

Affordances in Mobile Augmented Reality Applications

<http://dx.doi.org/10.3991/ijim.v8i4.4051>

Tor Gjørøster
University of Bergen, Norway

Abstract—This paper explores the affordances of augmented reality content in a mobile augmented reality application. A user study was conducted by performing a multi-camera video recording of seven think aloud sessions. The think aloud sessions consisted of individual users performing tasks, exploring and experiencing a mobile augmented reality (MAR) application we developed for the iOS platform named ARad. We discuss the instrumental affordances we observed when users interacted with augmented reality content, as well as more complex affordances rising from conventions from media content, AR and the traditional WIMP paradigm. We find that remediation of traditional newspaper content through the MAR medium can provide engaging, pleasing and exciting user experiences. However, the some of the content still suffers from being shoved onto the MAR platform without adapting it properly. Finally, we discuss what content was most successfully mediated to the user and how the content impacts the user experience.

Index Terms—augmented reality, mobile computing, human computer interaction

I. INTRODUCTION

Mobile augmented reality has become increasingly viable through less expensive (and recently free) systems development kits (SDK) and ever increasing technical capabilities of mobile devices. Wagner et al [1] pioneered "...the first completely stand-alone Augmented Reality client on a PDA." (p. 2) on handheld devices in the early 2000s. Since then stand-alone AR has gradually become more common as computing power on PDAs and smartphones increased.

A mobile augmented reality (MAR) or handheld mobile augmented reality (HMAR) application is in its most basic form a mobile device with a display to show an augmentation and registration technology and software that "...have the following three characteristics: 1) Combines real and virtual 2) Interactive in real time 3) Registered in 3-D". [2]

With the smartphone revolution as a backdrop, we decided to start developing a MAR application for distributing and presenting AR content in a newspaper – Verdens Gang – with a large circulation in Norway. The software we developed in collaboration with the newspaper iterated from concept and prototype, to the stable release now available on the AppStore. Through the iterations, we saw increasing requests for reuse of traditional media content the newspaper already had in its possession. These requests resulted in a final version of the application that supports remediating of content on markers in a newspaper.

The term remediation is coined in Bolter et al [3] book "Remediation – Understanding new Media". *Remediation* can be understood as taking existing content (pictures, videos, etc) adapted to a specific media outlet – for instance a printed newspaper – and remediated through a different medium, in our case an augmented reality application.

Trough a video based user study we have uncovered how remediation of traditional newspaper content has shaped our MAR application and how different affordances of the content itself affect the user experience of the application.

II. RESEARCHING MAR

Augmented reality as a term is broadly interpreted and to some degree diluted across media and academia alike. In this paper, we use Azuma's [2] definition of AR mentioned in the previous section to position this work. We can with confidence claim the ARad application to be augmenting reality as it incorporates Azuma's three characterizing properties.

In this paper, we focus on the interplay between the augmentation (the content) and how the user experiences the remediated content in our application.

This paper contributes to further the understanding of AR as a tool in a mobile multimedia context. Additionally it provides a detailed overview of the design of our application and provides design guidelines theoretically aligned towards affordances in human computer interaction (HCI).

III. MAR AS A MEDIUM

Bolter and Grusin [3] sees augmented reality as a hyper mediated visual space and "...the insistence that everything that technology can present must be presented at one time - this is the logic of hypermediacy." This definition is fitting since augmented reality seeks to enable us to see the world how we want it, with just the right amount of information. Macintyre et al [4] argues strongly for AR in general being a new media experience mainly because of a "...fluid blend of the physical and the virtual, and the inevitable tension between them, offers rich dramatic possibilities that are impossible in any other medium."

In Interactions [5], Bolter et al argues that the field of media studies is not responsible for providing design guidelines that the HCI community. However, as content and the interaction is becoming increasingly intertwined one might look to polyaesthetics [6] to shed light on aspects of new media. Polyastechics suggests media studies should be concerned with contributing knowledge

about the aesthetics of content since the interaction between content and interface creates the user experience.

We believe that *content in relation to the interface provides the complete user experience*. Poor content in an application with excellent user interface delivers a poor user experience and *visa versa*. This is paramount in the ARad application, as the content to some extent becomes the user interface itself.

AR technology promotes the ideal of a pure and fully transparent user experience, where the interaction with the medium is transparent to the user and just the experience of the content remains. Although some transparency is achievable on a handheld device, Rosenblum et al [7] believe "...if AR is to realize its full potential, hand-held form factors, despite much of the hype they are receiving now, simply are not adequate." (p. 445) We agree with Rosenblum to some degree. The handheld form factor can never create the same immersion that a totally transparent AR solution can. On the other hand, even with all its hype MAR has utility until this paradigm shifts.

Pavlik et al [8] gives an overview of the emergence of mobile AR in journalism in general. Content creators for the applications they identified (Aurasma, Layar, Wikitude World Browser, Junaio and Blippar) may benefit from the findings in this paper.

In Virtual Realities, Schmalstieg et al [9] discuss the first mobile augmented reality ad developed in 2007 by HIT Lab NZ for the Wellington Zoo. The application delivered a single static 3D model of a zoo animal augmented on a printed marker in the newspaper. They report that for mobile AR advertisement the "...most challenging aspects have been the content creation and application distribution, not the application programming." In addition, Schmalstieg et al points out, "Most of the published AR research has been on enabling technologies (tracking or displays, etc.), or on experimental prototype applications, but there has been little user evaluation of AR interfaces." [9].

In this paper we hope to contribute further to a field where there has "...to date been very little user-based experimentation in augmented reality." [10, p. 3]. Similarly, Zhou et al [11] call for more practical applications. In a more recent and extensive literature review De Sá [12] gather that "...despite the appeal and the growing number of services and applications, very few guidelines, design techniques and evaluation methods have been presented in the existing literature."

The motivation to write this paper comes from a wish to understand the nature of MAR applications, specifically the affordances in these applications. Accordingly, we aim to describe affordances and specify a set of design guidelines one may apply when *remediating* content on the mobile augmented reality platform.

IV. AFFORDANCES IN AR APPLICATIONS

Our application augments traditional media with virtual content. It enables us to create a new user experience for readers consuming traditional print media. By identifying how our MAR application *affords interaction* with this content, we seek to contribute prescriptive knowledge in the form of design guidelines intended for practitioners and academics alike in the field of interactive augmented reality mobile technologies.

To provide meaningful prescriptive knowledge about affordances we have adopted the sociocultural perspective of Kaptelinin et al [13]. This view, and its borrowing of "web of mediators" from Bødker [14] allows us to look at affordances in AR with a respect to the cultural aspect.

MAR is different from other media channels, and the content chosen for remediation needs to be adapted to the AR interaction paradigm.

Previous experience with different content types needs to be taken into account when designing user experiences for MAR. It requires the understanding of the combination of known usability conventions and conventions of aesthetics directly related to the content we are presenting. In this paper, we aim to identify how we can successfully present the content in an engaging manner on MAR.

The framework proposed by Kaptelinin et al [13] allows us to look at affordances from an individual perspective with regards to culture. Through Kaptelinin et al literature review, they find "A growing number of studies in HCI and related areas call for re-defining the notion of affordances to include social and cultural aspects of human interaction with the world." [13] Kaptelinin et al approach is named mediated action perspective and "is concerned with how humans act in their cultural environments, rather than with how animals act in their natural habitats." [13].

If we allow the discussion about affordances in HCI to include these aspects, we can study IT-artifacts with the cultural aspect in mind. This is fruitful in this study since we are creating an artefact relying on existing conventions from a wide range of interaction and content paradigms. These include the WIMP paradigm, the AR interaction technique, existing media content and smartphone conventions.

Through Bødker [14] they identify auxiliary affordances that take into account "complex relations within webs of mediation". This may be understood as there is often a need to perform indirectly or directly related actions to achieve an outcome. They identify maintenance affordances, aggregation affordances and learning affordances as examples of affordances technology may employ to achieve its intended purpose.

In this study, we will describe the *instrumental affordances* of our application. Further we look at the complex affordances that arise with bringing culture into understanding of applications.

Learning affordances: What steps do the users go through to learn how to interact with an AR application, and how we might improve these affordances.

Maintenance affordances: What needs to be taken care of to allow applications to function as effective mediators (maintenance).

Aggregation affordances: How is the application intertwined with analog and digital artifacts to achieve its outcome. Aggregation affordances illustrate the fact that some applications must be combined with other artifacts to achieve its intended purpose. AR's character is to use the environment to provide augmentations and in this regard we aim to identify the essential aggregation affordances of a MAR application.

In addition, we propose a domain specific affordance - *Remediation affordances*: This affordance encapsulates the affordances associated with the content. Specifically,

how we re-communicate the existing affordances of content through an interface. In our case, it deals with bringing the content from a known interface to a different medium. The act of remediation creates a number of boundaries and as well as new possibilities. The aim is to identify successful remediation affordances for a given medium or interface. This is at the heart of mediated action, taking into account the culturally bound when designing.

V. DEVELOPMENT OF ARAD

To address the problem formulated previously we developed several proof of concept prototypes on a variety of mobile platforms (Symbian, Windows Mobile and iOS) starting fall 2010.

An iterative process with several prototypes led to the application now freely available in the AppStore. In the following sections, we will describe the rationale behind the design choices we transferred to the interface, structure and content of ARad.

A. The structure and interface of ARad

ARad is intended to be a MAR application that enables users to seamlessly experience AR content in print media. The interface should be intuitive and enable the users to move rapidly between the different content types. ARad achieves this through a considerable amount of pre-fetching of data in load-screens. Rather than loading content for each marker, the user can download a collection of content from one content provider and experience it in a newspaper or any other printed medium by turning the page.

ARad is developed as a content browser; this means that we provide a range of content providers or campaigns represented as icons on the first screen. The application architecture is straightforward. When ARad starts up it checks for new content and content-providers. It loads new content-providers and links touchable icons to content. When the user touches an icon, it downloads the corresponding content and loads the media assets into memory while the user see a loading-screen. When the user has touched the play icon, a notification screen informs the user to direct the phone towards a marker. When the user points at a marker associated with the content, the augmentation appears. The user can leave this view by aiming the device at something untrackable and press a X icon. They then return to the screen representing the different types of content and may choose content from a different content provider (Figure 1).

This structure we believe enables the perception of ARad being a content browser for different types print media categorized as either campaigns or content-providers. Campaigns are not restricted to a template format, and commonly has some custom code implemented. Content-providers are identified as media outlets that generate new content rapidly based on templates, in contrast to campaigns that remove their custom content when its life cycle is complete.

VI. CONTENT IN ARAD

The specifications for the interface and content in ARad came from newspaper in Norway (Verdens Gang). The newspaper wanted a way of using existing images, videos, illustrations and 3D material already in their possession. Marshall McLuhan [15] remarked that the content of a

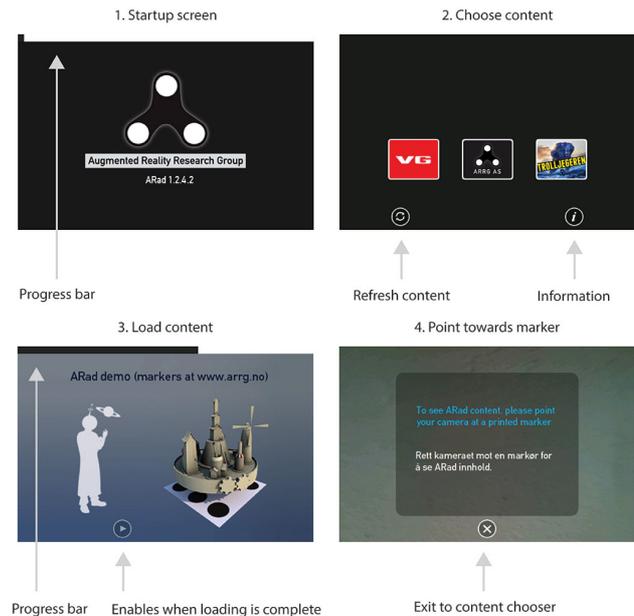


Figure 1. Different screen elements in ARad.

new medium is always the content of another medium. In keeping with this prophecy we inadvertently started the process of *remediation* old media through new media (MAR). Images, videos, illustrations and 3D material related to news or advertisement already in their possession was adapted to a format ARad could display and remediate.

A. Content types

ARad can handle several different types of content using templates. The templates can be updated on the dynamically using an XML-like specification language and supports the quick turnaround in a newspaper. When new content is available we flag it in an index file and when the user reload the content new content is downloaded to the application. ARad supports the following content types:

- Image slideshows with or without audio
- Videos with audio
- Dynamic 3D content with predefined points of interaction
- Custom content

The specification relays paths to marker-, texture-, image-, audio-, video- and mesh-data for the application to load and how to organize and orient it in relation to markers. The custom content needs to be submitted to AppStore for review as it contains executable code.

1) Image slideshows with or without audio

We configure this with a list of images and corresponding single ambient track for background music. Other soundtracks can be tied to an image to provide a narrative or sound effects to amplify the user experience. The user can use arrow icons to maneuver in the image carousel.

2) Videos with audio

The video template points to a MPEG-1 file with an attached audio file. The user can play or pause the video when it is playing on a marker or in full screen. The content is a short interview with a performing artist.

3) *Dynamic 3D content with predefined points of interaction*

The 3D content template loads a set of 3D objects onto a marker and allows tap interaction with the 3D model. The 3D objects have a predefined idle animation with audio. Animations for interaction can be accessed using tap interaction.

All the different content types can be viewed in something we have called full-screen mode. This means that the user can hit a button to view the images, movies and 3D content without the marker in the cameras view, and keep interacting by tapping and dragging the content. This gives the user freedom similarly to the Freeze-set-go interaction method proposed by Lee et al [16].

VII. METHOD

The aim of the evaluations is to provide descriptive knowledge about - how users use the mobile AR application - and prescriptive knowledge about design - in this case a reflection on what works and how to improve what already works - about this IT-artifact in accordance with Iivari [17] types of knowledge attainable from design science. This study utilizes the framework for studying and evaluating augmented reality games described by Gjørseter et al [18]. The framework details a step-by-step approach to conducting a study of AR games. This approach "...works particularly well with AR games since they combine virtual information with the physical environment: while video recordings in addition to screen captures allows the researcher to observe both what happens on screen as well as what happens in the physical space, Think aloud provides insights into the players' problem solving processes." (p. 78). While this study is not concerned with the gaming aspect of AR, we are concerned with what happens in virtual as well as physical space as well as insight into the users experience of the AR system.

A. *Data collection methods*

Think Aloud (THA) [19] is an evaluation technique that takes place during testing; users are instructed to verbalize their actions and thoughts throughout the evaluation. It is a more informal evaluation technique than other common evaluation techniques used in HCI such as cognitive walk-through and heuristic evaluation [20]. Rather than trying to uncover piecemeal design errors as these pragmatic methods excel at, this approach allows us to observe and investigate a uninterrupted user experience through the captured data.

Some previous studies of using think-aloud as an evaluation tool and observational evaluation technique for AR systems can be found. Dünser et al [21] employ think aloud to assess how young children interact with augmented reality books. Liarakapis et al [22] uses think aloud to evaluate and discuss implementation of different interaction methods to AR games.

We chose to do a concurrent Think Aloud, where the participants talked during gameplay. In addition, a retrospective Think Aloud and interview session was performed post session. This session served as a

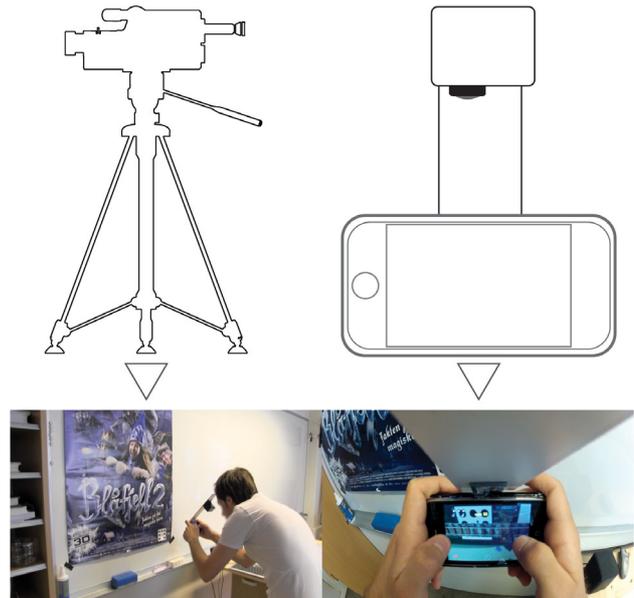


Figure 2. Tripod with custom camera rig

debriefing and allowed users to discuss their overall experience of the content.

B. *Capturing video of mobile augmented reality*

Video-based qualitative research has in recent years increased in popularity within the HCI field. A key advantage to this method is that it can help capture "...aspects of social activities in real-time: talk, visible conduct, and the use of tools, technologies, objects and artifacts" [23].

We used guidelines provided by Heath et al [23] and Gjørseter et al [18] to prepare for, and undertake the video recording. Care should be taken to ensure proper audio capture to facilitate transcription of interaction, clear video of the interaction in proper lighting and correct angles and other technical aspects of video recording. We digitized and synchronized the data in high resolution to allow us to see the participant's interaction with the context as well as the content (Figure 2). Some of the figures in this paper are formatted as cartoon strips. This format supports communication [18] of the activities, as well as the speech occurring during the sessions. The sessions were captured using a small camera attached to a rig at an angle to the handheld device screen. A portable tripod camera was used to film the user's interaction with the tangible printed markers.

We believe the combination of THA and video recordings provide a good foundation for revealing affordances and qualities with the interfaces. The video recordings provide detail about the worldly context and the user interface on the mobile device. In contrast to simply an audio recording of the Think aloud, where you would lose essential data about for instance how they are pointing and orienting the device. In some cases it may also be difficult to ascertain what user interface elements the users are referring just from audio recordings. Additionally, one can use still images, and still images in sequence from the video recordings to directly illustrate the system and underpin interesting findings.



Figure 3. A user exploring 3D content

C. Structure of evaluations

A pilot study with a female primary school teacher aged 25 preceded the final evaluations. The pilot study was conducted to rehearse and uncover problems with the structure of the evaluations. We gathered data about gender, age, profession and familiarity with information technology, smartphones and tables. The final evaluations were conducted with seven participants – five male and two female. With an average age of 29 years (min: 26, max: 34), the participants were faculty members (an accountant, three PhD students, a post doc student) and others (a high-school teacher and an engineer). When asked “How would you characterize your experience with information technology?” on a Likert scale from “1 – Very little” to “8 – Very much” the participants averaged 4.8 (4 min, 6 max). In regards to “How would you characterize your experience with tablets and smartphones?” the participants averaged 5.0 (3 min, 7 max).

We label the evaluators from R1 to R7. The author’s assistant is labeled A, and the author is labeled T in the transcripts. The sessions ranged from 25 minutes to 45 minutes of dual camera video. A free form retrospective think aloud followed each session, lasting between 10 and 30 minutes. The corpus for this study total 51 transcribed dual camera video clips.

During the sessions, the participants experienced six remediated content types.

- A video (Figure 4)
- An interactive image slideshow (Figure 6)
- A 3D interactive castle (Figure 3, 7)
- A 3D interactive troll (Figure 5, 7)
- A game related to the Troll (not described in this paper)
- A game related to a different campaign (not described in this paper)

VIII. FINDINGS

ARad is remediating in nature, and we choose to look at

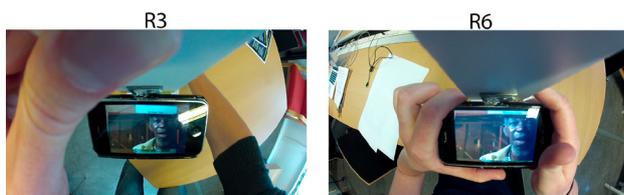


Figure 4. While interacting with 2D content, the users tried to align the content appearing on the marker to iPhone display.

it from this perspective in this study. When disseminating sequences of interaction that relate to the overall experience we may reveal how remediation of content can potentially provide new, better, more fun or engaging ways of interacting with print media.

A variety of other topics on the details of usability is readily available in the data. However, this study focused on the user experience and different affordances of content remediated through a handheld augmented reality application.

We use design research principles [24] in combination with established evaluation methods within the field of HCI to help us communicate and analyze the user experience of the content and interface of ARad through illustrations and quotations in context. The findings below are represented as direct quotes in italic and illustrations from the concurrent think aloud. Quotes are referenced in the text as Q with the quote number following.

A. Video content

During the interaction with the video content, we made observations in regards to how the users experienced this content.

The novelty factor of a digital video being experienced augmented on paper can not be understated (Q1, Q2). These utterances were spoken when the participants viewed the movie on the marker for the first time.

Q1 - R2: “Ah! Now it works. Cool!”

Q2 - R4: “It is attached to the paper there. That was pretty cool.”

Some of the evaluators found it to be unfamiliar to watch movies in this manner and tried to align it the movie naturally to the screen (Q3).

Q3 - R6: “The only thing here is that we need to aim to show the movie in the right size.”

Other participants did not utter this directly, but from the video material we can observe the participants align the video naturally after some time experimenting, and eventually bringing the movie to full screen as illustrated in Figure 4.

Some participants did not mind watching the video at odd angles. Others did not align the video to the display perfectly, but kept the movie in view for long sequences of interaction, and avoided getting the video out of view or at odd angles.

It may be beneficial to show content into alignment when we detect that the user is trying to align any flat content to the screen. Users want to see flat content in full screen mode. It may be some subtle property with AR-video we have yet to determine, but the evaluators seemed to be more excited about AR-video represented in this way than images.

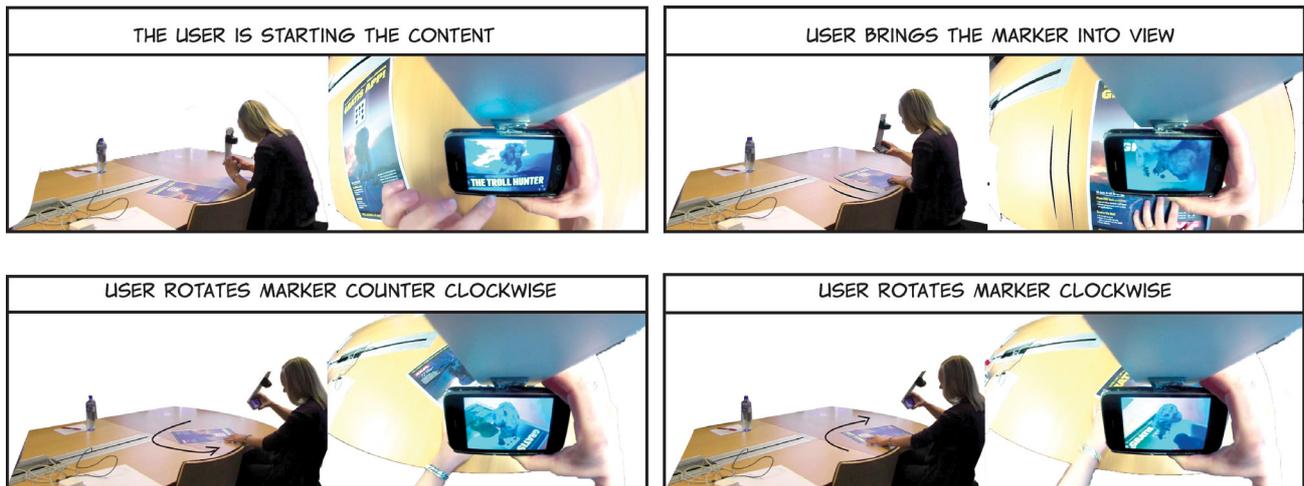


Figure 5. Observation of a user using the piece of paper with the marker on it to manipulate the 3D content.

B. Image content

We observed the participants recognize the image content as images. The concept of a slideshow we created was immediately recognized and articulated (Q5).

Q5 - R3: “*Sigh* now I’ve been through the entire slideshow.”

Q6 - R2: “I don’t know that it improves it, that the images are lying on top of the paper.”

The participant does understand the concept of the application, but he questions the value of AR-mode to display images, and clearly expresses boredom through body language and sighs (Q5). Q6 shows the user not seeing the point, expressing doubts about the experience this feature adds to the image content in any way.

Q7 - R6: “There is nothing wrong with experiencing it in this manner, but you are used to it being attached to the screen.”

This evaluator is decidedly neutral in discussing this way of viewing images (Q7). She walks through the motions of the task, but there is nothing suggesting fun or enjoyment when interacting with the image content. To view images in this manner does not add anything significant to experiencing of seeing image content.

We find that remediating flat still-images in this form serves little purpose as users express reservations about its merit. It serves its purpose of shoveling images to the users; they all understand the concept and diligently consume the images. We did not observe any immersive moments with the user using this content in contrast to the 3D content and video content where they conducted themselves differently.

C. Experiencing 3D content

When users interacted with augmented 3D content, we observed different reactions. We could also observe users perceiving the 3D content as significantly different than the 2D content. Technically, the image content is more or less the same, a mesh with a texture on it; particularly one user got excited about 3D content to a greater extent than 2D content (Q8, Q13, Q14, Q15).

Q8 - R7: “Wow! This is something completely different. This looks like game element perhaps? What is it?”

They spent a significantly more time exploring the 3D content than the flat image and video content. Users engaged with the model and looked at it from different angles, eager to discover features and details. This may

suggest that the exploration of the 3D model itself is a captivating and fun activity (Figure 3).

Q9 - T: “Is there an angle you prefer?”

R3: “This had depth, definitely.”

R3: “It’s more fun than standing over, it makes it flat. The fun part is that it is 3D.”

R3: “It’s tempting to touch it. I want to play with the windmill.”

The evaluator notes that watching the model from above gives the model less depth (Q9). This is a problem when developing AR content for print media. In this application, we achieve optimal tracking when the user has entire marker in view and a fair amount of high contrast pixels available for tracking. However, the 3D content itself achieves the most depth in conjunction with the print media at steeper angles. This is one of the main issues with freeform interaction with MAR content. It is difficult and time-consuming for a designer to develop content that is equally appealing from all angles and distances. Remediated content perhaps to a greater degree is created for other purposes, like effect shots in movies or informational content, rather than content to be viewed from above.

Q10 - R1: “It’s a 3D model of a troll.”

R1: “Stands still breathing.”

R1: “Waiting for something, Norwegian Christian’s mans blood.”

The users notice subtle movement (Q10) in the content and express more emotions and attention when interacting with the Raglefant troll.

Q11 - R1: “Being that it is an animal, it is tempting to look at him, in eye height.”

Q12 - R6: “I’m thinking I want to see him from a different perspective. I want to view him from the front.”

Some users want a deeper interaction with this content than the 2D content; they seem to require more interaction (Q11, Q12). We suspect the fact that it is a humanoid creature; the experience may be degraded by the fact that the content is not trying to interact with the users. It appears inanimate and behaves unnaturally.

Q13 - R5: “I’m trying to press it, since that is what the theme from before.”

R5: “YARRR!” (Participant mimics the noise of the troll)

R5: *Laughing*

Q14 - R7: “Oh oi! This is a Raglefant yes!”

The user exclaimed when the troll appeared in sight.

Q15 - R7: “Hello hello!”



1. Image slideshow on marker



2. Image slideshow in full screen

Figure 6. Image slideshow on marker (top), in full screen (bottom)

R7: "It got a bit annoyed when scratching it on the belly."

R7: "Nice!"

These users experienced fun, R5 (Q13) mimicking noises, almost forgetting the experimental setting and R7 (Q14, Q15) exclaiming and talking to the model. The 3D content, not surprisingly lends itself to greater immersion than the 2D content.

Despite a full screen icon being present throughout the interaction with the two content types, most of the users did not use it after first trying it out on the moving castle (Q16). One participant uttered this about going into full screen mode:

Q16 - R4: "Let's to full screen mode."

R4: "Now its in full screen, but that is not an improvement."

R4: "Its a lot cooler to not see it in full screen."

Those that used it found the interaction in this mode to be less compelling and to some degree pointless (Q17).

Q17 - A: "In contrast to what you mentioned before about that you want to see images filling the screen. How about this? Does this feel more natural?"

R6: "This is more like an object you should explore. You get the feeling that this object is sitting on the paper on the table, and that you are filming it in a way."

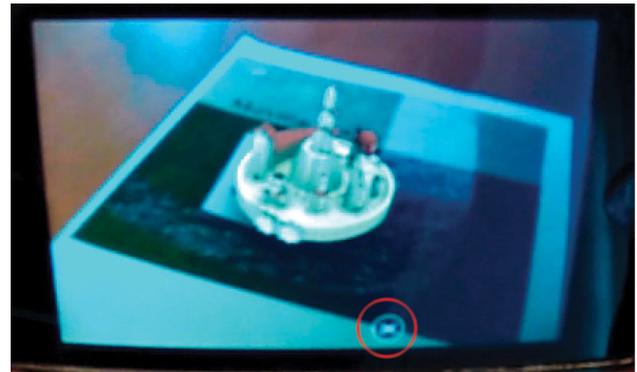
The users may feel this way because it is fun to "film" intriguing objects. If they think they are filming in 3D modus, this lends to the user experience while viewing it without its real world context lessens the immersion.

D. Physical interaction beyond the device with 3D content

How the users relate to the augmentations in physical space sheds light on how users perceive the content in combination with the markers.

1) Interacting with the troll

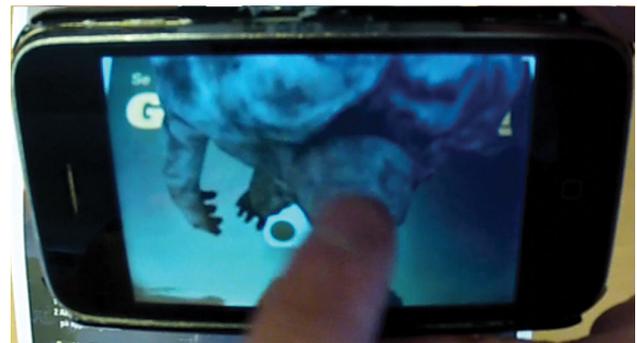
We observed different approach to interaction when the users engage with the troll than other content types. Two of the users tried to touch the augmentation in itself (Fig.



1. 3D castle on a paper marker



2. 3D castle in full screen mode



3. Raglefant on marker

Figure 7. 3D content in Arad

7), they swiped their fingers across the area where the augmentation appeared and expected some reaction. We could not observe this interaction with the 3D castle.

This leads us to believe that natural interaction for content on markers should also allow the users see some response from the interface when performing this type of interaction.

2) Exploring 2D content

During the evaluations of the flat content, all the users except one used their bodies and arms to align markers to the view of the mobile device. The users found the angle where the content was right side up and looked at it from that angle. None of the users wanted to see 2D content up side down (Fig. 4,6).

3) Exploring 3D content using markers

During the exploration of 3D content, only three participants actively engaged with the markers (Fig. 5).

- R7 rotates troll marker and castle marker.
- R5 rotates castle marker.

- R2 rotates the castle marker.

This observation may lead us to believe that many users may not engage with the marker at all, and their most natural way of interacting with 3D content will be by using the device when they are novice users of the technology. The affordance of markers and the content surrounding the markers should invite to interaction to enhance the user experience of 3D mobile AR content.

E. *Mixing paradigms*

Late in the development cycle the approach to experiencing "flat" content were changed. Instead of showing images and videos on the marker tracked in 3D first, the flat content – images and video will appear flat in full screen mode first.

This mimics the known paradigm of QR codes and came as a request by the media outlet. The intention was that users would find QR code interaction more familiar than the AR paradigm, where content appears tracked on the marker first. Users could touch a small icon, and 2D content would appear augmented on the marker. The evaluators had different views of this functionality. Through analyzing the videos we can observe how the participants chose to experience the video and image content.

The task presented to the users before experiencing the image and video data was as follows: They should point the phone at the marker and maneuver between full screen and marker mode (Fig. 6), tap through the slideshow and play/pause the video.

When we analyze one participant thinking aloud we observe that starting the movie in full screen mode is confusing (Q18, Q19).

Q18 - R4: "A film just appeared here. Did I do it correctly then do you think?"

Q19 - R4: "I have... I got it right there somehow. That was very cool."

A: "Do you know how? Can you reverse the process?"

R4: "I'm not sure. I got it right. Something happened afterwards. It did not happen before when I wanted to do it. Now it happened, now its fastened to the sheet of paper. And when I press play it plays"

A: "Ok"

R4: "That was pretty neat. If I press, then I get it in full screen there. Now I understand what I am doing. I didn't understand it previously. No, I have to find the marker there yes. And then I can click play, and take it to full screen from there. And it goes elegantly to full screen. It took some effort to understand."

The user does not understand how the film appeared in full screen. The connection between the marker and content is unnoticed by the user. After some trial and error, R4 manages to point the phone towards the marker when exiting full screen mode. Since he does not need to point the phone at the marker while in full screen mode, he has lost his point of reference

We asked the participants during the retrospective THA about the full screen-first approach "What do you think about the movie and image slideshows start in full screen mode". Some users responded negatively (Q20, Q21, Q22).

Q20 - R3: "It steals from the AR experience because the first thing that shows do not show the potential it has. If you didn't expect it you maybe hadn't... Why should I do anything with it?"

Q21 - R4: "I think it should have started down on the screen, not in full screen."

Q22 - R5: "I would have started it in world-modus because ... when a movie plays, or a slideshow is running you are experiencing the content already. I don't need to make more of itself, there and then."

Some participants were indifferent to it, but acknowledge its potential to confuse (Q23).

Q23 - R6: "It may be a little confusing in the start if you have to do something to make it be on the marker."

R6: "But I did not think about it really. It is perfectly fine."

However, two of the seven participants liked that the content started in full screen mode, and disliked the idea of having to point towards a marker (Q24, Q25).

Q24 - R1: "I thought it loaded very quickly, and if you are not needed to hold the camera towards the symbol, was very all right. So I liked that very much."

Q25 - T: "If you could choose, would you like it to start in full screen or on the marker?"

R7: "Full screen then."

The idea of having content pop up in full screen mode first seems not to be favored by the evaluators. However, the option to exit from AR-mode to full screen seems to be welcomed (Q26).

Q26 - R2: "I really like the functionality of being able to go in and out of full screen."

T: "What do you like about that?"

R2: "That you are independent of the marker when you have started it."

R2: "In addition to that the content is presented in a better way."

It may be tempting to align AR close to a medium like the familiar QR language where the user point a phone on a marker and content fills the screen. However, in this case we find that it creates inconsistencies for the user, and makes them confused.

Some experienced difficulties when pointing the camera towards a marker initially. Eventually, they made the cognitive leap from content to the marker. The idea of a marker did not become obvious until the content appeared on the marker itself. So having content appear in full screen right away may provide the user little chance of making the connection between the marker and content.

F. *Preferred content*

During the retrospective think aloud session users elaborated what content they found most enjoyable. The respondents answered that they favored the 3D content as well as the video content.

The users seemed to favor the 3D content. This is in line with the observations in regards to how they acted when experiencing the different types of content. The image content most users found uninteresting. However, we find it intriguing that the video content also was represented among the respondents as the favored content.

IX. IMPLICATIONS

By analyzing the findings we can summarize some design implications for future attempts to remediate content in this manner.

There is a need to develop a metaphor for interacting with content beyond the touch screen, people intuitively seeks to touch an augmentation. Hurst et al [25] also points out the need for gesture interaction in AR via finger tracking. The data reveals that interaction does not need to be complex. Simple tracking of a swipe gesture in front of the camera may be enough to support greater immersion with AR content.

We can see that the users attempt to engage more directly with humanoid and animal creatures and find it awkward that they appear lifelike, but do not interact naturally with the user. This is similar to the findings of Wagner et al [26] where they found that a virtual character not relating to them made them uncomfortable and could even be perceived as offending.

When creating 3D content to be displayed on markers in print media, we find that users will try to optimize their viewing angle to the content. However, when dealing with content directly remediated from another medium care should be taken – if possible – to reorient the content. This is to ensure that it appears nice when viewed from a downward angle.

When users interact with 2D content, we find that they prefer to start viewing it on a marker. In some cases, they bring it to full screen mode to take a closer look. However, when the user is trying to align the content to the display area, some automation should take place to optimize its viewing properties so the user can experience it in its native format on screen.

As it stands now, markers have not acquired the affordance for rotating. People recognize markers, but they do not use them to their full potential yet. Adding instructions to the markers themselves to afford rotation and manipulation may improve their affordance and encourage users to interact with content beyond maneuvering the mobile device itself.

We find that image content suffers most from being remediated in this manner; no users preferred this content over video or 3D content. Most of the users preferred the 3D content, and some users genuinely enjoyed watching and interacting with the video.

It has become increasingly clear that the user experience of remediated AR content is closely related to the affordances of the content itself. Text, images and video have clear and (perhaps culturally) deeply set affordances, 3D content on the other hand does afford a true new media experience. This may be self-evident to any practitioner in AR or MAR, to anyone else it may not.

X. AFFORDANCES

Based on analysis of the findings we have identified affordances related to Kaptelinin's instrumental affordances in addition to the learning, maintenance, aggregation, and our own niche affordance of remediation.

Firstly it is beneficial to clarify the basic instrumental handling and effector affordances of our application. This gives an impression of the possibilities for interacting with this technology.

The application supports handling through tangible markers, GUI widgets on the display, and the by maneuvering the mobile phone itself. These handling affordances let us effect the augmentation through interaction. We may handle the device (move the device), the GUI (use buttons to move to the next image) and the markers (in the world) to handle the augmentation to affect the augmentation layer as we observe the users do eventually. The effector affordances come through the device, GUI and marker. The users affect the augmentation primarily through maneuvering the device. The GUI allows users to play and pause videos, move between images and view animations inherent in the 3D

content. We find that users rarely handle the markers to affect the augmentations.

Ideally the handling and effector affordances should be easily and intuitively picked up, in some cases this how it transpires. Most users intuitively point the phone at the marker and starts interacting, while some users require directions to start the interaction itself. Hence we recognize the need to identify the "web of mediating" affordances that make MAR viable.

A. Maintenance affordance:

Maintenance affordance is related to tasks one must perform to operate the IT artifact. As we use CV in the tracking algorithm, both the markers and the environment the markers operates in requires maintenance. This comes in the form of maintaining good enough lighting and maintaining the integrity of the markers. This means they cannot be crumbled too much, and the lighting must be sufficient to support tracking. This is an affordance that is not particularly easy to communicate. Room lighting is not something one immediately thinks of as a problem when interacting with MAR. Human eyes adapt easily to changing lighting conditions, in sharp contrast to CV tracking algorithms.

B. Aggregation affordances:

Markers have yet to become a convention for point of interaction. Even though the participants interacted with printed surfaces with markers on them, not everyone understood that markers reveal AR content. Markers in themselves poorly afford the desired interaction, and the participants seldom read on-screen instructions. This is an obstacle for AR content in print media at this stage. Newspapers, and sheets of paper do not afford engaging manipulation.

C. Learning affordances:

Even with onscreen instructions telling the users to point the camera at the marker, this simple instruction would not be absorbed. This is similar to a finding in a study of AR games [18]. The study makes a point of users having trouble understanding what object in the world the AR application is referring. It must be made exceedingly clear what real world objects the application interacts upon.

When the content appeared on the marker, affordances related to translating and rotating the objects were immediately recognizable using the phone to rotate the view. Very few of the participants wanted to rotate the printed media to experience the content from different angles.

Seldom we find that other AR applications try to shed light on how they work. According to Bødker[14], some applications need to mediate how they work, in contrast to invisible computers. AR relies heavily on computer vision (CV), for a user to be able to correct errors in CV it is useful to understand the limitations of the CV algorithm. To enjoy an AR application we may need a representation of its internal status more clearly. We do provide users with instructions on where to point the marker initially, but we decided against giving the users information about how well they are tracking the markers. This was a conscious decision. We felt that users might focus on achieve optimal tracking, rather than consuming content if

we provided them with visualization of the current performance of the tracking.

D. Remediation affordances:

The findings emphasize the need for users to bring the content into a known interaction paradigm. Users often tried to align 2D content to the display insofar that it resembled a familiar way of consuming this type of content. Users liked the idea of being independent of the marker, by enabling them to enter full screen mode to move the device freely while still seeing the content. In the trials we had the content appear in full screen directly, some users found this to be unintuitive as they expected the content to be first in 3D.

E. Design guidelines for content remediated through handheld AR:

It is necessary to note that these guidelines are intended for handheld augmented reality. They may not have much merit in other AR applications. However, we word them to be usable in general AR applications.

1) Learning affordances

- Clearly afford what objects the application is referring to in the physical world to support learning.

This can be achieved by use visual clues that make it clear how the software understands the world.

- Represent the internal state insofar that users may learn how the CV algorithm performs satisfactory.

Signal the inner workings of the tracking algorithm.

2) Maintenance affordances

- Afford actions needed to adequately maintain an environment suitable for augmenting.

We believe this may be achieved by introducing functions to the hardware: external sensors for sensing light and providing information about the environment to enable users to take steps to improve the performance of the application.

3) Aggregation affordances

- Afford the relationship the device has to the trackable.

Visual clues can be used to make the users direct the device at the marker. Bear in mind that textual clues may not be perceived.

4) Remediation affordances

- Afford easy transition between the known interaction and the augmented interaction of remediated content.

Users will try to view 2D-content as they are used to in other media, we support this by allowing users to enter a mode where they are not dependent on the marker.

- 3D content should afford their value from natural angles.

As 2D content can easily be viewed from a bird's eye view, this is not the case for 3D content. 3D content needs to be designed explicitly to support viewing from a bird's eye view.

- Afford the same entry point for interaction even though the content is different

We experienced that our effort to align the interaction closer to a QR paradigm would confuse the users rather than improve the user experience.

- Humanoid character that afford interaction should allow interaction with the user

Users may become uncomfortable or provoked if humanoid characters that afford interaction are unresponsive.

XI. CONCLUSIONS

This study describes an application using AR to augment print media with remediated content. We find that remediation of video content in an AR context seems promising. Users find it at times awkward to view flat (2D) content in a marker-context and like the functionality allowing them to bring flat content into a full screen mode. Image slideshows in the AR context has little merit amongst the users and contributes little to an overall positive user experience of the application. 3D content provides users with the greatest level of immersion and users engaged more directly with this type of content.

A consequence of MAR being a new medium it may be tempting to leverage existing maladapted content to help saturate this new platform with content. Case in point being the idea of using a QR-code metaphor, and 2D images because it is familiar and easy. We believe this degrades the user experience, firstly because it makes the MAR medium to some degree pointless and it adds little to the user experience.

Affordances related to aggregation, learning, maintenance and remediation is described in this paper. These affordances and can be used in combination with design guidelines to assist design of remediating MAR applications. These guidelines can be utilized to create a better user experience of content for the MAR platform.

Overall we conclude that the ARad application gives a fun user experience, and we argue that the true potential in remediating content to MAR lies in the user experience of 3D content.

XII. FUTURE RESEARCH

In this study we have looked at remediated content from a newspaper to the MAR domain. Future research would be into the affordances of MAR games. The maintenance and learning affordances we expect to remain similar. However, in regards to aggregation and the remediation of game content and interaction we expect some differences. Game developers adapting games for the MAR domain may seek to remediate concepts and content from game development to the MAR application. It would be interesting to investigate if this is the case, and if so what shape and form the content and conventions take, and to identify design guidelines for game interfaces in MAR applications.

REFERENCES

- [1] D. Wagner and D. Schmalstieg, "First steps towards handheld augmented reality," presented at the ISWC '03 Proceedings of the 7th IEEE International Symposium on Wearable Computers, 2003, p. 127.
- [2] R. Azuma, "A survey of augmented reality," *Presence-Teleoperators and Virtual Environments*, vol. 6, pp. 355-385, 1997.
- [3] J. D. Bolter and R. Grusin, *Remediation - Understanding New Media*, Paperback edition. The MIT Press, 2000, pp. 1-290.
- [4] B. Macintyre, J. D. Bolter, E. Moreno, and B. Hannigan, "Augmented reality as a new media experience," presented at the

PAPER
AFFORDANCES IN MOBILE AUGMENTED REALITY APPLICATIONS

- Augmented Reality, 2001. Proceedings. IEEE and ACM International Symposium on, 2001, pp. 197–206.
- [5] J. D. Bolter, M. Engberg, and B. Macintyre, “Media Studies, Mobile Augmented Reality, and Interaction Design,” *interactions*, ACM, Jan-2013.
- [6] M. Engberg, “Writing on the world: augmented reading environments,” *Sprache und Literatur*, pp. 67–78, Feb. 2012.
- [7] L. J. Rosenblum, S. K. Feiner, S. J. Julier, and J. E. Swan, “The Development of Mobile Augmented Reality,” in *Expanding the Frontiers of Visual Analytics and Visualization*, no. 24, London: Springer London, 2012, pp. 431–448. http://dx.doi.org/10.1007/978-1-4471-2804-5_24
- [8] J. V. Pavlik and F. Bridges, “The Emergence of Augmented Reality (AR) as a Storytelling Medium in Journalism,” *Journalism & Communication Monographs*, 2013.
- [9] D. Schmalstieg, T. Langlotz, and M. Billinghurst, “Augmented Reality 2.0,” in *VIRTUAL REALITIES*, no. 2, Wien: Virtual Realities, 2011.
- [10] J. Swan and J. Gabbard, “Survey of user-based experimentation in augmented reality,” presented at the Proceedings of 1st International Conference on Virtual Reality, 2005.
- [11] F. Zhou, H. B.-L. Duh, and M. Billinghurst, “Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR,” presented at the ISMAR '08: Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, 2008, pp. 193–202.
- [12] M. De Sà and E. F. Churchill, “Mobile Augmented Reality: A Design Perspective,” in *Human Factors in Augmented Reality Environments*, no. 6, W. Huang, L. Alem, and M. A. Livingston, Eds. New York: Springer, 2013, pp. 139–164.
- [13] V. Kaptelinin and B. Nardi, “Affordances in HCI: toward a mediated action perspective,” presented at the CHI '12 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2012, pp. 967–976.
- [14] S. Bødker and P. B. Andersen, “Complex mediation,” *Human-Computer Interaction*, vol. 20, no. 4, Dec. 2005. http://dx.doi.org/10.1207/s15327051hci2004_1
- [15] M. McLuhan, *Understanding Media - The extensions of man*. McGraw-Hill, 1964.
- [16] G. A. Lee, U. Yang, Y. Kim, D. Jo, K.-H. Kim, J. H. Kim, and J. S. Choi, “Freeze-Set-Go interaction method for handheld mobile augmented reality environments,” presented at the VRST '09: Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology, 2009. <http://dx.doi.org/10.1145/1643928.1643961>
- [17] J. Iivari, “A paradigmatic analysis of information systems as a design science,” *Scandinavian Journal of Information Systems*, vol. 19, no. 2, 2007.
- [18] T. Gjørseter and K. Jørgensen, “Combining Think Aloud and Comic Strip Illustration in the Study of Augmented Reality Games,” presented at the NOKOBIT 2012, 2012, pp. 1–21.
- [19] K. A. Ericsson and H. A. Simon, “Verbal reports as data.,” *Psychological review*, vol. 87, no. 3, pp. 215–251, 1980. <http://dx.doi.org/10.1037/0033-295X.87.3.215>
- [20] A. Dix, J. E. Finlay, G. D. Abowd, and R. Beale, *Human-Computer Interaction (3rd Edition)*, 3rd ed. Prentice Hall, 2003.
- [21] A. Dünser and E. Hornecker, “An Observational Study of Children Interacting with an Augmented Story Book,” presented at the Edutainment'07 Proceedings of the 2nd international conference on Technologies for e-learning and digital entertainment, Berlin, 2007, pp. 305–315.
- [22] F. Liarokapis, L. Macan, G. Malone, G. Rebolledo-Mendez, and S. de Freitas, “A Pervasive Augmented Reality Serious Game,” presented at the Games and Virtual Worlds for Serious Applications, 2009. VS-GAMES '09. Conference in, 2009, pp. 148–155.
- [23] “Video in Qualitative Research,” 2010.
- [24] A. R. Hevner, S. T. March, J. Park, and S. Ram, “Design science in information systems research,” *Mis Quarterly*, vol. 28, no. 1, pp. 75–105, 2004.
- [25] W. Hürst and C. Wezel, “Gesture-based interaction via finger tracking for mobile augmented reality,” *Multimedia Tools and Applications*, pp. 1–26, Jan. 2012.
- [26] D. Wagner, M. Billinghurst, and D. Schmalstieg, “How real should virtual characters be?,” presented at the ACE '06 Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology, 2006.

AUTHOR

Tor Gjørseter is with University of Bergen, Norway.

Submitted 17 July 2014. Published as resubmitted by the authors 14 October 2014.