Human-Centered Recognition of Children's Touchscreen Gestures

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

ABSTRACT

Touchscreen gestures are an important method of interaction for both children and adults. Automated recognition algorithms are able to recognize adults' gestures quite well, but recognition rates for children are much lower. My PhD thesis focuses on analyzing children's touchscreen gestures and using the information gained to develop new, child-centered recognition approaches that can recognize children's gestures with higher accuracy than existing algorithms. This paper describes past and ongoing work toward this end and outlines the next steps in my PhD work.

CCS CONCEPTS

• **Human-centered computing→touch screens • Human-centered computing→gestural input**

KEYWORDS

Gesture; stroke; recognition; children; beautification; touchscreen

ACM Reference format:

A. Shaw. 2017. Human-Centered Recognition of Children's Touchscreen Gestures. In *Proceedings of the 19th International Conference on Multimodal Interaction, November 2017 (ICMI ' 17).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3136755.3137033

1 INTRODUCTION

The pervasiveness of mobile touchscreen devices in recent years has created an environment in which most children are regularly using touchscreen devices [14,17]. Thus, researchers have noted the importance of understanding how children interact with these devices and how their interactions compare to those of adults [5,11,19]. Gesture interactions, in particular, differ widely between adults and children, leading to large gaps in automated recognition rates between children and adults [4,5,30]. In this paper, the term 'gesture' refers to a series of one-finger strokes on a touchscreen to create a letter, number, shape or symbol [5,6,8,23,25,30]. In my prior work, we reported 94% recognition accuracy for adults' gestures compared to 84% for children ages 5 to 10 [30]. Furthermore, that study reported a large discrepancy between recognition rates of gestures produced by younger children compared to older children. Recognition accuracy of the youngest children's (5-year-olds) gestures was the lowest, at 64%. Fig. 1 illustrates the recognition accuracy by age in the study. Prior work in the domain of handwriting recognition has found that children are not satisfied with less than 91% accuracy [21], so it is clear that current recognition rates are not good enough, especially for the youngest children. Similar work for adults found that they were less tolerant of errors than children, accepting no less than 97% accuracy [18]. My PhD work focuses on understanding the reasons for the large discrepancy in recognition rates between adults' and children's gestures and developing new approaches that can better recognize children's gestures.

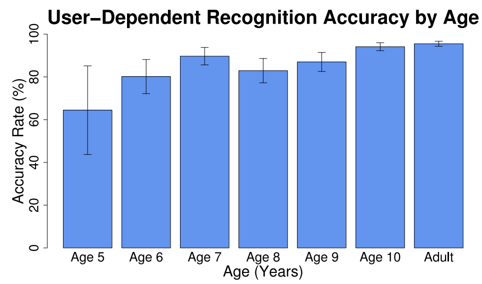

Figure 1. Recognition rates from our prior work [23,30]. Error bars represent the 95% confidence interval.

ICMI '17, November 13-17, 2017, Glasgow, United Kingdom

ACM ISBN 978-1-4503-5543-8/17/11…$15.00
http://dx.doi.org/10.1145/3136755.3137033

2 RELATED WORK

Prior work in analyzing touchscreen gestures has primarily focused on either feature-based analysis of the gestures or automated recognition. The feature-based analyses often help motivate the creation of new and improved recognition algorithms. I provide a brief overview of the work done in these areas, including work involving children's gestures.

2.1 Gesture Features and Analyses

Features of touchscreen gestures can be defined as measures or metrics that provide quantitative information about the gesture [7,22,26]. These features can fall into a number of different categories, such as, for example, size-based features (e.g., length or height) or time-based features (e.g., production time or average velocity). One of the first sets of features to be introduced was Rubine's [22] set of 13 features used for recognition, all of which were time- or distance-based. Rubine's features include, for example, the sine and cosine of the gesture's initial angle and the distance between the gesture's starting point and ending point.

The features presented by Rubine [22], as well as several others, were used as a basis by Anthony et al. [7] to analyze the differences between pen and finger gestures produced by adults as well as children ages 11 to 17. The feature set included geometric features, such as number of strokes and sharpness, and kinematic features, such as production time and average speed. Anthony et al.'s [7] paper also introduced GECKo (Gesture Clustering toolKit), a tool for visualizing similarity among stroke gestures using feature-based clustering.

Vatavu et al.'s [26] work on gesture features introduced a set of relative accuracy features, which quantify the differences between pairs of the same gesture type. Examples of these relative accuracy features include shape error (average difference in distance between corresponding points) and time error (average difference in articulation time of the two gestures). These features are therefore useful in quantifying the amount of variation within and between users' gestures of the same type. The paper used these relative accuracy features to investigate gestures from five publicly available datasets [8,13,27–29] (none from children).

Investigating and understanding features of children's touchscreen gestures is critical for improving gesture recognition since many recognizers use these features to compare template and candidate gestures when performing recognition tests [9,10,25,29]. Therefore, feature-based analysis is an important component of continued work in gesture recognition.

2.2 Gesture Recognition

Since the publication of Rubine's [22] 1991 paper on the GRANDMA recognizer, gesture recognition has become a widely studied topic. My work focuses primarily on template matching approaches to recognition, which classify gestures by comparing them to other pre-selected gestures called templates. Different template matchers use different methods of comparing the gestures; a common approach is to use Euclidean distance [9,25,29]. Template matchers may use multiple metrics to compare gestures [10,22]. Template matchers have high usability due to their relative ease of implementation compared to other types of recognizers like neural networks or Hidden Markov Models. Despite their simplicity, template matchers are able to achieve high recognition rates for adults' gestures with few training examples [30]. Thus, I focus on improving the performance of template matching recognition approaches for children.

Because I am primarily interested in recognition of children's gestures, I focus my discussion on studies including children. Anthony et al. [5] found that recognition accuracy of gestures produced by children ages 7 to 16 (81%) was lower than the recognition accuracy of gestures produced by adults (90%) using the $N-Protractor [9] recognizer. My work [30] used $N-Protractor's successor, $P [25], and found a recognition accuracy of 94% for adults and 84% for children.

Another study examining children's gestures was Kim et al.'s [16] developmental gesture classifier. The system did not attempt to recognize the gestures, but instead classified the developmental level of the children, as well as their gender, based on their gestures. The authors report 91% accuracy in classifying the children's ages and 71% accuracy in classifying their gender.

The small body of work, as well as the low recognition rates for children's gestures and the prevalence of touchscreen use in children, points to an important area for continued research. My PhD thesis work helps advance our understanding of children's gestures, leading to improved recognition rates.

3 COMPLETED WORK

In this section, I discuss some of the work that I have carried out thus far to motivate improved recognition of children's gestures.

3.1 Analysis of Articulation Features

To examine the quantitative differences between children's and adults' touchscreen gestures, we performed a study in which we compared both simple and complex features of gestures produced by adults and by children [23]. The simple features, adapted from prior work by Anthony et al. [7], included 10 geometric and kinematic features. Examples of these features included number of strokes, path length, and production time. The complex features included all 12 features from Vatavu et al.'s [26] relative accuracy features, such as length error (average difference in length of gestures of the same type) and time error (average difference in production time of gestures of the same type).

After selecting the features, we computed the average value of each feature for four age groups: 5- to 6-year-olds, 7- to 8-year-olds, 9- to 10-year-olds, and adults (18+ year-olds). For each of the 22 features we computed, we ran a one-way ANOVA on the value of the feature with a between-subjects factor of age group. These tests found a significant effect of age group on 6 of the 10 simple features and all 12 relative accuracy features. In particular, we found that some of the features commonly used in recognizing gestures, such as shape error and distance between corresponding points, were highly variable in children's gestures, indicating that they are not ideal for use in recognizing children's gestures.

3.2 Establishing a Target Accuracy through Human Recognition

To help direct future work, it is useful to have an established target accuracy that new recognizers can aim to achieve. Read et al.'s [21] study shows the level of accuracy that children find acceptable (91%), but obtaining this level of accuracy may not be practical. To get a better idea of what accuracy levels we should aim for, we conducted an experiment [24] wherein human participants were asked to classify the gestures in our dataset [30], enabling us to compare machine and human recognition. We employed human computation, described by Luis von Ahn as "a paradigm for utilizing human processing power to solve problems that computers cannot yet solve" [1]. Computers can partially solve the problem of recognition, but they cannot recognize gestures with perfect accuracy. We compared human and

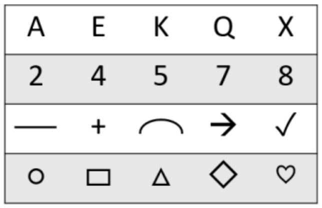

Figure 2. Gesture set from the data corpus we use [30].

computer recognition to get a reasonable target for accuracy in future work. Our recognition study was carried out using Amazon's Mechanical Turk [2] crowdsourcing tool. We had a total of 131 participants in the study. The participants each took an online survey in which they were asked to classify up to 120 different examples of gestures produced by children ages 5 to 10. The gesture types were those from the set depicted in Fig. 2; the participants were asked to choose which of the 20 gesture types the gesture they were presented with most resembled. We calculated accuracy rates by labelling each gesture as either correct (if the majority of participants correctly identified it) or incorrect (if the majority of participants incorrectly identified it). We then computed a per-writer accuracy rate and used these rates to compute overall accuracy as well as accuracy by age groups.

Our human recognition experiments found that humans were able to recognize gestures more accurately than a machine recognizer. Fig. 3 shows the accuracy rates by age for humans compared to machine recognition. On average, children's gestures were recognized with 90.60% accuracy by adults, significantly above the 84.02% accuracy of machine recognition, according to a repeated-measures ANOVA ($F_{1,20} = 42.197$, $p < 0.001$). Even for the youngest children (5-year-olds), the human recognition accuracy of 76.82% was higher than the machine recognition accuracy of 64.02%. Our work on human recognition establishes an empirical threshold by which we can judge our future work on improving automated recognition rates.

4 CONTINUED WORK

My work thus far has been focused on improving our characterization and understanding of children's touchscreen stroke gestures. Building on this work, my future plans for my PhD work include investigating new recognition techniques and eventually developing a new, child-centered recognizer.

4.1 Improving Recognition with Beautification

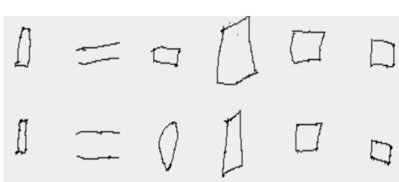

Figure 4. Example of "rectangle" gestures from different 6-year-olds.

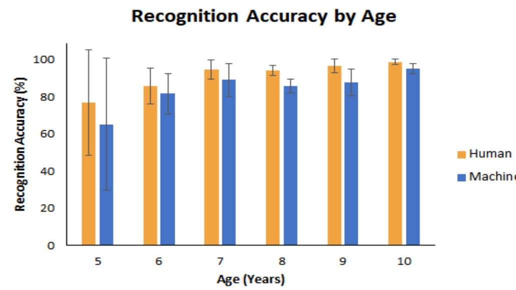

Figure 3. Effect of age and recognizer (human vs. machine) on recognition accuracy. Error bars are the 95% confidence interval.

Plotting the gestures produced by children shows wide variation in the ways gestures are drawn, as demonstrated in Fig. 4's depiction of rectangle gestures from several 6-year-olds. However, most gestures share in common that they can be broken down into a number of simpler units. The "X" gesture in Fig. 5, for example, can be represented as two intersecting lines. Sketching systems like SATIN [12] and PaleoSketch [20] have used such primitives to transform sketches into beautified, less noisy data. Uniform application of beautification can help noisy gestures to be more consistent in their appearance [13]. Thus, point-matching template-based gesture recognition approaches could potentially benefit greatly from the application of beautification as a preprocessing step.

Fig. 5 shows examples of gestures before and after a beautification process I am developing. My beautification algorithm first analyzes each stroke of the gesture and determines if they should be straight lines or curves based on the curviness of the stroke. If the stroke's curviness is below a threshold, it is snapped to a vertical, horizontal, or diagonal line depending on its slope. If the stroke is a curve, it is snapped to either an arc or a circle depending on its angle of curvature. After this initial transformation, the algorithm examines distance between the midpoints of the transformed strokes and determines whether the strokes should meet at their midpoints, adjusting them if they should. The beautification technique is applied purely based on the shape of each stroke, not the gesture type (that is, the beautification is agnostic to the gesture set being used). A preliminary application of beautification techniques to the gestures in our corpus produced by 5- to 7-year-olds improved recognition accuracy by approximately 10%. More complex gestures require accurate corner detection to be segmented correctly; I am developing a method of detecting corners that works well for children's gestures. A challenge my current beautification algorithm faces is that complex gestures like "heart" and "8" do not easily break down into the primitives used by the algorithm. My continued work will focus on improving beautification of these more complex gestures and improving the accuracy with which the beautified gestures are recognized.

While the primary purpose of beautifying the gestures in our set is to improve recognition rates, children may also feel that "live" beautification of gestures (as they produce them) improves their experience when using gesture-based applications. Prior

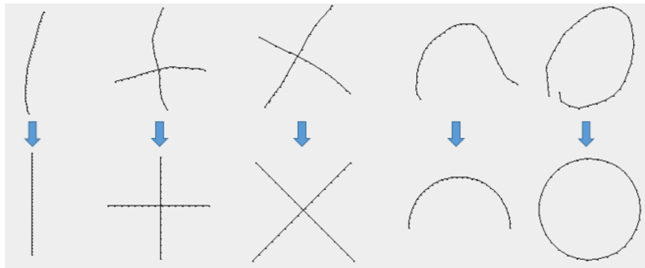

Figure 5. Illustration of gestures before (top) and after (bottom) beautification. All gestures were produced by 5- to 7-year-olds.

work demonstrated that both children and adults prefer gesture interactions with visual feedback compared to interactions without feedback [3,4]. Live beautification may be a next step in improving gesture interactions for users of all ages. Providing live beautification will require accounting for the inconsistencies and high levels of noise common in children's gestures. Segmentation of children's gestures also poses a challenge since traditional segmentation algorithms may not be effective. A long-term plan for my work is to provide segmentation and noise reduction algorithms for use with the beautification techniques I develop.

4.2 Developing a Child-Centered Recognizer

Synthesizing the findings from my previous work, I also plan to create a new recognizer specifically designed to recognize gestures produced by children. The information I gained from my work on articulation features helps illuminate common problems with children's gestures and how to account for them.

The high level of inconsistency in the values of the features I analyzed in children's touchscreen gestures suggests that new features could be developed to provide a better characterization of their gestures. Recognizers currently rely on consistency between gestures of the same type, but it is clear that the traditional comparison of Euclidean distance between corresponding points of two gestures is not optimal for children's gestures. Thus, a continuing part of my work will be to investigate new features that may better serve in recognizing children's gestures.

Our work on human recognition will also be useful in evaluating the effectiveness of the new recognizer that I will develop. The human recognition rate of 90.60% accuracy provides a target accuracy for continued work in recognizing children's gestures. Machine accuracy eventually may be higher than this accuracy rate, but it is a good goal for continued work in recognition, especially considering that Read et al. [21] showed that children are satisfied with 91% accuracy.

4.3 Verifying the Benefit of Improved Recognition

After developing a child-focused recognition algorithm, I plan to verify its effectiveness through a user study with children ages 5 to 10 years old. The participants in the study will use two different implementations of the same gesture-based application. The application will consist of a game in which the participants are asked to produce a series of gestures. In one implementation, an existing recognizer designed for adults' gestures will be used. In the other, the new child-centered recognition algorithm will be used. I will measure the accuracy rate of each implementation and I will also have the participants complete a brief survey about their experience with the two implementations. The survey will ask the participants which of the two implementations they liked better and they will be asked to rate their experience on each of the implementations using a Likert scale. I hypothesize that the participants will rate the implementation with the new child-centered recognizer significantly higher than the other recognizer. Such a study would not only validate the effectiveness of the new child-centered recognizer, but also help demonstrate the benefit of the new recognizer for user experience.

A long-term goal of improving recognition of children's touchscreen gestures is to use them to create adaptive applications, particularly in the domain of education. An application for teaching children to draw letters and numbers, for example, could use the new child-centered recognizer to improve personalized feedback for individual users, improving the learning experience for the child using the application. An example of an existing system that uses gesture recognition to aid in education is EasySketch[2] [15], which takes advantage of gesture recognition and motor control classification to provide feedback on gestures and guidance when a child struggles with a particular gesture. When the system detects performance under a pre-determined threshold, a connect-the-dots type tracing activity is used to help the child practice making the gesture. My future work will focus on building a similar system that can benefit from improved recognition to provide better feedback and enhanced learning.

5. CONCLUSION

My work thus far has been focused on improving characterizations of children's touchscreen gestures. It has been established that children are less consistent than adults in the way they create gestures, but my work quantifies some of the specific ways in which children are inconsistent. My continued work will focus on improving recognition of children's gestures through beautification, new features, and new algorithms.

ACKNOWLEDGMENTS

I would like to thank my advisors, Dr. Lisa Anthony and Dr. Jaime Ruiz, for their support on this work. I would also like to thank the members of the INIT lab at the University of Florida for their help.

REFERENCES

[1] Luis von Ahn. 2005. Human Computation. *PhD thesis, School of Computer Science, Carnegie Mellon University.*

[2] Amazon. Mechanical Turk. Retrieved from www.mturk.com

[3] 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 the International Conference on Interaction Design and Children (IDC '13)*, ACM Press, 157–164. http://doi.org/10.1145/2485760.2485775

[4] Lisa Anthony, Quincy Brown, Jaye Nias, and Berthel Tate. 2015. Children (and adults) benefit from visual feedback during gesture interaction on mobile touchscreen devices. *International Journal of Child-Computer Interaction* 6, 17–27. http://doi.org/10.1016/j.ijcci.2016.01.002

[5] Lisa Anthony, Quincy Brown, Jaye Nias, Berthel Tate, and Shreya Mohan. 2012. Interaction and recognition challenges in interpreting children's touch and gesture input on mobile devices. *Proceedings of the ACM international Conference on Interactive Tabletops and Surfaces (ITS '12)*, ACM Press, 225–234. http://doi.org/10.1145/2396636.2396671

[6] Lisa Anthony, Quincy Brown, Berthel Tate, Jaye Nias, Robin Brewer, and Germaine Irwin. 2014. Designing smarter touch-based interfaces for educational contexts. *Personal and Ubiquitous Computing* 18, 6, 1471–1483. http://doi.org/10.1007/s00779-013-0749-9

[7] Lisa Anthony, Radu-Daniel Vatavu, and Jacob O. Wobbrock. 2013. Understanding the Consistency of Users' Pen and Finger Stroke Gesture Articulation. *Proceedings of Graphics Interface (GI '13)*, Canadian Information Processing Society, 87–94.

[8] Lisa Anthony and Jacob O. Wobbrock. 2010. A lightweight multistroke recognizer for user interface prototypes. *Proceedings of Graphics Interface (GI '10)*, Canadian Information Processing Society, 245–252.

[9] Lisa Anthony and Jacob O. Wobbrock. 2012. $N-Protractor: A Fast and Accurate Multistroke Recognizer. *Proceedings of Graphics Interface (GI '12)*, Canadian Information Processing Society, 117–120.

[10] Rachel Blagojevic, Samuel Hsiao-Heng Chang, and Beryl Plimmer. 2010. The power of automatic feature selection: rubine on steroids. *Joint Session of the Seventh Sketch-Based Interfaces and Modeling Workshop and Eighth Symposium on Non-Photorealistic Animation and Rendering*, Eurographics Association, 79–86.

[11] Quincy Brown and Lisa Anthony. 2012. Toward comparing the touchscreen interaction patterns of kids and adults. *Proceedings of the SIGCHI Workshop on Educational Software, Interfaces and Technology*, 4pp.

[12] Jason I. Hong and James A. Landay. 2000. SATIN: a toolkit for informal ink-based applications. *Proceedings of the International Symposium on User Interface Software and Technology*, ACM Press, 63–72. http://doi.org/10.1145/354401.354412

[13] Heloise Hwawen Hse and A. Richard Newton. 2005. Recognition and beautification of multi-stroke symbols in digital ink. *Computers and Graphics* 29, 4, 533–546. http://doi.org/10.1016/j.cag.2005.05.006

[14] Hilda K. Kabali, Matilda M. Irigoyen, Rosemary Nunez-Davis, Jennifer G. Budacki, Sweta H. Mohanty, Kristin P. Leister, and Robin L. Bonner. 2015. Exposure and Use of Mobile Media Devices by Young Children. *Pediatrics* 136, 6, 1044–1050. http://doi.org/10.1542/peds.2015-2151

[15] Hong-hoe Kim, Paul Taele, Jinsilhwaryoung Seo, Jeffrey Liew, and Tracy Hammond. 2016. EasySketch2: A novel sketch-based interface for improving children's fine motor skills and school readiness. *Proceedings of the International Symposium on Sketch-Based Interfaces and Modeling (SBIM '16)*, Eurographics Association, 69–78.

[16] Hong-hoe Kim, Paul Taele, Stephanie Valentine, Erin McTigue, and Tracy Hammond. 2013. KimCHI: a sketch-based developmental skill classifier to enhance pen-driven educational interfaces for children. *Proceedings of the International Symposium on Sketch-Based Interfaces and Modeling (SBIM '13)*, ACM Press, 33–42. http://doi.org/10.1145/2487381.2487389

[17] Heather L. Kirkorian and Tiffany A. Pompek. 2013. Toddlers and touch screens: potential for early learning? *Zero to Three* 33, 4, 31–37.

[18] Mary LaLomia. 1994. User acceptance of handwritten recognition accuracy. *Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, 107–108. http://doi.org/10.1145/259963.260086

[19] Lorna McKnight and Daniel Fitton. 2010. Touch-screen technology for children. *Proceedings of the International Conference on Interaction Design and Children*, ACM Press, 238–241. http://doi.org/10.1145/1810543.1810580

[20] Brandon Paulson and Tracy Hammond. 2008. PaleoSketch: Accurate Primitive Sketch Recognition and Beautification. *Proceedings of the International Conference on Intelligent User Interfaces (IUI '08)*, 1–10. http://doi.org/10.1145/1378773.1378775

[21] Janet C. Read, Stuart MacFarlane, and Chris Casey. 2003. "Good enough for what?": acceptance of handwriting recognition errors by child users. *Proceedings of the International Conference on Interaction Design and Children (IDC '03)*, ACM Press, 155–155. http://doi.org/10.1145/953536.953565

[22] Dean Rubine. 1991. Specifying gestures by example. *ACM SIGGRAPH Computer Graphics* 25, 4, 329–337. http://doi.org/10.1145/127719.122753

[23] Alex Shaw and Lisa Anthony. 2016. Analyzing the Articulation Features of Children's Touchscreen Gestures. *Proceedings of the International Conference on Multimodal Interaction (ICMI '16)*, ACM Press, 333–340. http://doi.org/10.1145/2993148.2993179

[24] Alex Shaw, Jaime Ruiz, and Lisa Anthony. 2017. Comparing Human and Machine Recognition of Children's Touchscreen Stroke Gestures. *Proceedings of the International Conference on Multimodal Interaction (ICMI '17)*, ACM Press, To appear.

[25] Radu-Daniel Vatavu, Lisa Anthony, and Jacob O. Wobbrock. 2012. Gestures as point clouds: a $P recognizer for user interface prototypes. *Proceedings of the ACM International Conference on Multimodal Interaction (ICMI '12)*, ACM Press, 273–280. http://doi.org/10.1145/2388676.2388732

[26] Radu-Daniel Vatavu, Lisa Anthony, and Jacob O. Wobbrock. 2013. Relative accuracy measures for stroke gestures. *Proceedings of the ACM International Conference on Multimodal Interaction*, ACM Press, 279–286. http://doi.org/10.1145/2522848.2522875

[27] Radu-Daniel Vatavu, Daniel Vogel, Géry Casiez, and Laurent Grisoni. 2011. Estimating the perceived difficulty of pen gestures. *Proceedings of the IFIP TC International Conference on Human-Computer Interaction (INTERACT '11)*, Springer-Verlag, 89–106.

[28] Don Willems, Ralph Niels, Marcel van Gerven, and Louis Vuurpijl. 2009. Iconic and multi-stroke gesture recognition. *Pattern Recognition* 42, 12, 3303–3312. http://doi.org/10.1016/j.patcog.2009.01.030

[29] 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 the ACM Symposium on User Interface Software and Technology (UIST '07)*, ACM Press, 159–168. http://doi.org/10.1145/1294211.1294238

[30] Julia Woodward, Alex Shaw, Annie Luc, Brittany Craig, Juthika Das, Phillip Hall, Akshay Hollay, Germaine Irwin, Danielle Sikich, Quincy Brown, and Lisa Anthony. 2016. Characterizing How Interface Complexity Affects Children's Touchscreen Interactions. *Proceedings of the ACM International Conference on Human Factors in Computing Systems (CHI '16)*, ACM Press, 1921–1933. http://doi.org/10.1145/2858036.2858200