An Observational Coding Scheme for Detecting Children's Usability Problems in Augmented Reality

Iulian Radu
Georgia Institute of Technology
North Ave, Atlanta, GA, USA
iulian@gatech.edu

Simina Avram
Georgia Institute of Technology
North Ave, Atlanta, GA, USA
savram3@gatech.edu

Katherine Guzdial
Georgia Institute of Technology
North Ave, Atlanta, GA, USA
kguzdial@gatech.edu

Abstract
Handheld augmented reality (AR) is a unique medium that can be helpful for young children's entertainment and education, but to achieve the benefits of this technology, experiences need to be appropriately designed for young children's developing skills. In the current research, we are interested in identifying and quantifying the usability problems that are encountered by children using handheld AR applications. We have developed a qualitative coding scheme for detecting AR usability problems from video observations of user interactions. We then applied the coding scheme to extract usability problems encountered by children aged 5-10 years old as they played with a handheld AR game. Through triangulation with performance data, we have detected usability problems related to AR and non-AR components of the experience. Our analysis found positive and negative correlations between usability problems and child age, and found that children experienced a variety of problems such as grip and posture strains, inabilities to recover from various types of tracking loss, difficulties orienting their body around the gameboard, etc. This work is the first in a series of studies applying the qualitative coding for understanding young children's ability to use handheld AR applications.

IDC '17, June 27-30, 2017, Stanford, CA, USA

ACM ISBN 978-1-4503-4921-5/17/06.
http://dx.doi.org/10.1145/3078072.3084337

Mental & Emotional States
FRU - Indication of frustration / dislike
SUGG - Child makes suggestion
CONF - Indication of confusion
BOR - Indication of boredom
AHA - Indication of "aha" moment
TIRV - Indication of physical tiredness
SPACEV - Indication of 3D space awareness
General Movement
SCRATCH – Interrupts game to scratch
HSTR - Fingers / hand / arm stretched
HSHAKE - Hand shaken
BSTR - Body stretching
BSIT – Body sitting down
ELT - Elbow or hand is resting on table
PHONEDROP - Phone dropped or slips
PHONEDOWN - Puts the phone down
BUMP - Bumps or trips body
BSW - Back has switched its posture
Grips
GSW- Hand has switched its grip
Help and Interruptions
HLPME - Child asks for help
HLPV - Experimenter gives quick help
INTH1 – Gameplay interrupted while experimenter helped for a period of time, just verbally
INTH2 – Gameplay interrupted while experimenter helped for a period of time, and took away the phone / moved the paper / or touched the child
IGN - Child ignored instructions from the game or experimenter
INT – Child did something which caused their gameplay to be interrupted (e.g., looking at experimenter)
Tracking Events
TLW – Tracking lost while walking
TLF – Tracking lost, finger occlusion
TLA – Tracking lost, aiming away
TLC – Tracking lost, too close

Table 1. Behaviors tagged through the coding scheme.

Author Keywords
Augmented reality; children; mobile; qualitative method; video coding; usability; developmental psychology

ACM Classification Keywords
H.5.1 Artificial, augmented, and virtual realities.

Introduction
Augmented reality (AR) is a technology that enables computer-generated imagery to be superimposed on a view of the real world, and handheld devices, such as phones and tablets, will likely be the dominant method for delivering AR experiences to large number of children for the near future. AR has been shown to have measurable benefits to children's education over traditional approaches [1,2,3]. To achieve these benefits, AR experiences need to be usable by children. In this research, we are developing a method for identifying and quantifying usability problems in handheld AR experiences for children, permitting researchers to use observational video data to compare between different age groups and/or different AR designs.

Research in AR technology has typically measured usability through performance metrics and subjective reports [10,11]. Such metrics quantify certain aspects of usability, but they can miss user behaviors leading to the problem. Another method for quantifying usability problems is analysis of observational recordings. The popular DEVAN method [7] was designed for contexts where adults interact with desktop applications. Videos are annotated through a qualitative coding scheme, and coded behaviors are aggregated into usability problems [8,9]. Unlike adult users, children typically use digital games and are strongly affected by motivation, thus an adapted scheme was developed for such contexts [6].

In handheld augmented reality applications, different usability problems arise than compared to desktop computer interfaces, because handheld AR experiences require users to manipulate a mobile device through physical body motions in a three-dimensional space. Qualitative coding schemes have previously been used for understanding observational data of children playing with augmented reality [12,25,26,27,28]. However, such coding schemes are focused on studying specific aspects of the user experience other than usability. We do not know of any existing coding scheme specifically for measuring usability of children's AR experiences. This research intends to address this limitation, by answering:

1. What observable behaviors can serve as indicators of usability problems in augmented reality?
2. What types of usability problems are experienced by children using handheld augmented reality?
3. What is the relationship between children's age and the types of usability problems encountered?

To address these questions, we first developed a coding scheme for detecting usability breakdowns, then applied the coding scheme to 40 children playing with a handheld augmented reality game, identifying a set of usability problems and differences between age groups.

Coding Scheme and Experimental Design
We developed a first version of the AR usability problems coding scheme (Table 1). Codes are observable child behaviors that indicate a usability problem may be occurring. After participant videos are tagged with these codes, a set of usability problems can then be extracted, through the process explained in the next sections. The coding scheme was designed based on literature about children's usability problems, and then was refined through iterative application on pilot study videos. The

Figure 1. A child playing the augmented reality game.

Figure 2. Game items are exposed in open levels (left) and enclosed in tunnel levels (right).

Figure 3. Examples of strained body posture.

initial design was primarily informed by the research in [6], describing a coding scheme for detecting problems in children's desktop computer games, and by the research in [13], which describes links between child development and the usability of AR technology. The coding scheme was then refined by application to eight videos of children playing AR games. Researchers and undergraduate coders participated in this process, revising the scheme mainly by adding new codes and adjusting ambiguous codes (i.e. codes whose meaning would be interpreted differently by different people).

We then used the coding scheme to extract usability problems encountered by children of different ages as they played with a handheld augmented reality game. A total of 40 children participated, 14 children in each age group: 5-6 year olds (6M, 8F), 7-8 year olds (7M, 7F), and 9-10 year olds (7M, 7F). The game, previously been described in [21], was a typical tabletop augmented reality game where a 3D virtual world appears on top of the "gameboard" paper once it is viewed through a smartphone camera (Figure 1). The player took the role of a wizard who must collect a set of magical lemons, to create items for their virtual pets. The game was composed of a tutorial plus 4 levels of gameplay. Each level consisted of repeated trials of item collection, where children collected lemons by tapping the screen, or by using a crosshair activated by side buttons. The game was designed to test the relationship between usability and user movement around the gameboard. Half the game levels required changes in perspective, as the user had to collect targets enclosed inside tunnels (Figure 2). When games encourage changes in perspective, it may cause spatial and movement problems, such as bumping into objects, difficulties gripping the device, etc.

Two coders used the coding scheme to code the play sessions of all children, overlapping on 10% of the videos. To determine if coders agreed on each coded event, a 4-second overlap time window was used, similar to [6]. The coders achieved an inter-rater agreement validity of Cohen's Kappa equal to 0.80 or greater (percent agreement > 97%). Inter-rater agreement was calculated using Quera and Bakeman's GSEQ algorithm for continuous-timed event coding [22].

To extract types of usability problems, the coded events were first categorized into clusters of related behaviors, and a severity rating was assigned to each event according to how much assistance a child required to proceed. Clusters of events were then aggregated into usability problem types (Table 2.A and 2.B). Other types of events were also captured during video analysis, such as reasons why tracking loss was lost (Table 3.A and 3.B) and types of grips used by children (Figure 4). We then performed quantitative analysis on the counts of events for each child: Spearman rank correlations between children's age (in months) and number of event occurrences for each usability problem type; and Wilcoxon signed-rank tests comparing AR movement conditions.

USABILITY PROBLEMS DETECTED

Tables 2.A. and 2.B. list issues encountered by children under different age groups and under different AR movement conditions. The number of problems encountered by children was relatively similar between age groups, overall. Although these numbers are similar, younger children experience different problems than older children, as will be discussed below. Additionally, younger children experienced higher severity problems and required more assistance to fix the usability

Observed Problems in Tunnel Levels	Num. children aged 5-6/7-8/9-10
MANIPULATION	
Strained body posture OT	0% / 28% / 33%
Strained grip	0% / 21% / 16%
Dropping the phone	7% / 0% / 8%
SPACE	
Difficulty orienting body, while playing the game OT	21% / 50% / 0% **(7% / 0% / 0%)**
Difficulty orienting body, while fixing tracking OT	CN 50% / 7% / 0% **(42% / 7% / 0%)**
ABSTRACT THINKING	
Needing in-game instruction on crosshair	7% / 0% / 0% **(7% / 0% / 0%)**
Not understanding storyline	7% / 7% / 8%
Not understanding mechanics	14% / 14% / 0%
Difficulties interpreting or fixing tracking loss	CN 50% / 14% / 16% **(42% / 14% / 0%)**
ATTENTION	
Bumping or tripping OT	28% / 35% / 41%
Interruption from self-distraction	21% / 7% / 0%
Interruption from scratching	21% / 35% / 58%

Table 2.A. Usability problems encountered in Tunnel levels. Bold=Children requiring experimenter assistance. CP=positive correlation with age ($p<0.05$); CN=negative correlation. OT=Significant differences between open vs. tunnel levels.

problems. In the following discussion we present our qualitative findings and report triangulation with our performance data whenever relevant.

Problems Specific to Handheld Augmented Reality

When children experience usability issues in AR experiences, the issues may be related to different layers of the experience: game design and narrative, interaction design, AR tracking technology, hardware ergonomics, etc. In the current section, we present issues that are likely to be caused specifically by the presence of AR technology, and which may occur across other kinds of AR experiences.

Strained body postures in older children: Children aged 7-8 and 9-10 years old were observed bending their backs or necks in ways that appeared strained, or stretching their neck or back (Figure 3). The number of children experiencing this behavior increased with age; older children were observed to bend their body more often, and seemed more reluctant to move around the gameboard. This issue may, in part, be related to the fixed angle of tunnels, indicating that AR games should be designed to measure and adapt to player height. Although older children subjectively indicated enjoying the challenge of tunnel levels, they frequently noted a dislike for walking around the game, possibly indicating a sedentary lifestyle or familiarity with digital experiences that lack movement.

Difficulty orienting while playing the game: Performance metrics logged during gameplay showed that 5-6 year old children completed the tasks significantly slower than other age groups, but the performance data does not indicate why. Through the qualitative analysis, we see that children had most trouble with changing perspective while collecting lemons inside tunnels. This issue occurred more severely in 5-6 year old children, with 7% of children requiring researcher assistance. None of the older children, 9-10 years old, experienced this issue. This may be due to 5-6 year old children's developing abilities to understand space and coordinate their body [13], and can indicate that AR experiences for 5-6 year olds should be lenient if requiring perspective changes.

Difficulty orienting while fixing tracking: Losing the tracking of the paper is a significant issue because it interrupts the gameplay, and may occur for several reasons (Table 3.A and 3.B). Performance data showed that 5-6 year old children are significantly slower at fixing tracking. The qualitative data shows that children had trouble fixing tracking when required to reorient their body. In tunnel levels, the experimenter had to assist 40% of children in the 5-6 year old range, and 7% of children in the 7-8 year old range, indicating that this issue is problematic for young children when they are required to move around the gameboard. This is possibly because in such levels children changed their perspective to the gameboard, had likelihood of tracking losses, and it may be related to spatial and abstract thinking abilities of younger children [13].

Difficulties recovering from tracking loss: Abstract thinking is required to understand why tracking has become lost and how to fix it. We clustered all events which indicated that children either did not understand, were frustrated, or required assistance when the game lost tracking. This issue included instances of children being unable to orient themselves while recovering tracking, and included other instances of frustration, or difficulties fixing tracking, such as the need to reposition fingers away from the camera. This event was negatively

Observed Problems in Non-Tunnel Levels	Num. children aged 5-6/7-8/9-10
MANIPULATION	
Strained body posture OT	CP 0% / 14% / 33%
Strained grip	0% / 21% / 0%
Dropping the phone	0% / 7% / 8%
SPACE	
Difficulty orienting body, while playing the game OT	0% / 0% / 0%
Difficulty orienting body, while fixing tracking OT	0% / 0% / 0%
ABSTRACT THINKING	
Needing initial instruction on how to use crosshair	71% / 79% / 58% **(71% / 79% / 58%)**
Needing in-game instruction on crosshair	7% / 7% / 0% **(7% / 7% / 0%)**
Not understanding storyline	0% / 14% / 0%
Not understanding mechanics	7% / 0% / 0%
Difficulties interpreting or fixing tracking loss	14% / 7% / 0% **(7% / 0% / 0%)**
ATTENTION	
Bumping or tripping OT	7% / 0% / 0%
Interruption from self-distraction	0% / 0% / 0%
Interruption from scratching	CP 0% / 35% / 50%

Table 2.B. Usability problems encountered in Non-Tunnel levels. Bold=Children requiring experimenter assistance. CP=positive correlation with age ($p<0.05$); CN=negative correlation. OT=Significant differences between open vs. tunnel levels.

correlated with children's age. Some of the older children experienced this issue but none required experimenter assistance to resolve it.

Bumping or tripping: Children of all ages were observed either bumping into the table or tripping over their own feet as they changed their perspective on the gameboard. This event occurred significantly more in tunnel levels where users were required to change perspective. This event may indicate a focused attention on the game, and may cause problems if children play such games in dynamic environments such as in cluttered spaces, busy classrooms or outdoors.

Interruption due to self-distraction: Children were sometimes observed losing interest in the game and using the AR camera to look away at other pieces of the environment. This behavior was observed most often with 5-6 year old children, typically when they were challenged or frustrated by tunnel levels, possibly indicating mechanism for distraction from difficult tasks.

Types of Tracking Losses: The AR technology used in our application was based on image-based tracking. The experience would stop working if the smartphone camera (located on the left side of the phone) stopped seeing the gameboard. From performance metrics logged in the game, we discovered that children of all ages experienced tracking losses, and 5-6 year olds had significantly more tracking losses than both older age groups. With the current coding scheme, we identified multiple reasons: losing tracking while walking (negatively correlated with age), covering the camera with the finger (occurring in each age group), and aiming too close or away (negatively correlated with age).

Adverse Effects of Technology Exposure

Parents were asked to report the kinds of devices their child has been exposed to, and this metric was correlated with the number of tracking errors encountered by children (accounting for age effects in the analysis). The correlation was, surprisingly, a positive correlation: children with highest technology exposure had 83% increase in number of tracking losses due to fingers in front of the camera compared to children with lowest exposure. Performance metrics showed that children with more technology experience had 39% more selection errors when touching targets on the screen. This indicates a potential issue with introducing new technologies, namely that children who are experienced with technology will transfer expectations from old to new technologies, causing issues such as expectations of technology resilience to errors, and the use of familiar grips which may interfere with tracking technology.

Problems Not Specific to Augmented Reality

This section will discuss usability problems caused by other layers of the experience, such as the hardware ergonomics, game narrative, screen interaction types.

Grip Styles and Strains: Children aged 7-8 and 9-10 years old were observed such as stretching their hand, shaking their hand, or repeatedly shifting their grip in a way suggesting they are searching for a more comfortable grip. We identified a shift in grips over age (Figure 4), noting a possibility of transition period during age 7-8. As children age, their hands develop in both strength and size, therefore changing their ability to grip and manipulate the smartphone device. In our study, younger children showed a preference for the curled grip, but as children aged the use of the crab grip increased, whereby older children preferred the crab grip. There is a

Tracking Loss Reasons in Tunnel Levels	Num. children aged 5-6/7-8/9-10
Walking[OT]	[CN] 100% / 50% / 50%
Stationary, finger occlusion	50% / 64% / 41%
Stationary, aiming the camera away	[CN] 42% / 28% / 0%
Stationary, aiming too close	14% / 14% / 0%

Table 3.A. Children in each age group encountering tracking loss, in levels with tunnel-enclosed targets.

Tracking Loss Reasons in Non-Tunnel Levels	Num. children aged 5-6/7-8/9-10
Walking[OT]	[CN] 28% / 0% / 0%
Stationary, finger occlusion	71% / 57% / 41%
Stationary, aiming the camera away	28% / 7% / 8%
Stationary, aiming too close	7% / 0% / 0%

Table 3.B. Children in each age group encountering tracking loss, in levels with non-enclosed targets. Significant correlations between age and number of events are marked as "CP"=positive correlation; "CN"=negative correlation. Significant differences between open vs. tunnel levels marked as "OT".

strong use of the straight grip in 7-8 year olds, possibly indicating that the straight grip is an intermediary grip style where children's strength allows them to hold and control the phone by pushing on the sides; unfortunately this grip is quite unstable and leads to frequent tracking losses. A small number of children within each age group were observed dropping the phone, typically occurring as children were changing or relaxing their grip. All children recovered, but this may lead to mobile device damage.

Needing assistance with the game mechanics: All children knew how to touch on the screen with their finger. However, many young children did not know how to use the crosshair interaction, which required them to aim with the center of the screen. Children also had questions about game storyline and mechanics, such as why lemons are in tubes, how things can be made from lemons, how many lemons had to be collected, etc.

Limitations and Future Work

In upcoming research, we are interested to apply the coding scheme to a variety of AR experiences. Currently, our results are likely influenced by multiple variables, such as game interactions (ex: if users collect items by moving close/far vs. touching on screen), by hardware ergonomics (ex: mobile phones vs. tablets; camera in the center of the device vs. on the side), by AR tracking software (ex: image based tracking vs. depth-based markerless tracking), etc. As a next step, we will study children's usability problems with a non-AR version of the same game, allowing us to account for factors such as gameplay interactions and mobile device ergonomics. In the future, we will apply the coding scheme to detect usability problems in other types of AR experiences. We expect this process will refine the coding scheme, provide triangulation between metrics, identify a common set of usability problems encountered by children in AR, and ultimately lead to understanding relationships between AR designs and child development.

Finally, good games should be easy to learn and hard to master [6], and should keep players highly engaged by maintaining a dynamic balance between the player's abilities and the challenge offered by the game [23,24]. The current research has identified degree of difficulty which children experience with the AR game. Designers can use this knowledge to create AR games that provide appropriate challenge for children of specific ages.

Acknowledgements: We wish to thank the children and parents who participated in this study; Emory University Child Study Center and Spatial Cognition Lab; and the students who helped with video-coding in this study.

References

1. Radu, I. (2014). Augmented reality in education: a meta-review and cross-media analysis. Personal and Ubiquitous Computing, 18(6), 1533-1543.
2. Vincenzi D. A. , Valimont B. , Macchiarella N., Opalenik C., Gangadharan S. N., and Majoros A. E. (2003). "The Effectiveness of Cognitive Elaboration Using Augmented Reality as a Training and Learning Paradigm," Annual Meeting of the Human Factors and Ergonomics Society, Denver, CO, USA, pp. 2054-2058.
3. Lindgren R. and Moshell J. M. (2011). "Supporting Children's Learning with Body-Based Metaphors in a Mixed Reality Environment," Simulation, pp. 177-180
4. Dix, A., (2009). Human-computer interaction. Springer US.
5. Sharp, H., Rogers, Y. and Preece, J., (2007). Interaction design: beyond human-computer interaction.
6. Barendregt, W. and Bekker, M.M., (2006). Developing a coding scheme for detecting usability and fun problems in computer games for young children. Behavior research methods, 38(3), pp.382-389.

Figure 4.A. Grips used by children: Curl (top), Crab (middle), Straight (bottom).

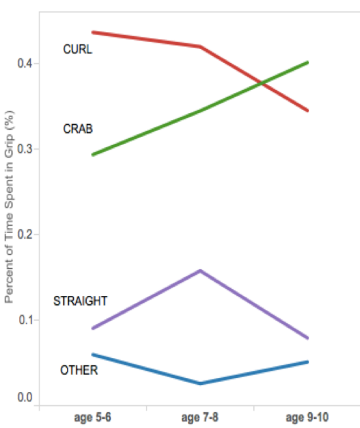

Figure 4.B. Percentage of time each grip type was used by children of different ages.

7. Vermeeren, A.P., den Bouwmeester, K., Aasman, J. and de Ridder, H., (2002). DEVAN: a tool for detailed video analysis of user test data. Behaviour & Information Technology, 21(6), pp.403-423.
8. Gamberini, L., Spagnolli, A., Prontu, L., Furlan, S., Martino, F., Solaz, B.R., Alcañiz, M. and Lozano, J.A., (2013). How natural is a natural interface? An evaluation procedure based on action breakdowns. Personal and ubiquitous computing, 17(1), pp.69-79.
9. Cockton, G. and Lavery, D., (1999), August. A framework for usability problem extraction. In INTERACT (pp. 344-352).
10. Bai, Z. and Blackwell, A.F., (2012). Analytic review of usability evaluation in ISMAR. Interacting with Computers, 24(6), pp.450-460.
11. Dünser, A., Grasset, R. and Billinghurst, M., (2008). A survey of evaluation techniques used in augmented reality studies (pp. 5-1). Human Interface Technology Laboratory New Zealand.
12. Hornecker, E. and A. Dünser (2009). "Of pages and paddles: Children's expectations and mistaken interactions with physical-digital tools." Interacting with Computers 21(1-2): 95-107.
13. Radu, I. and B. MacIntyre (2012). Using children's developmental psychology to guide augmented-reality design and usability. Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on, IEEE.
14. Nischelwitzer, A., F.-J. Lenz, et al. (2007). "Some Aspects of the Development of Low-Cost Augmented Reality Learning Environments as Examples." Universal Access in Human-Computer Interaction.: 728-737.
15. Dunleavy, M., C. Dede, et al. (2009). "Affordances and Limitations of Immersive Participatory Augmented Reality Simulations for Teaching and Learning." Journal of Science Education and Technology 18(1): 7-22.
16. Piaget, J. (1970). "Science of education and the psychology of the child. Trans. D. Coltman."
17. Rosser, R. A. (1994). Cognitive development: Psychological and biological perspectives, Allyn and Bacon Boston.
18. Wyeth, P. a. P., H.C. (2003). "Using developmental theories to inform the design of technology for children." Proceedings of the 2003 conference on Interaction design and children: 93-100.
19. Gelderblom, H. (2009). Designing technology for young children: guidelines grounded in a literature investigation on child development and children's technology. Doctor of Philosophy, University of South Africa.
20. Bekker, T. and A. N. Antle (2011). "Developmentally situated design (DSD): making theoretical knowledge accessible to designers of children's technology." Proceedings of the 2011 annual conference on Human factors in computing systems: 2531--2540.
21. Radu, I., MacIntyre, B. and Lourenco, S., (2016), Comparing Children's Crosshair and Finger Interactions in Handheld Augmented Reality: Relationships Between Usability and Child Development. In Proceedings of the International Conference on Interaction Design and Children (pp. 288-298). ACM.
22. Quera, V. and R. Bakeman (2000). GSEQ for Windows: New software for the sequential analysis of behavioral data, with an interface to The Observer. 3rd International Conference on Methods and Techniques in Behavioral Research. Nijmegen, Netherlands.
23. Shell, J. (2008). The art of game design.
24. Swink, S. (2009). Game feel. A Game Designer's Guide to Virtual Sensation. Burlington, MA, 1.
25. Zimmerman, Heather Toomey, et al. (2015) "Using augmented reality to support observations about trees during summer camp." Proceedings of the 14th International Conference on Interaction Design and Children. ACM.
26. Cassidy, B., Sim, G., Horton, M., & Fitton, D. (2015). Participatory design of wearable augmented reality display elements for children at play. In Computer Science and Electronic Engineering Conference (CEEC), 2015 7th (pp. 53-58). IEEE.
27. Escobedo, L., Tentori, M., Quintana, E., Favela, J., & Garcia-Rosas, D. (2014). Using augmented reality to help children with autism stay focused. IEEE Pervasive Computing, 13(1).
28. Fleck, S., Hachet, M., & Bastien, J. M. (2015). Marker-based augmented reality: Instructional-design to improve children interactions with astronomical concepts. In Proceedings of the 14th International Conference on Interaction Design and Children (pp. 21-28). ACM.