NARI: Natural Augmented Reality Interface
Interaction Challenges for AR Applications

Rui Nóbrega[1,2], Diogo Cabral[3], Giulio Jacucci[3,4] and António Coelho[1,2]
[1]*DEI/FEUP, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal*
[2]*Instituto de Engenharia de Sistemas e Computadores, INESC TEC, Porto, Portugal*
[3]*Helsinki Institute for Information Technology HIIT, University of Helsinki, Helsinki, Finland*
[4]*Helsinki Institute for Information Technology HIIT, Aalto University, Espoo, Finland*

Keywords: Augmented Reality, Computer Graphics, Interactive Applications, Multimedia.

Abstract: Following the proliferation of Augmented Reality technologies and applications in mobile devices it is becoming clear that AR techniques have matured and are ready to be used for large audiences. This poses several new multimedia interaction and usability problems that need to be identified and studied. AR problems are no longer exclusively about rendering superimposed virtual geometry or finding ways of performing GPS or computer vision registration. It is important to understand how to keep users engaged with AR and in what occasions it is suitable to use it. Additionally how should graphical user interfaces be designed so that the user can interact with AR elements while pointing a mobile device to a specific real world area? Finally what is limiting AR applications from reaching an even broader acceptance and usage level? This position paper identifies several interaction problems in today's multimedia AR applications, raising several pressing issues and proposes several research directions.

1 INTRODUCTION

Mixed and Augmented Reality techniques have vastly improved in the last years. What started as an interesting experiment to introduce polygons and virtual geometry mixed in real footage and videos has now evolved to be a part of several commercial products and multimedia applications. Although there are still open challenges in Augmented Reality (AR) rendering and tracking techniques it is important to start considering how to build meaningful interactive multimedia applications. How to engage users with these systems and make them feel that this is an important technology and not just a visual trick.

This paper addresses the user interaction implications of AR applications stating affordances, problems and challenges while pointing possible solutions for the future.

For years the challenges in AR were solely related with the computer graphics and computer vision fields. In order to introduce virtual objects mixed with the scene it was first required to recognize the real world through machine vision. Virtual objects were associated with real world markers or visual structural features such as walls, floors or ceilings. This was then used to produce interesting prototypes, which used tangible objects as controllers or presented virtual information or 3D content in front of the markers. It was an initial period where AR was scene in Museums or science fairs and was presented as a curiosity.

With the improvement of the graphic capabilities of computers, the superimposed graphics become more realistic and engaging. Publicly available AR applications started to appear in game consoles and computers. The main problem with these systems is that they relied on fixed or attached cameras. This fact limited AR systems to applications where the user interacts through one or more physical objects, the markers, in front of the fixed camera. This means that the application is camera-centered with the user revolving around the computer.

The rise of the smartphone and tablets changed entirely the dynamics of the interaction. These are mobile devices generally equipped with frontal and back cameras, accelerometers, gyroscopes and global positioning system (GPS) support. Additionally most have significant processing and graphical capabilities, which allow them to run complex multimedia applications. Mobile devices

Nóbrega R., Cabral D., Jacucci G. and Coelho A..
NARI: Natural Augmented Reality Interface - Interaction Challenges for AR Applications.
DOI: 10.5220/0005360305040510
In *Proceedings of the 10th International Conference on Computer Graphics Theory and Applications* (GRAPP-2015), pages 504-510
ISBN: 978-989-758-087-1

have the advantage of supporting AR applications in the wild, outdoors or out of controlled environments. In AR mobile applications the interaction is mostly centered in the user's surrounding environment and location. The advent of smart glasses will eventually change the interaction paradigm because the user will be able to use augmented reality seamlessly without having to point a mobile device.

Current AR commercial applications are mainly available for mobile devices, such as mobile phones or tablets, although with some exceptions in game consoles. The general reaction to these applications has been mixed (Piumsomboon et al. 2014). They usually cause a very positive first reaction but are quickly ignored in the long run. This leads us to the main research questions, which need to be addressed about Augmented Reality applications. It is important to define where it is useful to use AR and when it is not, additionally how the interaction takes place, and should AR be the main showcase technology or just a helping technique. Summarizing these are the main open research questions discussed in this position paper:

1. In what locations or events should AR be used? In the street, at home, outdoors or indoors.
2. What kind of AR applications are being used? Which ones are having success? Navigation, board games, advertising experiences?
3. How should interaction take place? How do people hold or wear their devices while using AR applications? What can they do while pointing a camera and holding a device?
4. Where is AR suitable? As a stand-alone graphical application or as a popup feature inside a larger application?

Now that AR applications are becoming more common it is important to have guidelines and metrics to understand what works and what does not. Former surveys such as Zhou et al. (Zhou et al. 2008) have fairly summarized the state-of-art of the AR technology, it is important now to study the interaction implications of such systems in order to create better, improved graphical and multimedia applications. In this paper we discuss the concept of Natural Augmented Reality Interfaces (NARI) based on Natural User Interfaces (NUI) and creating a parallel to the Tangible User Interfaces (TUI). The next sections will provide research directions that we believe should be followed in the future.

2 AR APPLICATIONS

Mixed and Augmented Reality systems have been around for many years (Azuma 1997). The first experimental prototypes date back from the 1970s but they lacked the graphical processing power to effectively implement the concept. In the last decade several commercial applications emerged in game consoles (e.g., EyePet[1]), design (e.g., Atelier Pfister[2]), and GPS navigation (e.g., Layar[3], Junaio[4]) or advertising. Augmented Reality (AR) has imposed itself mainly in non-critical environments but there is still a large discussion about its industrial applications (Fite-Georgel 2011). There are currently several frameworks such as ARToolKit[5], Vuforia[6] or Metaio[7], which enable the development of mixed and augmented reality applications using different techniques.

Every augmented reality system has to essentially solve two problems: registration/tracking and superimposition. In the registration phase it is required to acquire the location where the virtual information will be placed. Typically GPS/compass or feature processing are used. After the registration, the virtual information has to be rendered correctly in 2D or 3D (using pose estimation in 3D) in the correct spot. Depending on the type of registration and devices (hand-held or headset) the interaction with the AR system will be different. There are many different types of AR interactive applications. Some (Takeuchi and Perlin 2012) augment the reality by deforming it and giving more screen space to important objects such as important sites and monuments.

2.1 Sensor and GPS based AR

The simplest AR systems use the GPS position and the compass to track the location where the information will be rendered. The virtual information is usually displayed using the "bubble metaphor" (Takeuchi and Perlin 2012), where the size of the bubble depends on relevance and the distance of the information to the user. This means that several labels located at a certain geo-referenced place will be fighting for attention in the screen thus

[1] Eye Pet, Playstation game, http://www.eyepet.com/.
[2] Atelier Pfister, http://www.atelierpfister.ch/app.
[3] Layar, GPS AR application, http://www.layar.com/.
[4] Junaio, GPS AR application, http://www.junaio.com/.
[5] ARToolkit, http://www.hitl.washington.edu/artoolkit/.
[6] Vuforia, AR framework lib. https://www.vuforia.com
[7] Metaio, AR framework, http://www.metaio.com

creating a confusing interface for the user. Grasset et al. (Grasset et al. 2012) propose a solution, which simplifies the visual clutter of the labels by positioning them in more visible places according to the content of the image. Examples of this type of interaction are the previously mentioned AR platforms Layar and Junaio. Cabral et al. (Cabral et al. 2014) use a zoom mechanism for virtual elements for better visualization of distant augmentations. Others AR apps use special hardware such as head-mounted displays (e.g., Google Glass[8], Epson Moverio[9]) or accelerometers to detect the floor. The previously mentioned furniture design application Atelier Pfister[2] is an accelerometer based mixed reality application for furniture testing. The orientation of the floor is detected using the device accelerometer and the scene scale is adjusted by comparing the scene with a human figure.

2.2 Feature based AR

Another approach widely implemented is using fiducial markers (e.g., ARToolKit) or QR-codes (Turcsanyi-Szabo and Simon 2011) to locate where the virtual information should be displayed in the image. They usually have very distinctive characteristics such as: binary color system, simple geometric form (e.g., quadrilaterals, circles) and known physical size. The main disadvantage of marker systems is the need to introduce a marker in the captured scene, meaning that the user has to have/print a marker and necessarily has to have access to the scene area. Alternatively, the user sees the marker in the world and has to get the specific application to see it augmented.

Markers are ideal for augmented reality because they are simple to track, they can be used to setup a pose estimation algorithm to insert 3D content and the physical size of the marker can be linked to the scale of the virtual object. There are inumerous examples of projects using ARToolKit markers (Sukan et al. 2012; Myojin et al. 2012).

Most recently feature based systems have replaced ARToolKit style markers by image markers (Uchiyama 2011; Tillon and Marchal 2011). Metaio and Vuforia are two software libraries (previously mentioned) for smartphone development, which have successfully commercially explored the concept of image markers for augmented reality. An interesting example is the Ikea 2014 catalogue application[10] using AR furniture models which can be introduced in the scene by using the cover of the catalogue as an image marker. Several applications are being developed to augment paper publications such as magazines (Nguyen et al. 2012).

2.2.1 Scene Analysis

The most recent developments in AR are systems that do not dependent so much on markers and use instead scene analysis and visual element detection.

Using several images descriptors such as SIFT, SURF or FAST or 3D point clouds, several applications (Wagner et al. 2010) have been developed to track features in real-time in smartphones. These applications track different elements instead of visible markers such as planar spaces (Simon 2006) or pre-programmed images (Tillon and Marchal 2011; Nguyen et al. 2012).

The PTAM (Klein and Murray 2007) project automatically detects a plane in the scenario. This is achieved by translating the camera sideways. Using the structure from motion, the 3D scenario is acquired. Using the detected plane, virtual applications can take place on that plane, in the real world. The PointCloud[11] is a framework, which improves the PTAM concept by providing a library to create augmented reality applications in the user real space. The Ball Invasion[12] game is based on a similar system. It has a short tutorial with instructions and animations to illustrate how should the main plane of the game be acquired.

Additional analysis of the scene main lines is important to add other functionalities such as snap to line and automatic alignment of objects (Del Pero et al. 2011). Karsh et al. (Karsch et al. 2011) detect the scene structure from a single image by using a human assisted method where the user is constantly asked to refine the detection by annotating geometry and lights. Other alternatives, based on feature detection in single image (Nóbrega and Correia 2011) are presented by Gupta et al. (Gupta et al. 2011), where the scene layout is detected and free space is studied for the introduction of human models. Liu et al. (Liu et al. 2008) propose an AR system for videos which is based on scene transitions analysis.

8 Google Glass, http://www.google.com/glass/.
9 Epson Moverio BT-200, http://www.epson.com/cgi-bin/Store/jsp/Landing/moverio-bt-200-smart-glasses.do

10 Ikea 2014 catalogue,http://www.ikea.com/ms/en_AA/customer_service/catalogue/catalogue_2014.html
11 Pointcloud, 13thLab,http://pointcloud.io.
12 Ball Invasion, http://13thlab.com/ballinvasion/

3 AR ISSUES AND SOLUTIONS

The main pressing problem in augmented reality is making the general public aware of its existence and how it works. As part of a larger evaluation of an AR application one of the questions assessed the awareness of the public to mixed and augmented reality and to AR applications. The study took place in a classroom during an open science day in a University. For this reason, most of the 81 users were high-school students (80%, with ages between 15 and 19) and were visiting the campus. Before testing an AR application the students were asked the question stated in Table 1. In this young age the users, who were visiting a technology-related university, were mostly very tech-savvy. 83% answered that they possessed or had access to a smartphone or tablet. Even so, the results seen in Table 1 and the follow-up interviews show a high-degree of confusion. Many users had vaguely heard about Augmented Reality and AR apps but most admitted in the interviews that they never tried one. This may mean that AR applications have not yet reached a minimal critical mass capable of generating a follow-up movement and that there is still space for growth in AR.

Table 1: "I understand and I am confortable with the Augmented Reality concept."

	Disagree				Agree
	1	2	3	4	5
#	3	9	31	25	13
%	3.7	11.1	38.3	30.9	16

There are currently hundreds of AR applications, particularly for mobile devices, as stated in section 2. But there are a few areas where we believe that AR is especially suited for. We are currently studying and developing prototypes in the following areas:

1. **GIS and Tourism**: tourists are always curious to find more information about the destination where they are and AR may complement location-based information;
2. **Live Games and Gamification**, physical board or location games have huge potential because of the multimedia additional feedback AR can provide. Additionally on-site learning through gamification can take advantage of AR;
3. **Media Contextual Information and Advertisement**, providing information in the context of a location or a brand. Advertising and marketing are some of the main drivers of AR.

3.1 Graphics and Vision

Before discussing the interaction aspects of AR application it must be stated that there are still open questions graphics and vision techniques. Most vision recognition techniques are focused in finding real-world markers. One of the main challenges is to additionally detect the 3D structure surrounding the scene and making virtual objects interact with them. As an example, virtual objects should get occluded by real objects, labels and information should be pinned to walls and floor automatically and real world properties should apply to the virtual application (Nóbrega and Correia 2011). Additional integration realism between virtual and real world imposes several challenges in light source detection and reflection (Karsch et al. 2011), texture matching and shadow rendering (see Figure 1).

Figure 1: Example of developed image marker-based AR application simulating shadows of a small tree.

3.2 Interacting with AR

Following on the idea of creating an AR specific branch of NUI, Natural Augment Reality Interfaces should be a research area devoted to the study of interaction and guidelines for future AR development. Current AR applications are designed to be used outdoors (navigations and location finding) and indoors at home or public spaces (e.g., museums). Outdoors applications face problems (Jacob and Coelho 2011) such as GPS and compass accuracy errors and difficulty in visualizing the screen due to low contrast in sunlight. One of the main problems is related with the mobile device holding position. Users have to point their camera devices forward and hold them in an unnatural way with their arms stretched. This is something that, if repeated or prolonged, can lead to fatigue. One solution (without resorting to glass-based technology) would be to use a map to help the user approach a certain destination and then bring the AR interface only when the user is close enough or when the s/he deliberately presses an AR button. An alternative AR approach that we propose would be to create devices with top facing cameras so that users can look down naturally to the device while pointing the camera to the world.

Indoor applications have a much more controlled environment with smaller spaces and controllable illumination. They are ideal to create marker and feature based AR systems. Feature based systems such as previously mentioned Ball Invasion[12] or Atelier Pfister[2] can bring a large novelty to applications because they can be used in any non-predefined space. This brings a novelty factor that can be enacted in multimedia interactive productions for entertainment, gamming and advertisement. Without a physical marker games (Jacob et al. 2012) can be played anywhere taking advantage of different detected real world objects. One of the problems with these systems is that they are sometimes harder to initialize requiring a tutorial to capture a small video of the area or tuning the scale of virtual objects against the real world scale.

Marker-based technology is currently the most reliable form of real-time AR. It has been used in games and advertising but also has several interaction problems. Firstly there is the need to have a physical marker available for the user (e.g., the catalogue cover from Ikea). This is somehow complicated for the users because many times it is required for them to print a certain image. This is also the reason why AR is ideal for known brands; their logos are everywhere and are simple to convert into a marker.

In the real world it is somehow difficult to understand AR image markers. Everyone can recognize a QR-code but it is still difficult to understand if something is an AR marker. The app Augment[13] is an example of an AR browser where physical image markers have a distinctive logo next to them so that users can scan the image and get access to the virtual reality content.

Games and applications based in AR markers also (mostly in indoor applications) have an interaction problem. The user is usually asked to pick a mobile device with their hands and keep it pointed to a certain marker as seen in Figure 1. Additionally if it is an interactive application (e.g., game, or furniture app), users are asked to touch the screen while pointing the device. This is very complicated for most users because they have to hold and point with one hand and touch with the other. The solutions for these problems may rely on designing interfaces that rely solely on thumb interaction or designing a reality freeze feature, as shown in Silva et al. (Silva et al. 2012) for real-time video annotations. This function would allow the user to pause the interaction, bring the device to a rest position (not pointing), complete several actions in the touch interface and then unfreeze pointing the device again to the marker/world.

[13] Augment, http://augmentedev.com

4 FUTURE CHALLENGES

Users are generally very curious about AR technology but after the first impact of surprise they usually (with some exceptions) ignore the applications. Figure 2 presents a chart that represents the current development stage of several technologies.

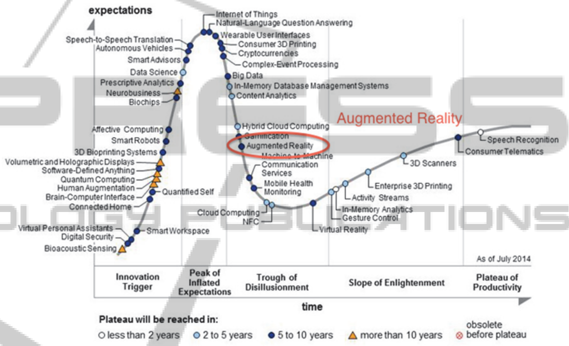

Figure 2: Gartner 2014 Hype Cycle (http://gartner.com/).

The chart assumes that every technology passes through five stages of public perception and acceptance. It can be seen that Augmented Reality after an initial hype, mostly related with the appearance of ARToolKit, is currently in the Trough of Disillusionment. This means that the technology has matured but the interest in it has diminished. What also can be observed is that it can enter the Slope of Enlightenment at any moment with the increase of reliable applications, which deliver consistent and useful results. Currently, new tools, new devices (hand-held and headsets), libraries and software development kits are having some commercial success (e.g., Vuforia). These developments will increase the opportunities for the development of interactive AR applications making this technology probably enter the Slope of Enlightenment.

Mixed reality applications will also improve with new solutions for the living room where televisions have cameras or are associated with game consoles, which have high-definition cameras. These interactive systems will probably have a larger impact in entertainment and gaming applications while mobile applications can have additional impact in advertising products and physical spaces.

The graphical visual of AR has evolved dramatically (as seen in Figure 1) and the level of recognition of markers and 3D structures from video as been increasing in performance and accuracy every year (Zhou et al. 2008).

To conclude, the purpose of this position paper was to raise questions about how AR applications are created and how do users interact with them. For this reason it is important to study and focus the attention of the graphics and vision community to the user interaction aspects of AR. This is a necessity so that in the future we continue to see AR graphic applications being able to be used in even more scenarios.

ACKNOWLEDGEMENTS

The Media Arts and Technologies project (MAT), NORTE-07-0124-FEDER-000061, is financed by the North Portugal Regional Operational Programme (ON.2 – O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF), and by national funds, through the Portuguese funding agency, Fundação para a Ciência e a Tecnologia (FCT). Research also funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n°601139 CultAR (Culturally Enhanced Augmented Realities).

REFERENCES

Azuma, R., 1997. A survey of augmented reality. *Presence-Teleoperators and Virtual Environments, MIT Press*, 4, pp.355–385.

Cabral, D., Orso, V., El-khouri, Y., Belio, M., Gamberini, L. and Jacucci, G.m 2014. The role of location-based event browsers in collaborative behaviors: an explorative study. In *Proc. of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI '14) 4*. ACM, 951-954.

Fite-Georgel, P., 2011. Is there a reality in industrial augmented reality? In *Proc. of ISMAR'11*. IEEE Computer Society, pp. 201–210.

Grasset, R. et al., 2012. Image-driven view management for augmented reality browsers. In *Proc. of ISMAR'12*. IEEE Computer Society, pp. 177–186.

Gupta, A. et al., 2011. From 3D scene geometry to human workspace. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'11)*. IEEE Computer Society, pp. 1961–1968.

Jacob, J. and Coelho, A. 2011. Issues in the Development of Location-Based Games, *International Journal of Computer Games Technology*, vol. 2011, Article ID 495437, 7 pages.

Jacob, J., Silva, H., Coelho, A. and Rodrigues, R., 2012. Towards Location-based Augmented Reality games. In *Procedia Computer Science*, Volume 15, pp. 318-319.

Karsch, K., Hedau, V. and Forsyth, D., 2011. Rendering synthetic objects into legacy photographs. *ACM Transactions on Graphics (TOG)*, 30(6), pp.1–12.

Klein, G. and Murray, D., 2007. Parallel tracking and mapping for small AR workspaces. In *Proc. of ISMAR'07*. IEEE Computer Society, pp. 1–10.

Liu, H. et al., 2008. A generic virtual content insertion system based on visual attention analysis. In *Proceedings of the 16th ACM international conference on Multimedia (MM '08)*. ACM, pp. 379–388.

Myojin, S., Sato, A. and Shimada, N., 2012. Augmented reality card game based on user-specific information control. In *Proceedings of the 20th ACM international conference on Multimedia (MM'12)*. ACM, pp. 1193–1196.

Nguyen, V. et al., 2012. Augmented media for traditional magazines. In *Proceedings of the Third Symposium on Information and Communication Technology (SoICT '12)*. ACM, pp. 97–106.

Nóbrega, R. and Correia, N., 2011. Design your room: adding virtual objects to a real indoor scenario. In *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*. ACM, pp. 2143–2148.

Nóbrega, R. and Correia, N., 2011. Magnetic augmented reality: virtual objects in your space. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI'11)*. ACM, pp. 332–335.

Del Pero, L. et al., 2011. Sampling bedrooms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'11)*. IEEE Computer Society, pp. 2009–2016.

Piumsomboon, T. et al., 2014. Grasp-Shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality. In Proc. of *ISMAR'14*. IEEE Computer Society, pp. 73–82.

Silva, J., Cabral, D., Fernandes, C. and Correia, N., 2012. Real-time annotation of video objects on tablet computers. In *Proc. of the 11th International Conference on Mobile and Ubiquitous Multimedia (MUM'11)*. ACM, p. n.19.

Simon, G., 2006. Automatic online walls detection for immediate use in AR tasks. In *Proc. of ISMAR'06*. Santa Barbara, CA, USA: IEEE Computer Society, pp. 4–7.

Sukan, M., Feiner, S. and Energin, S., 2012. Manipulating virtual objects in hand-held augmented reality using stored snapshots. In *2012 IEEE Symposium on 3D User Interfaces (3DUI'12)*. Costa Mesa, CA, USA: IEEE Computer Society, pp. 165–166.

Takeuchi, Y. and Perlin, K., 2012. ClayVision: The (elastic) image of the city. In *Proceedings of the 2012 ACM annual conference on Human Factors in*

Computing Systems (CHI '12). Austin, TX, USA: ACM, pp. 2411–2420.

Tillon, A.B. and Marchal, I., 2011. Mobile augmented reality in the museum: Can a lace-like technology take you closer to works of art? In *Proc. of the ISMAR-AMH*. IEEE Computer Society, pp. 41–47.

Turcsanyi-Szabo, M. and Simon, P., 2011. Augmenting experiences bridge between two universities. In *Proc. of ISMAR'11*. IEEE Computer Society, pp. 7–13.

Uchiyama, H., 2011. Toward augmenting everything: Detecting and tracking geometrical features on planar objects. In *Proc. of ISMAR'11*. IEEE Computer Society, pp. 17–25.

Wagner, D. et al., 2010. Real-time detection and tracking for augmented reality on mobile phones. *IEEE Transactions on Visualization and Computer Graphics*, 16(3), pp.355–368.

Zhou, F., Duh, H.B. and Billinghurst, M., 2008. Trends in Augmented Reality Tracking , Interaction and Display : A Review of Ten Years of ISMAR. In *Proc. of ISMAR '08*. IEEE Computer Society, pp. 193–202.