User-defined gestures for Agumented Reality with Smart Phones

Francesca Madeddu
Swansea University
803623@swansea.ac.uk

ABSTRACT
Augmented Reality is a promising technology which opens a universe of new possibilities. However, when introducing new technologies, the first step should be understand- ing the effect on the user interaction. Therefore we performed a guessability study to elicit end-user gestures for selection, zoom and rotation tasks with mobile phones on 15 participants. Findings from the experiment allowed to define a gestures taxonomy for AR reality environments.

INTRODUCTION
During the last decades new paradigms of interaction emerged along with the development of new technologies. It was only the 1963 when Ivan Shutherland developed the revolutionary *Sketchpad*, a system which used a light-pen to create and manipulate objects in engineering drawings [9], which is now considered the ancestor of modern CAD programs; only ten years later XEROX introduced the WIMP paradigm of interaction, later popularized by the Macintosh in the 1984 [11], which is now so deeply coded in our minds that it would be difficult to think of any other way to interact with elements of a desktop environment.
Examples of latest technologies include virtual (VR) and augmented reality (AR). These are not modern in the sense they have been invented recently, in fact the *Sensorama* machine [3], developed by the cinematographer Morton Heilig around the 1962, can be already considered as an good example of immersive, multi-modal technology implementation. They are considered modern because thanks to factors such as the higher processing power now available, the consequent prices drop, the spread of mobile devices and the Internet connection widely widespread, they can be finally exploited efficiently: today, even an amateur programmer can in fact develop a VR 3D application just using a smart phone mounted on a *Google cardboard*[1], a low cost headset equipped with two lenses to create the impression of a stereoscopic 3D image.
It is not a case that the biggest world IT companies are lately investing is AR products such as the *Google Glasses*[2], a wearable technology with an optical head-mounted display definition, and the Microsoft *HoloLens*[3], a smart glasses unit that is a cordless, self-contained Windows 10 computer.
New technologies demand for new paradigms of interaction, which aim to reduce the distance between the real world, where the user lives, and the physical implementation of the technology, where the digital information is encapsulated. The concept is not new, and it was firstly introduced by by Ishii with *Tangible Bits* [1]: the idea of Ishii's TUIs (Tangible User Interfaces) was to make the digital content tangible, which is translated in a coincidence of input and output spaces, where the input becomes space-multiplexed, meaning that multiple users can simultaneously manipulate the digital information.
As new technologies are introduced, a universe of new possibilities emerges, and it becomes important to ask ourselves: How can we use technology at its best? How can we make it efficient and effective for people? How can we use it to get the best out of people, supporting their natural capabilities?
Too often we forgot the simplest UCD practices, and we make the mistake of demanding people to be shaped towards the new technology instead of its contrary. Or even worse, we forgot that new technologies require new paradigms of interaction, and we trail old paradigms in new technologies, sentencing them to fail.

Therefore the purpose of this paper is investigating on user interaction with smart phones in AR environments. In particular here are presented the results of a *guessability study* performed involving 15 participants, designed to: 1) define a gesture taxonomy with a specific focus on selection, rotation and zooming tasks, 2) uncover insights on participants' mental model, 3) delineate implications for the development AR applications.

RELATED WORK
Research in AR firstly involves tracking techniques (20%) and secondary, with the same percentage (14.5%), interaction techniques, calibration and registration and AR applications [16].
Within the interaction techniques, for of the purpose of the current research, we distinguish between three main categories: *surface gestures* are those performed on the surface of the device which captures the interaction through the touch screen; *motion gestures* are those performed by moving the device, which captures the interaction through sensors such as accelerometer or gyroscope: finally, *direct manipulation gestures* are those performed directly on the virtual object and they usually require tracking techniques to recognise the user's fingers. The current research focuses on all the three categories of gestures.

[1] https://www.google.com/get/cardboard/
[2] https://developers.google.com/glass/
[3] https://www.microsoft.com/microsoft-hololens/en-us

Guessability

Guessability "is essential for symbolic input, in which users enter gestures or keywords to indicate characters or commands, or rely in labels or icons to access features" [13]. A high guessability is crucial for certain domains such as games: in fact, when playing, we usually face challenges that can be overtaken in only one way. For example, imagine a game where the user can reach the next level only when the main character jumps up to a certain hight: if the user cannot guess what kind of action he needs to perform on the system to make the character to jump, he is unable to carry on the game.
The problem is well known in the graph theory where the action goes under the name of *bridge*: a bridge affects the graph connectivity, in fact if deleted the graph will be disconnected [10]. If we think of an interactive system as a graph, where each vertex is a state and each arc is an action that the user can perform to change the system's state, disconnecting a graph means loosing the chance to reach a part of the graph: specifically, in the case of a game, it means that the user will no longer be able to reach the next level, hence to win the game.
Guessability is such an important characteristic for symbolic input that Wobbrock defined a methodology to calculate not only the guessability over a certain symbol but also the level of agreement. Several researches recognised the importance of the guessability methodology and applied it within the gestures' research to elicit symbolic input from end-users.

Eliciting end-user input

The importance of eliciting gestures from end-users is underlined by the work of Morris et. al who performed a study to compare two gesture set for interactive surfaces, one created by an end-user elicitation and one created by a group of HCI experts [2]. Results showed that participants preferred the first set of gesture as the HCI ones were usually perceived as "physically and conceptually complex" [2] .

The end-user elicited gestures used for the study came from a previous work of the same research group, which aimed to investigate on surface gestures with a specific interest for surface computing [14]. Study findings underlined that users usually prefer a simple interaction using only one finger, that the previous learned idioms affect the user mental model and that for some commands there is a so little agreement that the use of widget is highly recommended.
Another attempt to define a surface gestures taxonomy eliciting from end-users, but with a specific focus on multi-touch selection for interactive data graphics, was done by Willet et al. [12]. Once again, the analysis of the 720 user-elicited gestures highlighted the participants strong preference for simple one-hand gestures.

Moving from surface to motion gestures, Ruiz et. al narrowed their research on mobile interaction [7]. Their motivation arose by the consideration that, despite the fact that nowadays modern smartphones are equipped with a wide variety of sensors, which would allow sophisticated new ways of interaction, still a little has been done in this direction. Therefore their work towards a definition of a taxonomy of motion gestures proved that there can be a consensus that can be exploited to influence the best practices in designs.

Finally, other researches focused on direct manipulation gestures. For example Piumsomboonet al. tried to defined a direct manipulation gesture taxonomy for AR using a head mounted display [6]. A total of 800 gestures, captured during their study, allowed to identify a set of 44 user-defined gestures. However, because of the insufficient level of agreement between some of the gestures, they invited for a further investigation.

DEVELOPING USER DEFINED GESTURE SET

All the cited studies tried to define a taxonomy within a certain category of gestures: surface, motion or direct manipulation. However, as mentioned in the previous section, for some applications such as games, the high level of guessability and the high agreement is a key factor. Therefore in the current research gestures have been elicited from participants without distinction between categories.

Users

The study involved 15 participants, mostly men (60% of male against 40% of female) and in an age range of 19 and 40 years old. The 93% of participants were students, of whom 73% from the computer science department, hence with a strong technical background. In fact, 93% of them included themselves in a range from normal to expert mobile user and all of them declared to own at least one mobile device. The same percentage also declared that, when using mobile devices, they were usually able to think of the action which is needed to achieve a certain goal and to understand the given feedback, concepts formalised by Donald Norman as execution and evaluation gulfs [5]. Finally, even if the 40% knew what AR was, only the 26% had tried an AR application.

Data gathering

An initial structured questionnaires was used to gather demographic data and technological skills towards mobile technologies and augmented reality. This information was useful to provide a context to the the experiment's results.
During the experiment the researcher used a structured data sheet to jot down observations about the participant's interaction with the mobile phone, interaction which was also partially recorded by the preinstalled logging application. In particular the logging application registered surface gestures rather than the motion ones. The choice was justified by the fact that, being possible to graphically represent surface gestures as lines, they provided the researcher with a useful way to visually interpret the gathered data; motion gestures, instead, consisting in raw values coming from sensors such as gyroscope and accelerometer, are very hard and time consuming to be analysed, and they don't provide any immediate visual representation that can help the researcher to make sense of the collected data. Therefore the effort required to translate row data in meaningful information would not have been justified.
Finally, a semi-structured interview was used to gather information about the participants opinion on the performed gestures. All the stages of the experiment, where the participant

gave the consent, were audio-recorded and participants were asked to adopt a talk-aloud protocol.

Apparatus

The study was performed using a custom developed application installed on a Samsung Galaxy Note 3, measuring 151.2 mm x 79.2 mm x 8.3 mm, set to a 1080 x 1920 of resolution and running Android 5.0. The custom application, developed using Unity3D[4] and the Vuforia[5] external library, aimed to allow and record the interaction between the user to and the AR environment.
Each surface gesture was captured as a text file, a sequence of points for each fingers associated with timestamps to solve possible ambiguities, and as a image, a group of coloured lines where each colour represent a different finger (figure 1). Additionally, each time the participant touched the screen, a screenshot was captured: this was essential to give a context to the captured gesture.
The target object consisted instead in a rectangular box with stitched on the top the target image, that is the image which, once captured by the camera, triggered the mechanism to display the AR.

Figure 1. Images captured for a surface gesture. On the left the image showed by the phone's camera viewer at the screen touch, on the right the two coloured lines corresponding to the two fingers gestures.

Tasks

Participants were assigned with four main tasks. Given a scene displayed by the mobile phone's camera, augmented with a 3D model of a bag, and a 3D model of an Anubis statue, they were asked to perform the action they felt was appropriate to cause a specific effect, that was selecting the Anubis statue to drag it inside a bag, zooming on it and rotating it. Finally they were asked to explore autonomously the scene.

Procedure

The study took place in three different days and places, but under the same conditions. After having dealt with the consent procedure, the participant was delivered with a task description paper, to introduce him with the AR concept and the experiment procedure. Once the participant finished reading, they were given a questionnaire to complete. As the questionnaire was filled, the participant was given the smart phone and asked to perform the four tasks. Because investigating on the target object recognition lied outside the scope of the experiment, the researcher showed the participant how to make the AR appear the first time. Finally, the researcher proceeded with the final interview.

[4] https://unity3d.com/

[5] https://www.qualcomm.com/products/vuforia

During the whole process, which last from 15 to 30 minutes, the participant sat next to the researcher in front to a rectangular desk, with the target object positioned in front of them at a reachable distance.

Biases mitigations

When designing the experiment, several problems needed to be addressed to mitigate possible biases.

The first bias was related to a possible Hawthorn effect: that is the participant could have felt more comfortable when performing surface gestures, rather than motion or direct manipulation gestures, because more socially accepted. However, letting the participant know that he could have touched the target object or move the phone, would have biased the results: one of the experiment aim consisted in fact in understanding if the participant could *guess* by himself that touching the target object and moving the phone was a possible way to achieve to his task.
To solve the problem, during assignment of the task, the researcher moved the phone around and touched the target object, being careful not to reveal the actions feedback, to implicitly let the participant know that any interaction was allowed.

Similarly, another possible bias regarded the words chosen for the task questions. For example, asking a user to "drag an object inside a back" would have easily recalled the desktop the drag and drop action. Therefore words focused on the effect to achieve rather than on the action needed to obtain it. So for example, for the rotation task ,the user was asked to perform the action which would have allowed him to see the other side of the object.

The third possible bias was related to the target object's affordance: a piece of paper suggests less manipulation than a cube. Therefore it was possible that the user would have preferred surface over a direct manipulation gestures, only because of the target object's shape. To mitigate the problem, the target image was stitched on the top of a rectangular box, to improve the perceived performance in terms of manipulability.

Finally, the last bias involved the tasks' order: the more the participant used the application, the more he knew about it, so it was possible that the last task was made easier because of the gained experience. An easy solution would have been a latin square design to randomise the tasks' order. However, during the pilot study emerged that while selection and zoom where usually see as atomic action, when asked to "see the other side of the object" the participant was tempted to perform several action (such as first zooming, then rotate) rather then directly rotate the object. Because of this reason, as the selection appeared during the pilot study as the most straightforward, and the rotation as the most challenging, they were ordered respectively as first and third task.

RESULTS

In the followings sections are reported the 18 gestures elicited for selection, zoom and rotation tasks. Gestures are grouped into categories as surface, motion and direct manipulation;

a *mix* category has been added to include those cases where the participant performed a gesture which emerged from the combination of two different gesture's categories (such as for example moving the phone and the Anubis statue closer to each other at the same time).
Moreover, for each gesture it is indicated if it was the result of the first, second or third attempt: in fact, while some participants were immediately satisfied after the first, others needed more tries before coming to a satisfactory gesture. In particular, it frequently happened that, after the first attempt, participants thought about a new gesture, which was later recognised as the most *standard*: those actions are therefore reported in table as *Std* .

Selection
A total of 8 gestures was identified for the selection task. The same surface gesture resulted from the first attempt of 93% of participants: in fact they all tried the classic drag and drop selection from the Anubis statue to the bag, indicating a strong consensus. Moreover, the same number of participants also identified the drag and drop selection as the standard gesture. On the contrary, two participants tried the drag and drop selection in the opposite direction, meaning from the bag to the Anubis statue; however, this gesture was never tried as first attempt.
In terms of expected feedback, all participants declared that they would have expected the Anubis statue (or the bag) to move following their finger, as if it was anchored to it, therefore following the traditional drag and drop behaviour.

		1st	2nd	3rd	Std
Surface	One finger drag&drop from Anubis to Bag	**14**		1	**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			2	
	Single tap on background		1		
Motion	Move the Phone laterally from Anubis to Bag	1			
	Shake the Phone		2		1

Table 1. ***Selection*** **gestures.**

The remaining 4 surface gestures consisted in single or double tap on different areas of the screen; however participants saw their actions as the single step of a multi-step task. For example, one of the participant recursively tapped on all the screen's corners looking for hotspots and contextual information; similarly, another participant tapped on the bag to open it, to then move the Anubis statue into it, while another one tapped on the Anubis statue looking for a contextual menu with a selection option.

Only two participants tried a motion gesture, moving the phone laterally from Anubis to the bag or shaking the phone. When performing the gesture they verbally explained that they would have expected a direct cause-effect relationship between the phone and Anubis, so that sliding the phone, for example, should have caused Anubis to move towards the same sliding direction.

Finally, none of the participant tried a direct manipulation by touching the target object.

Zoom
A total of 5 gestures was identified for the zoom task. As first attempt, most of participants tried a surface gesture (60%), followed by a still significative number (26.6%) which tried a motion gesture. The remaining part tried mix gestures, such as a combination of movement and direct manipulation (moving the camera and the target object at the same time closer to each other) or a combination of surface and movement (moving the phone closer to Anubis and pinching to zoom at the same time). Once again, if we exclude the mix gestures, none of the participants tried to directly manipulate the target object.

		1st	2nd	3rd	Std
Surface	Pinch to Zoom	**8**	2		**11**
	Double Tap	1	1	1	
Motion	Move the Phone close to Anubis	4	1		2
Mix	Move the Phone + Grab Anubis to get them closer to each other	1	2		1
	Pinch to Zoom + Move the Phone	1			1

Table 2. ***Zoom*** **gestures.**

Two participants pointed out that the most natural gesture could have depended on their proximity to the target object: for close objects they would have preferred to move the phone, while for farer object they would have rather used the pinch to zoom gesture. Nevertheless, results showed a sharp preference for surface gestures, even if in the experiment settings the target object was close to participants.

In terms of expected feedback with surface gestures, the 86.6% of participants expected all the scene (Anubis and background) to be enlarged, however once again their expectation proved to be different from their preference: in fact, one third of participants declared that they would have preferred if only Anubis could have been enlarged, as it was the model, and not the background, the object of their interest.

Moreover the difference between expectations and preferences is underlined by the gesture chosen as standard. In fact, even if only the 53.3% of participants tried the pinch to zoom as first attempt, the percentage goes up to 73.3% when they needed to define the standard gesture.

Rotation

A total of 5 gestures was identified for the rotation task. Considering categories, gestures were fairly divided in a 40% for direct manipulation, 33% for motion and 22% for surface gestures. Surface gestures consisted mainly in a transposition of the rotation concept on screen: one or two fingers were used to draw a semi-circle shape. Clock and anti-clock wise gesture were grouped in the same one, under the assumption, mostly confirmed verbally by participants, that the gesture direction had a direct mapping on the rotation direction. However, for further studies, it would be recommended to suggest the participant a specific direction, so that the mapping could be further investigated.

		1st	2nd	3rd	Std
Surface	One finger circle rotation	2			1
	Two finger circle rotation	2	1	2	2
	One finger straight line	1			1
Motion	Move the Phone around Anubis	4	1	1	4
Direct Manipulation	Rotate Anubis	**6**	2		**7**

Table 3. ***Rotation*** **gestures.**

In terms of motion and direct manipulation gestures, participants actions were straightforward: in the first case they moved the camera around the target object, while in the second case they grabbed and rotate it. It has to be noted that, while a bias mitigation was provided for the direct manipulation, since the target image was stitched to a rectangular box to improve its graspable affordance, no mitigation was implemented for motion gestures. That is, being the participant sat on the side of a square table, it was inconvenient for him to physically move the phone around the target object. For any further study it is suggested to use a round table, and to invite the participant to stand up.

Overall, the direct manipulation of the target object was, not only the gesture mostly tried as first attempt, but also the one later recognised as standard.

First attempt, Standard and Preferred Gestures

During the experiment an additional gesture classification emerged. When the participant was given with the phone and asked to guess the right action to cause a specific effect, he performed his *first attempt gesture*, so the first gesture which came to his mind. However it frequently happened that the participant did not stop trying, gradually discovering new gestures which, in some cases, were defined as more satisfactory than the first one, hence defined as *preferred gestures*. Finally, the *standard gesture* was the one that, according to the participant opinion, most of people would have tried when asked to achieve a specific task.

For example, for the zoom task, several participants tried to move the phone close the Anubis as first attempt gesture, and they recognised it as their favourite; however, when they were asked about their opinion on which kind of gesture other people would have done, some of them chose the pinch to zoom surface gesture (figure 2).

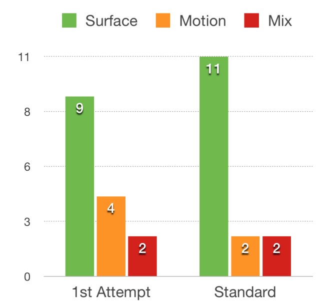

Figure 2. ***Zoom*** **gestures grouped by category.**

Similarly, for the rotation task, it was noticed a difference between first attempt and standard gestures (figure 3).

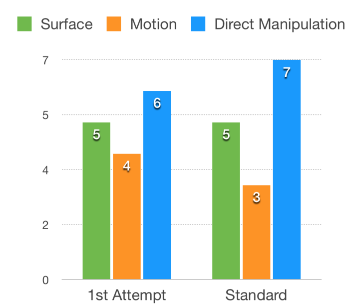

Figure 3. ***Rotation*** **gestures grouped by category.**

Finally no difference between standard and first attempt gesture was declared for the selection task, which underlines the strong consensus toward the elicited gesture (figure 4).

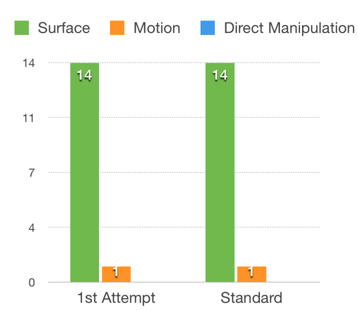

Figure 4. ***Selection*** **gestures grouped by category.**

Guessability and Agreement

Wobbrock defines the *Guessability* G as:

$$G = \frac{\sum_{s \epsilon S} |P_s|}{|P|} * 100\%$$

where P is the set of proposed symbols for all referents, and P_s is the set of proposed symbols using symbol s, which is a member of the resultant symbol set S [13].
Moreover Wobbrock defines the *Agreement* A as:

$$A = \frac{\sum_{r \epsilon R} \sum_{P_i \subseteq P_r} \left(\frac{|P_i|}{|P_r|} \right)}{|R|} * 100\%$$

where r is a referent in the set of all referents R, P_r is the set of proposals for referent r, and P_i is a subset of identical symbols from P_r [13].
So for example the A value for the selection task would be:

$$A = \left(\frac{8}{15}\right)^2 + \left(\frac{4}{15}\right)^2 + \left(\frac{1}{15}\right)^2 + \left(\frac{1}{15}\right)^2 + \left(\frac{1}{15}\right)^2$$

In figure 5 are showed Guessability and Agreement values for the most performed gesture of each task. The selection gesture, a drag and drop 1 finger surface action from Anubis to the bag, is characterised by high guessability and agreement. The zoom gesture, the surface pinch to zoom, is characterised by a lower value of guessability, but it still is slightly over the 50% threshold; the agreement over the gesture, however, is only the 36.8%. Finally, the rotation gesture, a direct rotation manipulation on the target object, is characterised by a even less certain guessability and agreement, respectively 40% and 27%.

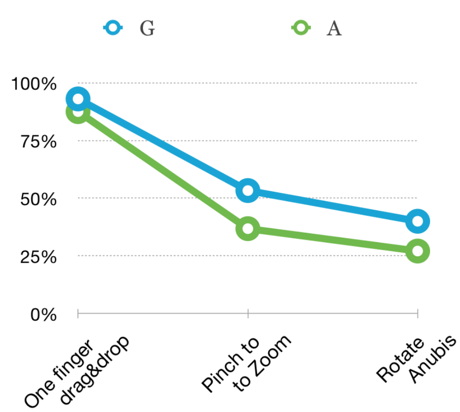

Figure 5. Guessability and Agreement values for the elicited gestures.

DISCUSSION

In the following sections quantitative results are discussed in the context of the qualitative findings captured thanks to the think-aloud protocol during the experiment, and during the final interview.

A Dual Participants Approach

In general it was possible to different group of participant which approached the interaction in a different way.

For a group of participants, as already highlight by Wobbrock [14], the interaction seemed to be strongly influenced by their background. Therefore, because touch is the primary mobile phone modality, they try to interact with the scene through surface gestures: in particular they applied the most common gestures, such as the drag and drop and the pinch to zoom, to the AR environment.
For example one of the participant, when asked if it was easy to think about a gesture to achieve the selection task answered:

> *Yes, it was easy but I always actually think of the previous experience [...] it's like simulate what we are doing when we use a computer when you put a file into a folder and you drag one into each other.*

In some cases participants expressed the concern that interacting with the target object could be complicated: one participant mentioned the possibility that the target object could have been fixed somewhere, while another was concerned that the target object could have been too heavy to be moved. However it has to be noted that in the environment settings the target object was close to the participant, light and easy to manipulate. All these observations provide an explanation for the preference towards surface gestures over the others. Moreover some people stated to prefer surface gestures because, being the mobile phone already in their hands, were more immediate (assumption confirmed by the Fitt's law).

For the other group of participants, even if they recognised that it was easy to think about surface gestures to achieve the given tasks, movement and direct manipulation gestures were preferred. The given reason was that they perceived the AR as equal to the reality, therefore they felt that the interaction should have been as real as possible. For example one of the participant stated:

> *I like the idea of augmented reality and I think that the interaction should be augmented based [...] (to interact) shouldn't be difficult because it is the same of the real world [...] I think this is the point of augmented reality.*

The contrast between the two approaches emerged with the rotation task. As a participant stated:

> *The zoom and the selection, it's something we are used to every day, the zoom with the camera for example, but the rotation is not so common so maybe other people will try different gestures.*

Needs of consistency, coherence and control

In general, once that the scene was accepted as real, consistency was required. For example a participant tried for few minutes to move Anubis around the table to see if the lighting of the object was consistent with the lighting of the room. The same participant also moved a pen first between the camera and the target object, then behind the target object to check, again, if the consistency hold. Another participant instead stated that he would have expect Anubis to fall down from the rectangular box if he tried to till it.
Another interesting observation was that, while when interacting trough common gestures (pinch to zoom, drag and

drop) participants seems to suspend they critical judgment, they look for coherence and consistency when they had to think about a custom gesture. For example, participant weren't disturbed by the fact that pinch to zoom and drag and drop were 2D gestures, but caused a 3D effect; however when trying to achieve the rotation task with surface gestures, most of participants underlined the incongruence.
Along with the desire of consistency and coherence, participants often expressed the of control. For example several participants stated that they would have liked to be able to make the augmented object disappear once they had enough of it.
These findings are not surprising, in fact control and consistency are basic usability requirement suggest by both the Nielsen's 10 Heuristics [4] and the 8 Schneiderman's Golden rules [8].

Abstract and Concrete gestures

A further classification on gestures can be made according to their abstract or concrete nature. For example, shaking the phone to achieve selection is an abstract gesture as the relationship cause-effect is not clear whereas moving the phone around the object or rotating the target object itself, are a concrete gestures, in the sense they emulates the action we would perform in a real world.

Ambiguous gesture and alternative interaction

Some elicited gestures caused a conflict. For example, the double tap was used both to achieve zoom but also selection. In general, it suggests that, when a gesture does not share a high agreement, it is better to provide the user with extra information, to solve the ambiguity.
During the free exploration task several participants pointed out that they would have expected other GUI elements to appear to help them interact with the scene. For example, for the zoom task often it was hypothesised a *plus* or *minus* button for the zoom-in and zoom-out while for the rotation task it was often suggested the visualisation of the dimensional axis for a manipulation of the object.

DESIGN IMPLICATION AND CONCLUSION

In this paper we presented findings from a guessability study which aimed to elicit end-user gestures for selection, zoom and rotation tasks in AR enviroments with mobile phones. No restriction was imposed in terms of category so that participants were free to perfrom surface, motion or direct manipulation gestures.

Results from the experiment showed that, once a category was individuated, the choice of a specific gesture was straightforward. In particular, for the selection task almost the totality of participants (93%) 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, the general trend looked inverted and for the first time the highest agreement was towards a direct manipulation gesture, that is participants directly rotate the target object. However, the agreement for the gesture was the lowest for the three tasks, in fact only 40% participants tried the gesture in their first attempt.
As highlighted by other studies [14] participants behaviour seemed in general highly influenced by their previous experience, especially where a common gesture already existed for the task. In those cases, the fact that they were interacting with a AR environment did not affect on their capacity to think of a gesture for the task, which was always considered as natural and intuitive. However, when a standard gesture did non exist, as for the case of the rotation task, their behaviour was affected by their critical judgment. In those case emerged a need of coherence and consistency previously ignored.

There are several implication for these findings. Firstly, as the biggest indecision is about which category of gesture the user will choose, and not about the specific gesture within a certain category, when developing an application we should put an effort in two different directions. On one hand attention should be paid towards making the underrated category more visible: that is, if the user is not tempted by the direct manipulation because he is afraid to touch the object, or because he thinks it is fixed, then the target object affordance for manipulation should be increased. On the other hand, because for some tasks, such as the rotation one, agreement is divided between several categories, then an application should probably implement a gesture for each category. Finally, when the agreement is too low, additional GUI element should be provided to help the user to interact with the application. In fact, as Wu et al. state, a system where the learning only refers to gestures, is a system hard to learn [15]. Participants' suggested ideas such as contextual menus, buttons for zooming and axis slider for rotation, are good example of how to achieve the goal.

The current research also confirmed that participants prefer a simple interaction, in fact except one case during the pilot study, where a participant tried a grabbing surface gesture involving all the 5 fingers, participants always interact with the screen using 1 or 2 fingers. During the experiment also emerged that, when no common gesture exists for the required task, participants tend to use their judgment to figure out a new gesture.
Implications are that we should always prefer simple common used gestures over complex new ones. Also, when a common gesture is not already present, the new gesture should be a concrete one.

Finally, participants showed a need for consistency, coherence and control. Implication are that we should pay attention to the relationship between the AR rules and the real world rules. This problem is not new in HCI, and it is appears when dealing with metaphors. For example, in a desktop environment, the basket is positioned above instead of under the desktop: breaking the metaphor is needed or the basket would not be invisible. However, the more the metaphor is close to the reality, the more a breakage confuses the user's mind. As AR is a metaphor for reality, we should be always careful when breaking the rules. Example of design choices that could break the metaphor are: implementing the zoom function in a way which enlarges only the models and not the background or implementing the rotation through surface gestures as we allow a 2D gesture to cause a 3D effect. As in

the desktop case, breaking the metaphor does not have to be considered wrong, but it has to be done consciously and the choice need to be justified.
A deep discussion on relationship and constraints between the reality and AR lies outside the purpose of the current research, but it would be worthy to be investigated.

The next important step for the current research is the evaluation of the elicited gestures on the field. In fact, despite the effort put to mitigate and control biases, there must be recognised that the laboratory environments could significantly have affected on the results.

REFERENCES

1. Ishii, H. Tangible bits: beyond pixels. In *Proceedings of the 2nd international conference on Tangible and embedded interaction*, ACM (2008), xv–xxv.
2. Morris, M. R., Wobbrock, J. O., and Wilson, A. D. Understanding users' preferences for surface gestures. In *Proceedings of graphics interface 2010*, Canadian Information Processing Society (2010), 261–268.
3. Morton, H. Sensorama simulator, Aug. 28 1962. US Patent 3,050,870.
4. Nielsen, J. 10 usability heuristics for user interface design. *Fremont: Nielsen Norman Group.[Consult. 20 maio 2014]. Disponível na Internet* (1995).
5. Norman, D. A. Cognitive engineering. *User centered system design: New perspectives on human-computer interaction 3161* (1986).
6. Piumsomboon, T., Clark, A., Billinghurst, M., and Cockburn, A. User-defined gestures for augmented reality. In *Human-Computer Interaction–INTERACT 2013*. Springer, 2013, 282–299.
7. Ruiz, J., Li, Y., and Lank, E. User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2011), 197–206.
8. Shneiderman, B. *Designing the user interface: strategies for effective human-computer interaction*, vol. 3. Addison-Wesley Reading, MA, 1992.
9. Sutherland, I. E. Sketch pad a man-machine graphical communication system. In *Proceedings of the SHARE design automation workshop*, ACM (1964), 6–329.
10. Thimbleby, H. *Press on: principles of interaction programming*. The MIT Press, 2010.
11. Van Dam, A. Post-wimp user interfaces. *Communications of the ACM 40*, 2 (1997), 63–67.
12. Willett, W., Lan, Q., and Isenberg, P. Eliciting multi-touch selection gestures for interactive data graphics. In *Short-Paper Proceedings of the European Conference on Visualization (EuroVis)*, Eurographics (2014).
13. Wobbrock, J. O., Aung, H. H., Rothrock, B., and Myers, B. A. Maximizing the guessability of symbolic input. In *CHI'05 extended abstracts on Human Factors in Computing Systems*, ACM (2005), 1869–1872.
14. Wobbrock, J. O., Morris, M. R., and Wilson, A. D. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2009), 1083–1092.
15. Wu, M., Shen, C., Ryall, K., Forlines, C., and Balakrishnan, R. Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. In *Horizontal Interactive Human-Computer Systems, 2006. TableTop 2006. First IEEE International Workshop on*, IEEE (2006), 8–pp.
16. Zhou, F., Duh, H. B.-L., and Billinghurst, M. Trends in augmented reality tracking, interaction and display: A review of ten years of ismar. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, IEEE Computer Society (2008), 193–202.