EG UK Computer Graphics & Visual Computing (2015)
Rita Borgo, Cagatay Turkay (Editors)

User-defined gestures for Augmented Reality with Smart Phones

Francesca Madeddu[1], Daniel Archambault[1], Rita Borgo[1] and Lewis Hancock[2]

[1]Swansea University, UK
[2]Codemasters, UK

Abstract

The purpose of this research is to investigate user interaction with smart phones in an augmented reality context. In particular, we present the results of a guessability study performed with 15 participants in order to: 1) define a gesture taxonomy with a specific focus on selection, rotation and zoom tasks, 2) uncover insights on the participants' mental model, 3) discover implications for the development of augmented reality applications.

Categories and Subject Descriptors (according to ACM CCS): H.5.1 [INFORMATION INTERFACES AND PRESENTATION]: Multimedia Information Systems—Artificial, augmented, and virtual realities

1. Introduction

Augmented Reality (AR) is a promising technology that opens a universe of new possibilities. However, when introducing new technologies, the first step should be understanding their effects on user interaction.
In particular, when trying to define gestures for new technologies, it is preferable to elicit gestures from end-users rather than letting HCI designers define them. Designers tend to design gestures that are perceived by the final user as "physically and conceptually complex" [MWW10]. Wobbrock [WARM05] formalised a methodology for eliciting end-users gestures popularised as guessability study: in a guessability study, users are asked, given a certain *effect*, to guess and perform the action which they think will *cause* the given effect.

2. Related Work

To recognise the importance of involving users in the process, several research groups have performed guessability studies to define a gesture taxonomy in different domains. All these studies investigate new interactions required by the introduction of new technologies. Wobbrock et al. [WMW09] [MWW10] focused their research on *surface gestures* with tabletops; these results were used for a later study comparing the elicited gestures set with another set designed by HCI experts. Willet et al. [WLI14] refined this research on multi-touch surface selection for data-graphics. Ruiz et. al investigated *motion gestures* with mobile devices. Despite the fact that smart phones are equipped with a wide variety of sensors allowing sophisticated new ways of interaction, surface gestures are seen as primary input. For *direct manipulation gestures*, Connell et al. [CKLP13] preformed a study to elicit child-defined gestures with Kinect. Finally, with a specific focus on AR environments, Lee et al [LWP*15] and Piumsomboon et. al [PCBC13] investigated interaction with virtual mirrors and head mounted displays respectively.
Findings from these studies generally suggest that users prefer simple interactions, and that the previous learned idioms affect the user mental model. For some interactions there is little agreement on the preferred gestures.

3. Developing user defined gesture set

The aim of the current research is to investigate user interaction in AR environments using smart phones, as nothing has been done so far in this direction. Moreover, previous studies try to define a taxonomy within a certain category of gestures (surface, motion or direct manipulation). The current research focuses on eliciting end-user gestures with no distinction among categories.

3.1. Users

The study involved 15 participants (60% male 40% female) with an age range of 19 to 40 years old. Most participants were students (93%), of whom 73% were from the computer science department. In fact, 93% of them included themselves in a range from a normal to expert user on mobile devices and all of them owned at least one mobile device. Finally, even though 40% knew what AR was, only the 26% had tried an AR application.

3.2. Apparatus

For the purpose of this study, a custom application was developed to let the user to interact with objects in an AR environment and to record this interaction. The smartphone used to deploy the custom application was a Samsung Galaxy Note 3, measuring 151.2 mm x 79.2 mm x 8.3 mm with a 1080 x 1920 of resolution. For each surface gesture, the application captured the coordinates of the user's fingers and saved then as images, where the colour of each line represents a different finger. Each time the participant touched the screen a screenshot was also saved to contextualise the captured gesture. The application was developed to give no feedback, as the expected feedback was part of the information we wanted to elicit from participants.
The target object was a rectangular box with an image on top (figure 1). This image, once captured by the camera, triggered the AR display. The choice of using a box, rather than a sheet of paper, improved the interactive affordance of the object, so that the participant was implicitly made aware that they could touch the object. In this way, the bias of surface and motion gestures over direct manipulation was mitigated.

Figure 1: *Example images captured by the logging application for a rotation task.*

3.3. Data gathering

An initial structured questionnaire was used to gather demographic information and technological knowledge towards mobile technologies and AR. During the experiment the researcher used a structured data sheet to jot down observations about the participant's interaction with the smart phone. The interactions were also partially recorded by the preinstalled logging application. Finally, a semi-structured interview was conducted to gather information about the participants' opinion on the performed gestures. All stages of the experiment were audio-recorded and participants were asked to adopt a talk-aloud protocol.

3.4. Tasks

Participants were asked to complete four tasks. Given a scene displayed by the smart phone's camera, a 3D model of an Anubis statue and a bag appeared, and participants were asked to perform the action they felt was most appropriate to cause a specific effect. The tasks were selecting the Anubis statue and dragging it inside the bag, zooming on it, and rotating it. Finally, they were asked to explore the scene.
To avoid bias in the explanation of the task, words focused on the effect to achieve rather than on the action needed to obtain it. So for example, instead of asking the participant to rotate the Anubis model, they were asked how they would see the other side of the model.

3.5. Procedure

Participants were given a task description paper to introduce them to the AR concept and the experimental procedure. Once the participant finished reading, they were given a questionnaire to complete. After these instructions, the questionnaire was completed and the participant was given a smart phone and asked to perform the four tasks. As target recognition lies outside the scope of the experiment, the researcher demonstrated to the participant how to make the AR appear the first time. Finally, the researcher conducted with the final interview.
The experiment lasted 15 to 30 minutes and the participant sat next to the researcher in front of a rectangular desk. The target object was positioned in front of them at a reachable distance.

4. Results

Table 1 lists the 18 gestures elicited for selection, zoom and rotation tasks. Results from the experiment show that, given a specific category, the choice of a gesture was straightforward. In particular, for the selection task, 93% of the participants chose a surface drag and drop gesture. For the zoom task, even if the agreement was lower, more than a half (53%) chose the pinch to zoom surface gesture. Finally, for the rotation task, this trend was inverted and the highest agreement was towards a direct manipulation, meaning participants directly rotated the target object. However, the agreement for the gesture was the lowest of the three tasks. In fact, only 40% of the participants tried the gesture first.

5. Discussion and Conclusion

As previously highlighted by the work of Wu et al. [WMW09], participant behaviour seemed highly influenced by previous experience, especially when a standard gesture existed for the task. In these cases, the fact that they were interacting with a AR environment did not affect the gesture for the task. However, when a standard gesture did not exist, as for the case of the rotation task, their behaviour was

			1st	2nd
Selection	Surface	One finger drag&drop from Anubis to Bag	**14**	
		One finger drag&drop from Bag to Anubis		2
		Single tap on Anubis		2
		Double tap on Anubis		1
		Single tap on the Bag		
		Single tap on background		1
	Motion	Move the Phone laterally from Anubis to Bag	1	
		Shake the Phone		2
Zoom	Surface	Pinch to Zoom	**8**	2
		Double Tap	1	1
	Motion	Move the Phone close to Anubis	4	1
	Mix	Move the Phone + Grab Anubis to get them closer to each other	1	2
		Pinch to Zoom + Move the Phone	1	
Rotation	Surface	One finger circle rotation	2	
		Two finger circle rotation	2	1
		One finger straight line	1	
	Motion	Move the Phone around Anubis	4	1
	Direct Manipulation	Rotate Anubis	**6**	2

Table 1: *Elicited gestures from each task are grouped for category. For each gesture it is reported if the participant tried it as first of second attempt. Not all participants tried more than once.*

affected by their critical judgment. Moreover, the current research confirmed a preference for simple gestures. In fact participants always interacted with the screen using one or two fingers. In general, there was a difference among the first participant's gesture, the gesture recognised as standard, and the preferred gesture.
The next step in this research will be to perform an evaluation against elicited gestures.

6. Acknowledgements

The authors thank Tom Owen and Harri Mansikkamaki from Leadin Companay for their insightful comments.

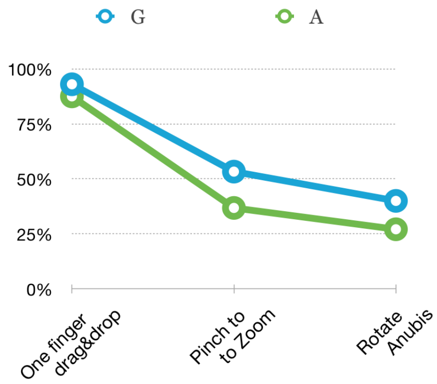

Figure 2: *In the graph are reported guessability (G) and agreement (A) values for the most performed gestures of each task. Values are calculated referring to the Wobbrock methodology [WARM05].*

References

[CKLP13] Connell S., Kuo P.-Y., Liu L., Piper A. M.: A wizard-of-oz elicitation study examining child-defined gestures with a whole-body interface. In *Proceedings of the 12th International Conference on Interaction Design and Children* (2013), ACM, pp. 277–280. 1

[LWP*15] Lee G. A., Wong J., Park H. S., Choi J. S., Park C. J., Billinghurst M.: User defined gestures for augmented virtual mirrors: A guessability study. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (2015), ACM, pp. 959–964. 1

[MWW10] Morris M. R., Wobbrock J. O., Wilson A. D.: Understanding users' preferences for surface gestures. In *Proceedings of graphics interface 2010* (2010), Canadian Information Processing Society, pp. 261–268. 1

[PCBC13] Piumsomboon T., Clark A., Billinghurst M., Cockburn A.: User-defined gestures for augmented reality. In *Human-Computer Interaction - INTERACT 2013*. Springer, 2013, pp. 282–299. 1

[WARM05] Wobbrock J. O., Aung H. H., Rothrock B., Myers B. A.: Maximizing the guessability of symbolic input. In *CHI'05 extended abstracts on Human Factors in Computing Systems* (2005), ACM, pp. 1869–1872. 1, 3

[WLI14] Willett W., Lan Q., Isenberg P.: Eliciting multi-touch selection gestures for interactive data graphics. In *Short-Paper Proceedings of the European Conference on Visualization (EuroVis)* (2014), Eurographics. 1

[WMW09] Wobbrock J. O., Morris M. R., Wilson A. D.: User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2009), ACM, pp. 1083–1092. 1, 2