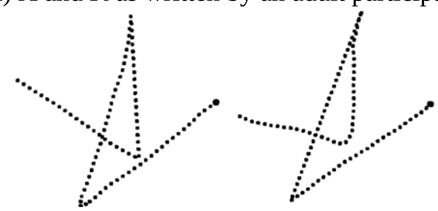
	Tested Gesture	Confused Gesture	No. of Confusions	% Tests Confused
Adults	A	K	268	6.5%
	E	[no result]	246	6.0%
	rectangle	E	195	4.8%
	arrowhead	7	184	4.5%
	diamond	triangle	167	4.1%
Children	plus	X	503	14.4%
	A	K	353	10.1%
	K	4	206	5.9%
	X	plus	203	5.8%
	triangle	diamond	191	5.5%

Table 2. Top five most highly confused gestures for children and adults. Total number of tests for adults per gesture type is 4100 and for children is 3500.

Figure 7 shows an example (for an adult) of how A and K as written (a) can be confused after $N-Protractor performs its pre-processing steps (b) used to match multistroke gestures [1,2]. Because they are confused for both adults and children, this finding indicates that gesture sets must be defined based on knowledge of what gestures are confusable by particular recognizers.

Also, for children, plus and X are confused for each other in both directions, *i.e.*, plus is often recognized as X and X is often recognized as plus, but the same is not true for adults. Based on knowledge of how the $N-Protractor recognizer works [1,2], we would expect these two symbols to be highly confusable because they are rotations of each other. In our tests, $N-Protractor was configured with a bounded rotation invariance of ±45° (meaning that it would only consider two gestures to be the same if they were within 45° of rotation from each other). For children, plus and X were often within this rotation bound, but for adults, this was not common. Figure 8 shows one child's plus and X (a)

(a) A and K as written by an adult participant.



(b) Recognizer pre-processing steps for A and K yield two gestures that it cannot distinguish.

Figure 7. Illustration of how the recognizer's pre-processing steps can create highly confusable pairs of gestures.

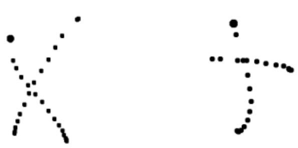
that were confused for each other by the recognizer, and one adult's (b) that were not. This finding may indicate that children do not tend to write their gestures oriented to an expected upright rotation. This factor can be used to tweak parameters for gesture recognizers and to select the best gesture set so as to avoid such rotation-induced conflicts, increasing confidence that we will know what was meant.

DISCUSSION AND DESIGN IMPLICATIONS

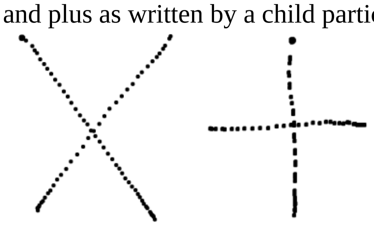
The results of this work have important implications both for designing touchscreen interfaces for children and adults and for implementing recognizers for children's gestures. The technical challenges we identify focus on understanding and interpreting what the user meant to do:

- Prevalence of *holdover* touches and *misses* create challenges for identifying the user's intended target; and,
- Low recognition accuracy on some gestures creates challenges for interpreting the user's intended gesture.

These challenges are more pronounced across the board for input from kids *vs.* adults. We now discuss implications for design and implementation that follow from our findings.



(a) X and plus as written by a child participant.



(b) X and plus as written by an adult participant.

Figure 8. Example of a pair of gestures (X and plus) that were highly confused for children but not for adults.

Touch Interactions

The target task results lead to several recommendations for designing touch interfaces for kids.

Use timing and location of touches and interface widgets to identify and ignore holdover touches (corollary: improve interface responsiveness to prevent holdovers). Consider the *holdover* touches observed for many of the targets, especially the "very small" ones. Although the holdover phenomenon of intentional touches in the location of the previous target was exhibited by both children and adults, the vast majority of occurrences (96%) were caused by children. This difference suggests that there may be cognitive processing delays preventing children from noticing as quickly that their target has activated. This phenomenon also occurs in real world use of mobile devices, from the authors' own experiences, and makes it challenging for the system to understand the user's intended action: was it an intentional or unintentional touch? If the touch was very close to the last activated target in space or in time, it is probably a holdover and can be ignored.

Use consistent, platform-recommended target sizes. Our data revealed that the "very small" targets, 0.125" square, are too small for both children and adults. However, we found that the other target sizes (0.25", 0.375", and 0.5") were large enough that on average only one attempt was needed, pointing to a possible floor effect. Because of the limited screen space on a mobile device such as a smartphone, smaller interactors allow more information to be onscreen at once. It is possible that there may be a minimum acceptable target size between 0.125" and 0.25" that has yet to be discovered. The Android design guidelines specify targets should be between 7 to 10 mm in size (0.275" to 0.39"), which is supported by our findings.

Increase active area for interface widgets to allow slightly out-of-bounds touches to count. The target size impacts the number of recorded misses for both children and adults. For the smallest targets, touches appear in a scattered distribution centered on the target itself; making this space also an active area will improve touch accuracy.

Align targets to edge of screen, or count edge touches. Our data also shows that many of the missed touches for edge-padded targets appear in the "gutter" between the target and the edge of the screen. Thus, we find Fitts' law's recommendation (for desktop pointing applications) to have interactors at screen edges also applies to mobile devices.

Gesture Interactions

The gesture task results also lead to recommendations for implementing recognition for kids.

Train age-specific recognizers to improve accuracy on kids' gestures. Recognition and interpretation of children's gestures was more difficult than of adults' gestures, and recognition accuracy was positively correlated with participant age: the recognizer had the most trouble interpreting the youngest kids' gestures. Previously we have found significant differences in the characteristics of children's and adults' gestures such as height, width, and stroke count [4]. We analyzed similar features of the gestures collected in this study and did not find any significant discriminators between children's and adults' gestures. However, we did observe higher recognition error and gesture confusion among children's input than among adults'. This data suggests that there are other characteristics, yet to be discovered, of gestures that differ between adult and child users that do affect recognition accuracy. We intend to investigate other potential feature spaces to identify these characteristics in order to implement recognition algorithms that can accommodate them. For example, measures of the total curvature, sharpness, or global orientation of the gesture [24], or movement variability within the gesture (*e.g.*, "wiggliness") [19] may be effective discriminators for children's gestures.

Design gesture sets to avoid confusions caused by recognizer limitations (corollary: train recognizers specifically to problematic pairs). Our data shows that children's writing and drawing styles can affect recognition and interpretation of what they meant to draw. We found that the recognizer's most frequently confused gestures differed for kids and for adults, both in terms of frequency and in terms of whether they were reciprocal (*i.e.*, in both directions). Since we found that recognition accuracy is related to age, we expect that, as children age and improve their fine motor skills and penmanship, not only will the recognition accuracy gap narrow, but recognizer confusions will be more similar to those for adults.

LIMITATIONS

This study represents an important next step in our research agenda of understanding how children use touch and gesture interaction differently than adults, in order to better design and implement such interfaces for kids. Building on prior work, this study expands the ages of the children and the task we have studied. This study does have some limitations, which we acknowledge here and plan to address in future work. First, this study expands on our prior study [4] by looking at older children. It may be that younger children will pose more challenges for touchscreen interfaces and recognizers, and wider recruitment is needed to find kids with varying touchscreen expertise levels. Therefore, we plan to expand to younger children. Working with younger children will require further modification of the tasks to engage young kids during the study session.

Second, we focused on a relatively narrow age range for adults (18 to 29) in this study. This choice was partly because we expect this age range of adults to have the smoothest performance compared to older demographics, and we wanted to compare children to adults for whom the biggest performance differential would result. However, based on prior work on mouse pointing performance [17], we anticipate that touchscreen interaction performance will reach a peak in early to middle adulthood and then begin declining with age. We plan to explore this relationship in future studies by including older adults, people with disabilities, and users with less touchscreen expertise.

FUTURE WORK

We previously found significant differences in other touch and gesture characteristics, as well as target accuracy and gesture recognition [4]. In this paper, we focused on understanding and interpreting what the user meant by his or her input. The visual feedback provided in this study in the gesture task may have impacted our findings, since we did not see the same gesture feature differences. Plans for future studies include comparing the task both with and without visual feedback to understand the potential effects of providing such feedback. Also, 93% of participants in this study considered themselves either average or expert users of touchscreen devices. In future studies, we plan to include users who consider themselves beginners to explore the effects of users' familiarity with touchscreen devices on gesture characteristics and touch target acquisition. To investigate possible impact of the recognition approach, we also plan to explore other recognizers, such as the Microsoft Tablet PC recognizer. Ultimately, tailored recognition may be required to confidently interpret kids' gestures.

CONCLUSION

We have presented results of a study that identifies specific challenges in understanding what kids meant to do in touch and gesture tasks on mobile touchscreen devices. We have found frequent intentional and unintentional touches outside of onscreen targets for children, and age-related challenges in recognizing children's gestures, both of which will impact the success of children's interactions. For example, children miss a greater proportion of targets than do adults, and generate a larger amount of *holdover* touches after an onscreen target has been selected. Also, gesture recognition of children's gestures is more challenging than for adults, especially for younger children. These challenges indicate a need for tailored interaction for children. The results of this work inform the types of interaction or recognition adaptations that would be necessary for children's successful interaction with mobile touchscreen devices.

ACKNOWLEDGMENTS

This work was partially supported by Department of Education HBGI Grant Award #P031B090207-11 and National Science Foundation Grant Awards #IIS-1218395/ IIS-1218664. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect these agencies' views.

REFERENCES

1. Anthony, L. and Wobbrock, J.O. A lightweight multistroke recognizer for user interface prototypes. *Proc. GI 2010*, Canadian Information Processing Society (2010), 245–252.
2. Anthony, L. and Wobbrock, J.O. $N-protractor: a fast and accurate multistroke recognizer. *Proc. GI 2012*, (2012), 117–120.

3. Beery, K., Buktenica, N., and Beery, N.A. *The Beery-Buktenica Developmental Test of Visual-Motor Integration, 5th Edition*. Modern Curriculum Press, New Jersey, 2004.
4. Brown, Q. and Anthony, L. Toward comparing the touchscreen interaction patterns of kids and adults. *ACM SIGCHI EIST Workshop*, 4pp.
5. Brown, Q., Bonsignore, E., Hatley, L., Druin, A., Walsh, G., Foss, E., Brewer, R., Hammer, J., and Golub, E. Clear Panels: a technique to design mobile application interactivity. *Proc. DIS 2010*, ACM (2010), 360–363.
6. Chiong, C. and Shuler, C. *Learning: Is there an app for that? Investigations of young children's usage and learning with mobile devices and apps*. The Joan Ganz Cooney Center at Sesame Workshop, New York, NY, 2010.
7. Donker, A. and Reitsma, P. Aiming and clicking in young children's use of the computer mouse. *Computers in Human Behavior 23*, 6 (2007), 2863–2874.
8. Guyon, I., Schomaker, L., Plamondon, R., Liberman, M., and Janet, S. UNIPEN project of on-line data exchange and recognizer benchmarks. *Proc. ICPR 1994*, IEEE (1994), 29–33.
9. Hammond, T.A., Logsdon, D., Paulson, B., Johnston, J., Peschel, J., Wolin, A., and Taele, P. A sketch recognition system for recognizing free-hand course of action diagrams. *Proc. IAAI 2010*, AAAI Press (2010), 1781–1786.
10. Harris, A., Rick, J., Bonnett, V., Yuill, N., Fleck, R., Marshall, P., and Rogers, Y. Around the table: Are multiple-touch surfaces better than single-touch for children's collaborative interactions? *Proc. CSCL 2009*, International Society of the Learning Sciences (2009), 335–344.
11. Hinrichs, U. and Carpendale, S. Gestures in the wild. *Proc. CHI 2011*, ACM (2011), 3023–3032.
12. Hourcade, J.P., Bederson, B.B., Druin, A., and Guimbretière, F. Differences in pointing task performance between preschool children and adults using mice. *ACM Transactions on Computer-Human Interaction 11*, 4 (2004), 357–386.
13. Hse, H.H. and Newton, A.R. Recognition and beautification of multi-stroke symbols in digital ink. *Computers & Graphics 29*, 4 (2005), 533–546.
14. Inkpen, K.M. Drag-and-drop versus point-and-click mouse interaction styles for children. *ACM Transactions on Computer-Human Interaction 8*, 1 (2001), 1–33.
15. Joiner, R., Messer, D., Light, P., and Littleton, K. It is best to point for young children: a comparison of children's pointing and dragging. *Computers in Human Behavior 14*, 3 (1998), 513–529.
16. Jones, T. An empirical study of children's use of computer pointing devices. *Journal of Educational Computing Research 7*, 1 (1991), 61–76.
17. Keates, S. and Trewin, S. Effect of age and Parkinson's disease on cursor positioning using a mouse. *Proc. ASSETS 2005*, ACM (2005), 68–75.
18. Li, Y. Gesture search: a tool for fast mobile data access. *Proc. UIST 2010*, ACM (2010), 87–96.
19. MacKenzie, I.S., Kauppinen, T., and Silfverberg, M. Accuracy measures for evaluating computer pointing devices. *Proc. CHI 2001*, ACM Press (2001), 9–16.
20. Norris, C. and Soloway, E. Envisioning the handheld-centric classroom. *Journal of Educational Computing Research 30*, 4 (2004), 281–294.
21. Pitrelli, J.F. and Perrone, M.P. Confidence-scoring post-processing for off-line handwritten-character recognition verification. *Proc. ICDAR 2003*, IEEE (2003), 278–282.
22. Read, J.C., MacFarlane, S., and Casey, C. Pens behaving badly-usability of pens and graphics tablets for text entry with children. *Adj. Proc. UIST 2002*, ACM Press (2002), 21–22.
23. Rick, J., Harris, A., Marshall, P., Fleck, R., Yuill, N., and Rogers, Y. Children designing together on a multi-touch tabletop: An analysis of spatial orientation and user interactions. *Proc. IDC 2009*, ACM (2009), 106–114.
24. Rubine, D. Specifying gestures by example. *SIGGRAPH Computer Graphics 25*, 4 (1991), 329–337.
25. Ryall, K., Morris, M.R., Everitt, K., Forlines, C., and Shen, C. Experiences with and observations of direct-touch tabletops. *Proc. Tabletop 2006*, IEEE (2006), 8pp.
26. Shin, N., Norris, C., and Soloway, E. Effects of handheld games on students learning in mathematics. *Proc. ICLS 2006*, International Society of the Learning Sciences (2006), 702–708.
27. Shuler, C. *Pockets of Potential: Using Mobile Technologies to Promote Children's Learning*. Joan Ganz Cooney Center at Sesame Workshop, New York, NY, 2009.
28. Vatavu, R.-D., Vogel, D., Casiez, G., and Grisoni, L. Estimating the perceived difficulty of pen gestures. *Proc. INTERACT 2011*, Springer (2011), 89–106.
29. Vogel, D. and Balakrishnan, R. Direct pen interaction with a conventional graphical user interface. *Human–Computer Interaction 25*, 4 (2010), 324–388.
30. Willems, D., Niels, R., Van Gerven, M., and Vuurpijl, L. Iconic and multi-stroke gesture recognition. *Pattern Recognition 42*, 12 (2009), 3303–3312.
31. Wobbrock, J.O., Aung, H.H., Rothrock, B., and Myers, B.A. Maximizing the guessability of symbolic input. *Ext. Abstracts CHI 2005*, ACM (2005), 1869–1872.
32. Wobbrock, J.O., Wilson, A.D., and Li, Y. Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. *Proc. UIST 2007*, ACM (2007), 159–168.