Analyzing the Articulation Features of Children's Touchscreen Gestures

Alex Shaw
Department of CISE
Gainesville, FL, USA 32611
alexshaw@ufl.edu

Lisa Anthony
Department of CISE
Gainesville, FL, USA 32611
lanthony@cise.ufl.edu

Figure 1. Circle gestures from four 8-year-old participants in the gesture corpus we used.

ABSTRACT
Children's touchscreen interaction patterns are generally quite different from those of adults. In particular, it has been established that children's gestures are recognized by existing algorithms with much lower accuracy than are adults' gestures. Previous work has qualitatively and quantitatively analyzed adults' gestures to promote improved recognition, but this has not been done for children's gestures in the same systematic manner. We present an analysis of gestures elicited from 24 children (age 5 to 10 years old) and 27 adults in which we calculate geometric, kinematic, and relative articulation features of the gestures. We examine the effect of user age on 22 different gesture features to better understand how children's gesturing abilities and behaviors differ between various age groups, and from adults. We discuss the implications of our findings and how they will contribute to creating new gesture recognition algorithms tailored specifically for children.

Categories and Subject Descriptors
• Human-centered computing~Touch screens • Human-centered computing~Gestural input • Social and professional topics~Children

General Terms
Human Factors, Measurement.

Keywords
Gesture interaction; child computer interaction; gesture articulation

1. INTRODUCTION
It has been well established in the literature that children's touchscreen interactions are quite different from those of adults [7,17,22,27]. In particular, previous work has shown that many recognition algorithms recognize children's gestures with much lower accuracy than adults' gestures [4]. In this paper, we use the term 'gesture' to refer to a series of one-finger strokes on a touchscreen to create a letter, shape, number, or other symbol [2-6,23-27]. A recent study by Woodward et al. [27] reports recognition rates as low as 64% for 5-year-old children, but in a handwriting recognition study, Read et al. [18] showed that children are not satisfied with recognition rates below 91%.

In-depth studies of the characteristics of adults' gestures have been carried out using a number of tools and features [5,23,25]. These studies offer suggestions for how to improve the design of gesture sets and gesture-based touchscreen interfaces for adults, but since they do not study the gesture articulation patterns of children, it is not clear how well the suggestions will generalize. To support better interfaces, gesture set design, and gesture recognition rates for children, more work is needed to systematically characterize the ways children make gestures across different ages. We tackle this goal by examining children's gestures based on several features.

This paper presents an analysis of gestures (a subset is shown in Figure 1) made by 24 children (ages 5 to 10 years old) and 27 adults. These gestures were collected during a user study presented by Woodward et al. [27], but that paper only reported overall recognition results on this corpus. We focus on characterizing these gestures based on two categories of features: (1) simple features, which are calculated on a single gesture, and (2) relative features, which are calculated between two gestures of the same type made by the same user, giving a measure of a user's consistency between different iterations of gestures of the same type. Our simple features are based on those presented by Anthony et al. [5], and our relative features are based on those in Vatavu et al.'s GREAT (Gesture RElative Accuracy Toolkit) [23].

By studying the way these features are correlated to user's age, we provide a systematic look at the differences in gesture articulation patterns across children of various ages, and how they are different from adults. In doing so, we highlight features that can be used as the basis to create new recognition algorithms and gesture sets that are more tailored toward children than existing ones.

2. RELATED WORK
A number of previous studies have investigated recognition and design of touchscreen gesture interactions, though most have not

ICMI '16, November 12-16, 2016, Tokyo, Japan © 2016 ACM. ISBN 978-1-4503-4556-9/16/11…$15.00

focused on children. We divide our examination of prior work into three categories: (1) general work on children's gestures, (2) gesture recognition, and (3) gesture features. We discuss prior work in the context of applicability to kids' gestures.

2.1 Children's Gestures

Much of the existing previous work on children's gestures focuses on providing insight into how designers can improve children's touchscreen experiences when using gesture-based interfaces. A study by Hiniker et al. [11] found that children respond better to gesture prompts that are designed specifically for younger users than to standard prompts. Research by Anthony et al. [2] has shown the importance of providing visual feedback for children when making gestures, and a study by McKnight and Fitton [14] demonstrated the benefits of compensating for errors that children tend to make when articulating gestures, such as unintended touches. Aziz et al. [1] examined appropriate gesture sets for children ages 2 to 4 years old, and Nacher et al. [15] offered a set of guidelines for designing multi-touch gestures for children based on a study of kids ages 2 to 3 years old. Hamza and Salivia [8,9] showed that 4 and 5 year-olds' ability to use gestures like zoom-in and drag-and-drop is affected by target size and position. We add to the existing literature by characterizing gestures elicited from children ages 5 to 10 based on a number of different geometric, kinematic, and relative accuracy features.

2.2 Gesture Recognition

Since Rubine's seminal 1991 work [19], gesture recognition has been the focus of a large body of research. Many new recognizers have been designed, but the majority of them have been designed for and tested on adults, not children [6,10,12,13,19,20,26]. A study by Anthony et al. [4] examined the recognition rates of some of these algorithms for children ages 7 to 17 years old, demonstrating that accuracy rates are lowest for the youngest children and increase for older children. Woodward et al.'s [27] study on children's touchscreen interactions confirmed this pattern for children ages 5 to 10 years old using $P, a popular multi-stroke gesture recognition algorithm [24]. However, prior work has not investigated the features of children's gestures that could explain why recognition rates are so much lower for younger children. Our work fills this gap by characterizing the gestures made by kids of various ages.

2.3 Gesture Features

In addition to work on gesture recognition, a number of researchers have investigated users' touchscreen gestures based on a variety of features. Anthony et al. [5] examined agreement rates across adult participants' gestures using clustering algorithms based on 12 geometric features, and released GECKo (GEsture Clustering toolKit) to allow for visualization of the clustering. Vatavu et al. [23] examined relative accuracy features of groups of gestures produced by adults, and released GREAT (Gesture RElative Accuracy Toolkit) to encourage using these features to characterize other data. Another tool that has been developed by Vatavu et al. [25] is GHoST (Gesture HeatmapS ToolkiT), which provides visualizations of how a number of factors, such as speed and turning angle, change during a user's execution of a gesture. These toolkits provide useful information and allow visualizations of users' articulation of touchscreen gestures, but they have not been systematically used to characterize gestures elicited from children. By using these features to understand children's gestures, we enable future work on designing new recognizers tailored toward children and on designing child-focused gesture sets.

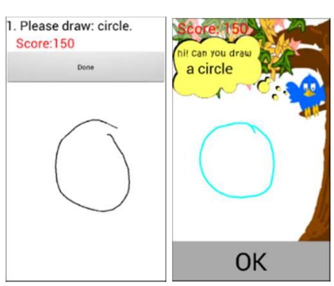

Figure 2. The gesture collection apps used for the corpus [27].

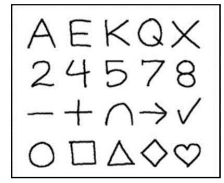

Figure 3. The gesture set from the data corpus we use [27].

3. DATA CORPUS

The gestures used in our analysis were collected as part of a study by Woodward et al. [27] which investigated the effect of interface complexity (e.g., animations, graphics, and so on) on children's touchscreen interaction. The gestures in that study were collected using two different smartphone applications – one with a simple interface and one with a more complex interface (Figure 2). The gestures produced in both cases were the same, and were based on previous work by Anthony et al. [4]. These gestures, shown in Figure 3, are common letters, numbers, signs and symbols, and the set was originally designed based on developmental education and psychological literature on children's drawing development [16].

There were a total of 60 participants in the study, including 30 adults and 30 children. The age range, 5 to 10 years, was chosen because it is a time when children undergo important psychological and developmental growth [16]. Figure 4 shows the recognition rates of the Woodward et al. study [27]. The study reported a significant effect of age on recognition accuracy, and no significant effect of interface complexity on recognition accuracy [27]. For this paper, we also ran a repeated measures ANOVA on recognition accuracy to test the interaction between age group and complexity, and found no significant interaction ($F_{6,50} = 0.455$, *n.s.*). Based on this evidence, we group the gestures produced from the two apps together for our analysis. In each application, the participant provided 6 samples of each of the 20 gesture types. Nine participants did not completely finish the study, and as such did not have complete gesture sets. We excluded these 9 participants from our study, leaving a total of 51 (24 children and 27 adults). Thus, our corpus consisted of 12,240 gestures (2 apps x 20 gesture types x 6 samples x 51 participants). The 24 children consisted of three 5-year-olds, four 6-year-olds, three 7-year-olds, four 8-year-olds, five 9-year-olds, and five 10-year-olds.

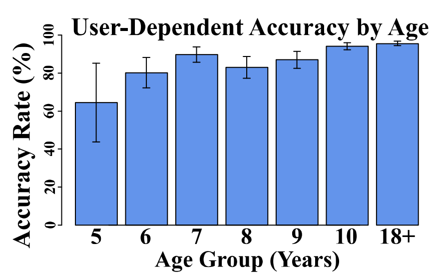

Figure 4. Recognition rates of the gestures used in our study, from Woodward et al. [27].

4. METHOD

As mentioned earlier, we group our analysis of children's gestures into two main categories: (1) simple features, and (2) relative accuracy features.

4.1 Simple Features

The simple features in our analysis are taken from a paper by Anthony et al. [5] which analyzed gestures from adults using 12 geometric and kinematic features. We implemented each of these features and calculated their values for each gesture in our corpus.

A definition of each of these features, as well as an explanation of its significance, is presented in Section 5.

4.2 Relative Accuracy Features

The relative accuracy features in our analysis are taken from a paper by Vatavu et al. [23] in which they introduce 12 relative accuracy features. These features are computed by their Gesture RElative Accuracy Toolkit, which we employ in our analysis as well.

Relative accuracy features are computed between two different examples of the same gesture. These features quantify how similar two gestures are based on the particular feature. The gesture being used as the basis for comparison is called the task axis [23]. GREAT offers two options for choosing a task axis: either the centroid of a user's gestures of a particular type can be computed, or the gesture most similar to the centroid (of that type by that person) can be used. In either case, there is only one task axis per gesture type per person. Thus, the relative accuracy features are not computed between every possible pair of gestures. Since children's gestures are more inconsistent than those of adults (as we will show), systematic analysis of a corpus with this model of GREAT might have the problem of under- or overestimating feature values based on choosing a particularly non-representative task axis.

Therefore, to get a more consistent picture of the relative accuracy features, we modified GREAT such that each possible pair of gestures of the same type from the same user are used as inputs for each of the features. Since each participant produced 6 of each gesture on two types of apps, this gives a total of $2 * \binom{6}{2} = 30$ comparisons per person per gesture. For each feature, an average was computed for each gesture of each participant. The per-gesture rates were then averaged for each user, and finally the per-user rates were averaged to get the average per-age-group. The results presented in this paper were calculated using this modified version of GREAT to find the per-age-group averages for each feature.

For all features, we computed the values of each feature on four different age groups: (1) 5 to 6-year-olds, (2) 7 to 8-year-olds, (3) 9 to 10-year-olds, and (4) adults. We grouped the users because some of the age groups had relatively low numbers of participants (e.g., three 5-year-olds), so grouping allows us to have a higher level of confidence in our results. We picked these particular groupings because they have roughly equal numbers of participants, and grouping adjacent ages together still gives a picture of how the feature is affected by the age of the participant. We graphed each feature to visualize the trend across the ages (Figure 5).

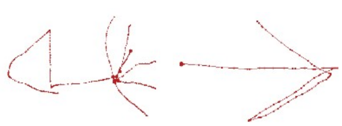

Figure 6. An arrow gesture from a 5-year-old (left) and a 10-year-old (right) in the gesture corpus we used. The 5-year-old uses a much larger number of strokes (6 strokes for 5-year-old compared to 1 for 10-year-old).

5. ANALYSIS

We now discuss our findings for each of the features in our analysis, starting with the simple features and continuing with the relative accuracy features. For each feature, we report a one-way ANOVA on the effect of age group on the features. Where significant, we report post-hoc tests with Bonferroni corrections.

5.1 Simple Features

We now provide a brief description of each simple feature in our study and discuss our findings. While Anthony et al.'s [5] study discussed 12 features, we consider only 10 of them for our analysis. The two that we exclude are *cosine of starting angle* and *cosine of ending angle*. We leave out these features because they are largely dependent on the type of gesture being executed, so averaging the value over the 20 gestures in our experiment would 'wash out' the individual differences. Future work may consider looking at these values across age groups for individual gestures in larger corpora.

Number of Strokes (Figure 5a). Average number of strokes is defined as a user's average number of pen or finger down-up events per gesture. A one-way ANOVA showed a marginal effect of age group on average number of strokes ($F_{3,47} = 2.208$, $p = 0.0996$). The average number of strokes is highest for the youngest children, and decreases for older participants. Figure 6 illustrates this pattern.

Path Length (Figure 5b). Path length refers to the sum of the distance between each adjacent pair of points in a gesture. In other words, it is a measure of the amount of ink used in creating a gesture. A one-way ANOVA showed a significant main effect of age group on average path length ($F_{3,47} = 4.124$, $p < 0.05$). Post-hoc tests found a significant difference between 5-6 year-olds and adults ($p < 0.05$). Average path length is highest for the youngest participants, and decreases for older participants. This behavior may be related to children's tendency to overtrace gestures.

Area of Bounding Box (Figure 5c). The area of the bounding box of a gesture refers to the area of the smallest box that fully encloses all points of a gesture. A one-way ANOVA showed a marginal main effect of age group on average area of bounding box ($F_{3,47} = 2.566$, $p = 0.0657$). Average area of bounding box is highest for the youngest participants, and decreases for older children. Thus, younger children tend to use a wider area of the canvas.

Line Similarity (Figure 5d). Line similarity is a measure of how similar the strokes in a user's gestures are to a straight line. The maximum possible value, 1, indicates a perfectly straight line. A one-way ANOVA showed a significant main effect of age group on average line similarity ($F_{3,47} = 5.622$, $p < 0.001$). Post-hoc tests found a significant difference between 5-6 year-olds and adults ($p < 0.05$). Average line similarity is lowest for the youngest children, and increases for older participants. Thus, younger participants tended to draw strokes that were less similar to straight lines than older participants. It is important to note, however, that this value is highly influenced by the gesture set. We see this pattern because the gesture set that was used contains many gesture types made primarily of straight lines (e.g., line, plus, X, E), but the finding may not hold for another gesture set consisting of more curved gestures. Taking this feature to the extreme and analyzing only the line gesture in our set, for example, shows the same pattern, with the youngest children having the lowest values for line similarity.

Global Orientation (Figure 5e). A gesture's global orientation is equal to the angle of the diagonal of the gesture's bounding box. A very tall and thin gesture would have a high global orientation, while a short and wide gesture would have a low one. A one-way ANOVA showed a significant main effect of age group on average

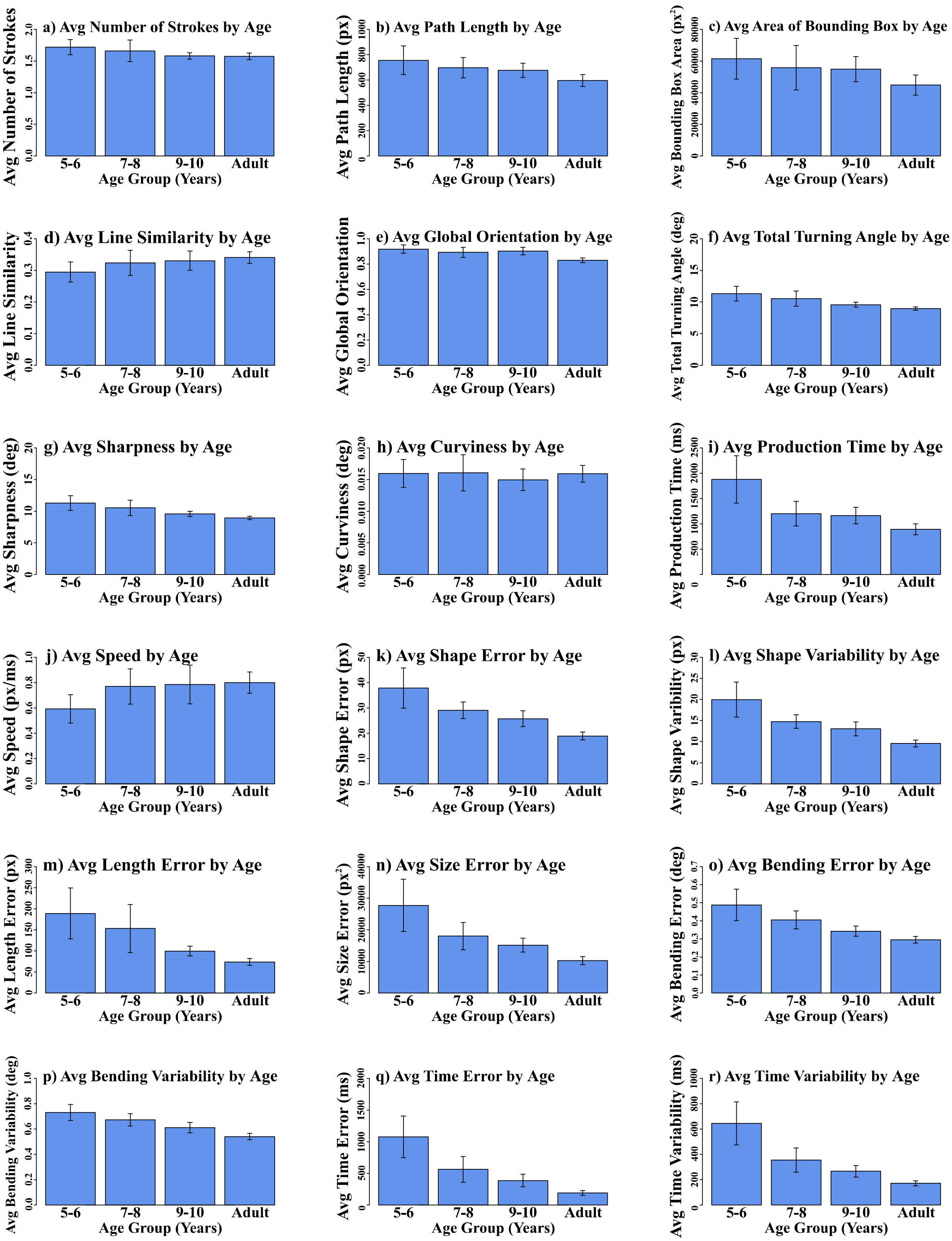

Figure 5. The effect of age group on the features examined in our analysis. Error bars are 95% confidence intervals.

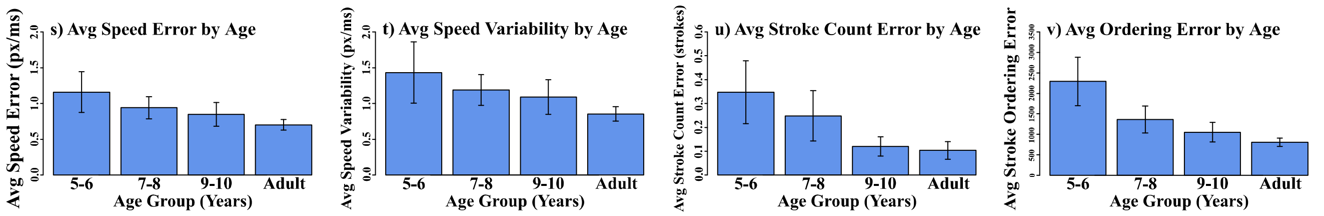

Figure 5 (continued). The effect of age on the features examined in our analysis. Error bars are 95% confidence intervals.

global orientation ($F_{3,47} = 10.14$, $p < 0.001$). Post-hoc tests found a significant difference between 5-6 year-olds and adults ($p < 0.001$), between 7-8 year-olds and adults ($p < 0.05$), and between 9-10 year-olds and adults ($p < 0.05$). Average global orientation is greatest for the youngest participants, and generally decreases for older participants. It should be noted that the values of average global orientation for the three groups of children in our analysis are very similar, but that adults have a much lower average value. This feature does not appear to be highly correlated with recognition rates. However, since previous work has shown that recognition rates after age 13 are similar to those of adults [2,3], future research may consider examining this feature in children ages 11 to 12 to see where this behavior begins to resemble that of adults.

Total Turning Angle (Figure 5f). The total turning angle of a gesture refers to the sum of the absolute value of the angle made at each point on each stroke of a gesture. A one-way ANOVA showed a significant main effect of age group on average global orientation ($F_{3,47} = 12.89$, $p < 0.001$). Post-hoc tests found a significant difference between 5-6 year-olds and 9-10 year-olds ($p < 0.05$), between 5-6 year-olds and adults ($p < 0.001$), and between 7-8 year-olds and adults ($p < 0.05$). Average total turning angle is highest for the youngest participants, and decreases for older participants. This corroborates the finding that older participants tend to draw straighter lines as seen in the line similarity feature. Younger children tend to have more (and sharper) angles in their gestures.

Sharpness (Figure 5g). The sharpness of a gesture is equal to the sum of the squares of the angles at each point of the gesture. A one-way ANOVA showed a significant main effect of age group on average sharpness ($F_{3,47} = 10.66$, $p < 0.001$). Post-hoc tests found a significant difference between 5-6 year-olds and 9-10 year-olds ($p < 0.05$), between 5-6 year-olds and adults ($p < 0.001$), and between 7-8 year-olds and adults ($p < 0.001$). The youngest children had the highest average value for sharpness, and it decreases for older participants. As with the previous feature, this behavior is likely due to younger children's tendency to have more frequent, more pronounced angles in their gestures.

Curviness (Figure 5h). The curviness of a gesture is equal to the total turning angle of the gesture divided by the path length. A one-way ANOVA showed no significant main effect of age group on average curviness ($F_{3,47} = 0.233$, *n.s.*). The value of average curviness is roughly the same across all age groups in our analysis, so it does not appear to be a good feature for discriminating across different ages. It may seem strange that curviness does not exhibit the same pattern as total turning angle and path length, but it shows that the *ratio* of the two features stays roughly the same across ages, so they both show approximately the same variation with age.

Production Time (Figure 5i). Production time of a gesture is equal to the total time taken to create it, including the time between strokes. A one-way ANOVA showed a significant main effect of age group on average production time ($F_{3,47} = 14.94$, $p < 0.001$). Post-hoc tests found a significant difference between 5-6 year-olds and 7-8 year-olds ($p < 0.05$), between 5-6 year-olds and 9-10 year-olds ($p < 0.001$), and between 5-6 year-olds and adults ($p < 0.001$). Production time is highest for the youngest age group, and decreases for older participants. We hypothesize that this is likely due to younger children's tendency to take breaks between strokes, as evidenced by the timestamps of the strokes in their gestures. This behavior causes them to take longer to produce a gesture. Table 1 shows the average amount of time between strokes for each age group, which confirms this pattern clearly.

Average Speed (Figure 5j). Average speed of a gesture is defined as path length divided by production time. A one-way ANOVA showed no significant main effect of age group on average speed ($F_{3,47} = 1.75$, *n.s.*). However, the average speed for the youngest age group is much lower than that of the other age groups, likely due to the fact that their average production time is so much higher.

5.2 Relative Accuracy Features

The relative accuracy features used in this analysis were developed by Vatavu et al. [23] in their study of relative accuracy of adult gestures, in which they introduce the Gesture REIative Accuracy Toolkit (GREAT). Each of these features is computed between two gestures. In our analysis, we computed the average of these features for every possible pair of gestures of the same type from the same person in our data set (the corpus contained 12 samples per person of each gesture type). Many of these features build on the previously discussed simple features, but focus on the ways they *differ* between two gestures.

Shape Error (Figure 5k). Shape error refers to the average deviation between two gestures based on Euclidean distance. A one-way ANOVA showed a significant main effect of age group on shape error ($F_{3,55} = 22.92$, $p < 0.001$). Post-hoc tests found a significant difference between 5-6 year-olds and 7-8 year-olds ($p < 0.05$), between 5-6 year-olds and 9-10 year-olds ($p < 0.001$), between 5-6 year-olds and adults ($p < 0.001$), between 7-8 year-olds and adults ($p < 0.001$), and between 9-10 year-olds and adults ($p < 0.05$). Shape error is highest for the youngest age group, and decreases for older participants. Thus, younger children are less consistent in the way they draw the same gesture multiple times.

Shape Variability (Figure 5l). Shape variability is equal to the standard deviation of the distances between the points of two gestures. A one-way ANOVA showed a significant main effect of age group on average shape variability ($F_{3,55} = 24.81$, $p < 0.001$). Post-hoc tests found a significant difference between 5-6 year-olds and 7-8 year-olds ($p < 0.05$), between 5-6 year-olds and 9-10 year-olds ($p < 0.001$), between 5-6 year-olds and adults ($p < 0.001$), between 7-8 year-olds and adults ($p < 0.001$), and between 9-10 year-olds and adults ($p < 0.05$). Shape variability is highest for the youngest age group, and decreases for older participants. This behavior shows that not only do younger children have a higher level of shape error, they have a larger *range* of shape errors. Thus,

Table 1. The average time between strokes of gestures (in milliseconds) for each of the age groups in our corpus.

5 to 6	7 to 8	9 to 10	Adult
1025.13	482.21	437.25	247.63

Figure 7. Two E gestures drawn by the same 5-year-old participant. The two gestures have a high value for length error due to the traceovers of the strokes in the left gesture.

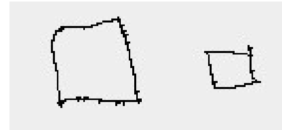

Figure 8. Two rectangle gestures drawn by the same 6-year-old, scaled uniformly. Though they are the same type, the left gesture uses more ink and has a larger bounding box.

not only are children inconsistent in the way they make shapes, they are even inconsistent about the ways they are inconsistent.

Length Error (Figure 5m). Length error is a measure of the inconsistency of lengths of strokes between two gestures. The higher the value of length error, the more inconsistent the lengths are. A one-way ANOVA showed a significant main effect of age group on length error ($F_{3,55} = 12.9$, $p < 0.001$). Post-hoc tests showed a significant difference between 5-6 year-olds and 9-10 year-olds ($p < 0.05$), between 5-6 year-olds and adults ($p < 0.001$), and between 7-8 year-olds and adults ($p < 0.05$). Length error is highest for the youngest participants, and decreases for older participants. Younger kids, therefore, have more inconsistency in the amount of ink used in their gestures. This behavior could be due to children's tendency to vary their gesture by, for example, writing in block letters, embellishing their gestures with curly tails, or tracing over their strokes. Figure 7 shows an example of two gestures from the same participant with a high length error.

Size Error (Figure 5n). Size error is a measure of the inconsistency between the areas of the bounding boxes of two gestures. The higher the value of size error, the more inconsistent the areas of the bounding boxes. A one-way ANOVA showed a significant main effect of age group on size error ($F_{3,55} = 18.56$, $p < 0.001$). Post-hoc tests showed a significant difference between 5-6 year-olds and 7-8 year-olds ($p < 0.05$), between 5-6 year-olds and 9-10 year-olds ($p < 0.001$), between 5-6 year-olds and adults ($p < 0.001$), and between 7-8 year-olds and adults ($p < 0.05$). Size error is the highest for the youngest participants, and decreases for older participants. Thus, there is a greater discrepancy between the sizes of gestures of the same type elicited from younger participants than from older participants. Figure 8 shows an example of two gestures from the same participant with a high size error.

Bending Error (Figure 5o). Bending error refers to the average of differences between corresponding turning angles of two gestures of the same type. A one-way ANOVA showed a significant main effect of age group on bending error ($F_{3,55} = 18.17$, $p < 0.001$). Post-hoc tests showed a significant difference between 5-6 year-olds and 9-10 year-olds ($p < 0.001$), between 5-6 year-olds and adults ($p < 0.001$), and between 7-8 year-olds and adults ($p < 0.001$). Bending error is highest for the youngest age group, and decreases for older participants. Young children are less consistent in the angles made during the articulation of their gestures, indicating that they often do not take the same path when drawing a gesture multiple times.

Bending Variability (Figure 5p). Bending variability refers to the standard deviation of differences between corresponding turning angles of two gestures of the same type. A one-way ANOVA showed a significant main effect of age group on bending variability ($F_{3,55} = 18.78$, $p < 0.001$). Post-hoc tests showed a significant difference between 5-6 year-olds and 7-8 year-olds ($p < 0.001$), between 5-6 year-olds and 9-10 year-olds ($p < 0.001$), between 5-6 year-olds and adults ($p < 0.001$), and between 7-8 year-olds and adults ($p < 0.001$). Bending variability is highest for the youngest participants and decreases for older participants. Thus, younger children tend to not only have the highest average difference in corresponding angles on their gesture articulation path, but they also tend to have a wider variety of values.

Time Error (Figure 5q). Time error refers to the difference in the amount of time taken to articulate two gestures. A one-way ANOVA showed a significant main effect of age group on time error ($F_{3,55} = 29.44$, $p < 0.001$). Post-hoc tests showed a significant difference between 5-6 year-olds and 7-8 year-olds ($p < 0.001$), between 5-6 year-olds and 9-10 year-olds ($p < 0.001$), between 5-6 year-olds and adults ($p < 0.001$), and between 7-8 year-olds and adults ($p < 0.001$). Time error is the highest for the youngest group of participants and decreases for older participants. Younger children have a much larger average discrepancy in the time taken to produce different samples of the same type of gesture.

Time Variability (Figure 5r). Time variability refers to the standard deviation of the differences of the timestamps of each individual point in a gesture. A one-way ANOVA showed a significant main effect of age group on time variability ($F_{3,55} = 33.05$, $p < 0.001$). Post-hoc tests showed a significant difference between 5-6 year-olds and 7-8 year-olds ($p < 0.001$), between 5-6 year-olds and 9-10 year-olds ($p < 0.001$), between 5-6 year-olds and adults ($p < 0.001$), and between 7-8 year-olds and adults ($p < 0.001$). Time variability is the highest for the youngest group of participants and decreases for older participants. Thus, not only do younger children tend to have the most inconsistency in the overall amount of time taken to produce a gesture, they also show the most inconsistency in the amount of time it takes them to articulate each individual point along the path of the gesture.

Speed Error (Figure 5s). Speed error refers to the difference in speed of production of two gestures. A one-way ANOVA showed a significant main effect of age group on speed error ($F_{3,55} = 7.289$, $p < 0.001$). Post-hoc tests showed a significant difference between 5-6 year-olds and adults ($p < 0.001$). Speed error is highest for the youngest age group, and decreases for older participants. There is more variation in the speeds of children's gestures than adults'.

Speed Variability (Figure 5t). Speed variability refers to the standard deviation of differences in the speed of production of two gestures. A one-way ANOVA showed a significant main effect of age group on speed variability ($F_{3,55} = 6.02$, $p < 0.001$). Post-hoc tests showed a significant difference between 5-6 year-olds and adults ($p < 0.05$). Speed error is highest for the youngest age group, and decreases for older participants. This is consistent with our findings for time variability, and it should also be noted that the calculation of speed variability depends on speed error.

Stroke Count Error (Figure 5u). Stroke count error refers to the difference in number of strokes of two gestures of the same type. A one-way ANOVA showed a significant main effect of age group on stroke count error ($F_{3,55} = 9.94$, $p < 0.001$). Post-hoc tests showed a significant difference between 5-6 year-olds and 9-10 year-olds ($p < 0.05$), between 5-6 year-olds and adults ($p < 0.001$), and between 7-8 year-olds and adults ($p < 0.05$). Stroke count error is highest for the youngest children, and decreases for older users. Younger children tend to have more variation in the number of strokes used.

Stroke Ordering Error (Figure 5v). Stroke ordering error is a measure of the inconsistency in the order that different strokes of a gesture are drawn between two samples of that gesture. A one-way ANOVA showed a significant main effect of age group on stroke ordering error ($F_{3,55} = 22.71$, $p < 0.001$). Post-hoc tests showed a significant difference between 5-6 year-olds and 7-8 year-olds ($p < 0.001$), between 5-6 year-olds and 9-10 year-olds ($p < 0.001$), between 5-6 year-olds and adults ($p < 0.001$), and between 7-8 year-olds and adults ($p < 0.05$). Stroke ordering error is highest for the youngest age group, and decreases for older participants. Younger children tend to have more variation in the ordering of their strokes.

6. DISCUSSION

Our analysis supplements the existing body of work on understanding gestures by adding an empirical analysis of gestures produced by children and how they can be characterized based on a number of different features. These features have previously been used to characterize adults' gestures, but not those of children, so our work offers new insight into the ways in which children make gestures. We now discuss the implications of the findings of our analyses, particularly in the context of how they can enable improved gesture set design and recognition for younger children.

6.1 Simple vs. Relative Features

In general, relative accuracy features appear to be much more discriminative among different age groups than simple geometric or kinematic features. In our analysis, all 12 of the relative accuracy features showed a significant main effect of age group on the value of the features, but only 6 out of 10 of the simple features did. Relative features also showed more post-hoc differences by age. This finding indicates that the discrepancy in recognition rates between children's and adults' gestures is not a function of individual gestures in isolation, but rather of the way different age groups show different levels of consistency. All of the relative features show the same general pattern as recognition rates for those age groups [27], indicating they may be strongly associated with recognition accuracy.

6.2 Time Features

Several of the features in our analysis examined the amount of time taken to produce a gesture (production time, time error, time variability, speed error, and speed variability). All of these features show a much higher value for the youngest age group. Even though the gestures in our analysis are fairly commonplace and simple shapes that they are likely to be exposed to in school, younger children still took a much longer time, on average, to produce them. Some of them even asked for guidance on how to draw some of the symbols, such as the diamond, saying they had not yet learned them [Woodward, personal correspondence]. Thus, gesture set designers should carefully consider the complexity of the gestures they use, aiming to make them intuitive and close to shapes and symbols that children of the target age are likely to have encountered frequently before. In gesture-based applications, children should be expected to have highly variable input times for gestures, with much longer overall average times. This is an especially important consideration for applications such as games where time is an important factor.

Designers should keep in mind, too, that children are likely to take longer breaks between strokes than adults, so the number of strokes in gestures should be kept low. This finding also has implications for real-time gesture recognition in which segmentation of strokes into individual gestures is necessary. Previous work has suggested time-based thresholds to segment characters [21], but based on Table 1, we show that it would not be possible to use the same threshold for adults as for younger children.

6.3 Distance & Size Features

Many of the features in our analysis are related to the amount of ink used (total distance of strokes) and the size of a gesture (based on the bounding box). All of these features indicate more inconsistency in younger participants as well as bigger gestures and longer strokes for younger participants. The wide variation seen in the sizes of kids' gestures indicates the importance of scale-invariance in template matching gesture recognition approaches. Furthermore, in applications with gesture-based interfaces, designers should account for the fact that children's gestures may extend out of the expected area. The fact that younger children tend to have so much variety in the amount of ink in their gestures indicates that template matching gesture recognizers will not work as well for kids as for adults, since there is a large variation in the way gestures are drawn. A recognizer that could detect common idiosyncrasies of children's gestures, like traceovers of strokes, could potentially account for these issues and improve recognition.

6.4 Impact on Gesture Recognition

Many of the features examined in our analysis could have an impact on gesture recognition. We specifically discuss features that could impact $P [24], the template-based recognizer used on these gestures by Woodward et al. [27]. Because $P is size-invariant, features such as size error, size variability, and area of bounding box would likely not impact recognition rates. The features that would impact recognition rates are those that describe the articulation path taken by the user, including such features as bending error and variability, average path length, and global orientation. If we could characterize the user's articulation path using different features that *are* consistent for children, we might improve recognition rates for children. We have also shown that children are inconsistent about the ways in which they are inconsistent, indicating the potential for further work in developing new features for kids' gestures that do show better consistency. Future work can capitalize on these new features to design new recognition algorithms.

LIMITATIONS

Our analysis expands previous work by characterizing children's gestures based on a number of different features. Despite this, there are some limitations to our work. First, the gestures were collected in a laboratory setting, which may not be generalizable to natural circumstances of use. Another limitation is the fact that the children in the original study were all in the age range of 5 to 10 years old, so it is possible that these findings may not generalize to other children of different age groups. Future work may consider characterizing gestures of other age groups to get a fuller picture of how children's gestures differ across ages. Finally, the gestures in this corpus were collected using a smartphone, so the findings may not generalize to devices with larger screen sizes. In particular, due to the relationship that we found between size features and children's gesture inconsistencies, more work is needed to see if these findings are also indicative of children's gesture interaction patterns with larger screens.

7. CONCLUSION AND FUTURE WORK

We presented an empirical analysis of the gestures of 24 children and 27 adults based on 22 different features. Of these 22 features, 10 were simple geometric or kinematic features, while 12 were relative accuracy features. We examined the effect of age on each of these features to better understand the ways in which children make gestures and how future work can focus on improving touchscreen gesture interactions for children. We found that 6 of 10 simple features and all 12 relative features showed a significant main effect of age on the value of the feature. By providing a systematic characterization of the ways different ages of users make gestures, we set the stage for further work in creating child-centered gesture recognition algorithms. We show that some of the traditional features employed in gesture recognition, such as distance between points and shape error, are subject to high variability in children, so they are not ideal for recognizing children's gestures. Our analysis of how simple and relative accuracy features vary with age will help designers in accounting for these inconsistencies when creating child-centered algorithms.

8. ACKNOWLEDGEMENTS

This work uses images from a previous work (Figures 2, 3): Julia Woodward, Alex Shaw, Annie Luc, Brittany Craig, Juthika Das, Phillip Hall, Jr., Akshay Holla, Germaine Irwin, Danielle Sikich, Quincy Brown, and Lisa Anthony. 2016. Characterizing How Interface Complexity Affects Children's Touchscreen Interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1921-1933. DOI: http://dx.doi.org/10.1145/2858036.2858200 © 2016 Association for Computing Machinery, Inc. Reprinted by permission.

This work is partially supported by National Science Foundation Grant Award #IIS-1433228. Opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect these agencies' views.

9. REFERENCES

1. Nor Azah Abdul Aziz, Firat Batmaz, Roger Stone, and Paul Wai Hing Chung. 2013. Selection of touch gestures for children's applications. *Proceedings of SAI*, IEEE, 721–726.
2. Lisa Anthony, Quincy Brown, Jaye Nias, and Berthel Tate. 2013. Examining the need for visual feedback during gesture interaction on mobile touchscreen devices for kids. *Proceedings of IDC*, ACM Press, 157–164.
3. Lisa Anthony, Quincy Brown, Jaye Nias, and Berthel Tate. 2015. Children (and adults) benefit from visual feedback during gesture interaction on mobile touchscreen devices. *IJCCI* 6, 17–27.
4. Lisa Anthony, Quincy Brown, Berthel Tate, Jaye Nias, Robin Brewer, and Germaine Irwin. 2014. Designing smarter touch-based interfaces for educational contexts. *JPUC* 18, 6, 1471–1483.
5. Lisa Anthony, Radu-Daniel Vatavu, and Jacob O Wobbrock. 2013. Understanding the consistency of users' pen and finger stroke gesture articulation. *Proceedings of GI*, Canadian Information Processing Society, 87–94.
6. Lisa Anthony and Jacob O. Wobbrock. 2012. $N-protractor: a fast and accurate multistroke recognizer. *Proceedings of GI*, Canadian Information Processing Society, 117–120.
7. Ahmed Sabbir Arif and Cristina Sylla. 2013. A comparative evaluation of touch and pen gestures for adult and child users. *Proceedings of IDC*, ACM Press, 392–395.
8. Zainab Hamza. 2014. Study of Touch Gesture Performance by Four and Five Year-Old Children: Point-and-Touch , Drag-and-Drop, Zoom-in and Zoom-out, and Rotate. Master's Thesis. Minnesota State University - Mankato.
9. Zainab Hamza and Guarionex Salivia. 2015. Study of Touch Gestures Performance in Touch Devices by Young Children. *IJRITCC* 3, 3, 1395–1400.
10. J. Herold and Thomas F. Stahovich. 2012. The 1¢ Recognizer: a fast, accurate, and easy-to-implement handwritten gesture recognition technique. *Proceedings of SBIM*, Eurographics Association, 39–46.
11. Alexis Hiniker, Kiley Sobel, Sungsoo Ray Hong, Hyewon Suh, India Irish, Daniella Kim, and Julie A. Kientz. 2015. Touchscreen prompts for preschoolers: designing developmentally appropriate techniques for teaching young children to perform gestures. *Proceedings of IDC*, ACM Press, 109–118.
12. Levent Burak Kara and Thomas F. Stahovich. 2005. An image-based, trainable symbol recognizer for hand-drawn sketches. *Computers & Graphics* 29, 4, 501–517.
13. WeeSan Lee, Levent Burak Kara, and Thomas F. Stahovich. 2007. An efficient graph-based recognizer for hand-drawn symbols. *Computers & Graphics* 31, 4, 554–567.
14. Lorna McKnight and Daniel Fitton. 2010. Touch-screen technology for children. *Proceedings of IDC*, ACM Press, 238–241.
15. Vicente Nacher, Javier Jaen, Elena Navarro, Alejandro Catala, and Pascual Gonzalez. 2015. Multi-touch gestures for pre-kindergarten children. *IJHCS* 73, 37–51.
16. Jean Piaget. 1983. *Piaget's Theory*. Wiley & Sons, New York, New York, USA.
17. Janet C. Read, Stuart MacFarlane, and Chris Casey. 2003. What's going on?: discovering what children understand about handwriting recognition interfaces. *Proceedings of IDC*, ACM Press, 135-140.
18. Janet C Read, Stuart MacFarlane, and Chris Casey. 2003. "Good enough for what?": acceptance of handwriting recognition errors by child users. *Proceedings of IDC*, ACM Press, 155.
19. Dean Rubine. 1991. Specifying gestures by example. *ACM SIGGRAPH Computer Graphics* 25, 4, 329–337.
20. Tevfik Metin Sezgin and Randall Davis. 2005. HMM-based efficient sketch recognition. *Proceedings of IUI*, ACM Press, 281–283.
21. Steve Smithies, Kevin Novins, and James Arvo. 1999. A handwriting-based equation editor. *Proceedings of GI*, Canadian Information Process Society, 84–91.
22. Tiffany Y. Tang, Maldin Yifan He, and Vince Lineng Cao. 2016. "One Doesn't Fit All": A Comparative Study of Various Finger Gesture Interaction Methods. *Proceedings of DUXU*, Springer International Publishing, 88-97.
23. Radu-Daniel Vatavu, Lisa Anthony, and Jacob O Wobbrock. 2013. Relative accuracy measures for stroke gestures. *Proceedings of ICMI*, ACM Press, 279–286.
24. Radu-Daniel Vatavu, Lisa Anthony, and Jacob O. Wobbrock. 2012. Gestures as point clouds: a $P recognizer for user interface prototypes. *Proceedings of ICMI*, ACM Press, 273–280.
25. Radu-Daniel Vatavu, Lisa Anthony, and Jacob O. Wobbrock. 2014. Gesture Heatmaps: Understanding gesture performance with colorful visualizations. *Proceedings of ICMI*, ACM Press, 172–179.
26. Jacob O. Wobbrock, Andrew D. Wilson, and Yang Li. 2007. Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. *Proceedings of UIST*, ACM Press, 159–168.
27. Julia Woodward, Alex Shaw, Annie Luc, Brittany Craig, Juthika Das, Phillip Hall Jr., Akshay Holla, Germaine Irwin, Danielle Sikich, Quincy Brown, and Lisa Anthony. 2016. Characterizing How Interface Complexity Affects Children's Touchscreen Interactions. *Proceedings of CHI*, ACM Press, 1921-1933.