An experimental study on the role of AR content type
in an outdoor site exploration

Abstract
Overlaying a building with images from the past can be an engaging way to explore a historic site. However, little is known about what type of content functions well when using augmented reality (AR) in outdoor contexts. This research investigates how different types of AR content – such as text or image - can affect the flow experience as well as other cognitive, affective and behavioral responses. We ran an experimental study with 85 participants at a university campus, where three groups used different apps – AR app with overlaid textual information, AR app with overlaid both textual information and images and an app with audio guide - to explore the area’s historic and cultural background. The results show that overlaying images in addition to the text was clearly the most successful way of attracting attention and providing stronger flow in comparison to the apps that overlaid only text or delivered audio. However, we also discovered that users occasionally interpreted the overlaid imagery to be cues for something else other than what they were designed to represent. Finally, we discuss how AR content can enlarge otherwise invisible details, depict historic elements and unveil interiors of buildings.

Keywords: augmented reality, flow, in situ study, outdoor exploration, knowledge retention
1 Introduction

Popularity of AR has grown in multiple areas, amongst which are: commerce, tourism, entertainment, education and art. Application designers are continuously inventing new content types that can in some way enrich the physical environment. SnapChat updates AR filters on a daily or weekly basis (Telegraph, 2017), Pokémon Go attracted millions of users catching virtual Pokémon in real environments (Rauschnabel, Rossmann, and tom Dieck 2017) and commercial applications for virtual try-ons are close to becoming a standard for some product categories like make-up\(^1\) or glasses.

Because AR overlays the physical environment in real time and is deployed on mobile smart devices, it holds a strong potential for on-the-go discovery of historically and culturally relevant sites in a compelling manner\(^2\) – for example pointing one’s phone towards a cathedral to see how it used to look like centuries ago or to read particular information overlaid on a city skyline. Studies have shown that such applications can enable a deeper appreciation of artifacts in a museum context (Chang et al. 2014). However, because the AR content gets overlaid onto a view of the real world, the amount of information and modality of content requires special consideration. The two realities exist differently. Physical reality is tangible and is visible without a smart device, while the augmented content exists solely in digital form on the screen. But in AR they are visualized together. The question is what type of augmented content thus functions well as part of such view? Is textual information on its own the most appropriate way to supplement a particular historic site with additional facts? Would adding images help or confuse the user? Pictorial material can offer visualization of what would be hard to explain with words and is in general consid-

\(^1\) An example of make-up AR try-on is Rimmel Getthelook https://itunes.apple.com/gb/app/get-the-look-rimmel-london/id1120244830?

\(^2\) One of the first such apps implemented in a city context was UAR, developed in 2010 Rotterdam https://www.layar.com/news/blog/2010/06/30/uar-urban-augmented-reality-experience-the-past-present-and-future/
ered to have a stronger appeal than only textual content. However, an overlay with excessive information or a confusing combination of content types can potentially appear unclear or overwhelming in an AR view where the two realities visually blend.

Much research has explored the effectiveness of using different content modalities such as image, text, video or audio (Peeck 1993) for learning. Also, AR has been demonstrated as an effective tool for various educational purposes (Kurilovas 2016). However, little is known how content representations affect consumer appreciation of AR experience and absorption into outdoor exploration. Given that AR fuses digital content with real world, the content is viewed and perceived in a distinctive manner as information – be it imagery or written text - is much more embedded into physical surrounding and thus corresponds to specific points in environment. When one is learning new pieces of information about a particular site through an AR app, is there a certain type of content which functions best as an overlay on the real world? Can two different types of content – images and text – interrupt an immersion in the exploration if presented simultaneously or do they, on the contrary, lead to more favorable responses on the user’s side when so combined?

Mobile applications and audio information guides designed for exploration of cultural heritage, such as in museums and galleries, are now common (Bujari et al. 2016; Chiu, Tseng, and Hsu 2017; Julier et al. 2016; Smirnov, Kashevnik, and Ponomarev 2016). AR is becoming an increasingly used feature in such apps, especially when designed for indoors use (Brancati et al. 2016; Capuano et al. 2016; tom Dieck and Jung 2017). However, historic and cultural sites are often located in outdoor contexts, which is an environment with many interfering elements. If AR apps are to be designed and used also outdoors, we need to know more about how AR works within the specific features and context of an outdoor setting. For example, what are the effects of external factors on a user’s attention, such as
lighting, ambient sound and other people? How well are users able to combine the digital content (be it textual, pictorial or both) with what they are looking at in the real world?

We report on a study here that investigates how two different types of content - (i) image and (ii) image and text - affect the immersion into flow when using an AR app developed for outdoor exploration. In particular, we compare the differences between two types of AR content. By way of control we also used a traditional audio guide. The obtained insights can help AR designers in crafting enriching experiences for those who take to smart devices during their outdoor discoveries.

2 Background

Earlier studies on AR in tourism and learning focused more on the technical challenges (Kounavis, Kasimati, and Zamani 2012; Ahn, Ko, and Yo 2014), such as how to set up an efficient system and what are the requirements needed for an application to run smoothly - for example how to achieve compatibility of applications developed by SDKs across different operating systems. Designing for AR apps required resolving issues such as what is the best way for users to search for the content and what functions are most appropriate for user interface (Yovcheva, Buhalis, and Gatzidis 2012). While technical challenges still remain salient, it is the different facets of user experience with AR that have attracted more attention of late due to the increased use of this technology (Yovcheva et al. 2014; Kourouthananassis et al. 2015; Javornik et al. 2016).

2.1 AR features and content

Both content and features of AR apps depend on the value that the apps aim to deliver. For tourist AR apps, for example, Yovcheva et al. (2014) underlined two key design issues. Firstly, they emphasized that AR content should display the salient characteristics of the
physical target - features such as a building’s name should be in some way represented in the AR view. Secondly, they point out that AR apps should satisfy the information needs of the users. For instance, if there is an important landmark a user would be interested in, she expects the app to provide such location-based information. Despite the relative novelty of AR, users hold certain expectations about AR that should be fulfilled if an application is to deliver a favorable experience (Olsson et al. 2013; Yovcheva et al. 2014).

In an unfamiliar outdoor location, navigational cues are for instance a highly valued feature for that reason. In their study on AR app in tourist sites, Kourouthanassis et al. (2015) reported that users thought their touristic experience was improved because of having the camera-view on a smart device overlaid with useful navigational cues, delivering a functional purpose. Usability and reliable performance were rated as crucial for the quality of that particular experience, while personalization of the content (based on the user profile content preferences) did not yield significantly better evaluations of the app in comparison to non-personalised content (Kourouthanassis et al. 2015).

While user satisfaction can be strongly determined by such functional benefits, the hedonic, experiential value can also play a prominent role depending on the context. Jung, Chung and Leue (2015) investigated an uptake of AR in a Korean theme park, where visitors were invited to play in a showroom with a marker-based AR application overlaying interactive virtual characters onto the surroundings. The study unveiled that, besides the system’s quality related to the mechanics of AR, such as tracking accuracy, it was users’ appreciation of the content quality that had a significant impact on their satisfaction and intentions to recommend this application to others. Similarly, when Chung, Hang and Joun (2015) investigated users’ attitudes towards using AR during their visit to a historic palace, the users underlined the importance of AR annotations’ visual appeal in the sense that they cared strongly about the aesthetics and the design of the overlaid content.
Studying the impact of AR functionalities on overall AR experience remains relevant also for other AR devices, such as head-mounted displays. tom Dieck, Jung and tom Dieck (2016) showed in their study in an art gallery that an AR guide on Google Glass led not only to enjoyment, inspiration and creativity, but it was also nudging the users into exploring new routes for a museum visit.

Progressively more efforts have thus been directed towards designing appropriate AR features. Investigations into the differences among AR content types, however, have been less prominent so far, especially in relation to different contexts.

2.2 **Content type**

Little is known of how different types of AR content contribute to the quality of outdoor exploration. According to media studies, content modality impacts a variety of user responses to technologies (Sundar 2000). The two most common modalities for displaying information are visual and textual (Chau, Au, and Tam 2000). Visual information relates to images, which can be displayed in multiple ways with different sizes, styles and perspectives and may feature movement or animation. In contrast, textual information relies on written or spoken words to convey information about the object of interest. Combination of visual and textual information is commonly adopted when designing the information environment (Blanco, Sarasa, and Sanclemente 2010; Hong, Thong, and Tam 2004; Kim and Lennon 2008). Earlier research has demonstrated different effects for imagery versus discursive information on content perceptions, primarily with images having demonstrated more pronounced effects on a variety of user responses (Friedman 1979; Kim and Lennon 2008; Liu and Stout 1987; MacInnis and Price 1987).

While imagery has been generally shown to facilitate knowledge retention, it can in some cases also be misleading. Pictures can for instance obscure factual information (Crisp
and decrease a reader’s focus (Peeck 1993). A combination of static (such as text, image) and dynamic representations (where content can be modified with various interactive features) can in some cases impede learning in computer environment (Bodemer et al. 2004). Also, people often view pictures in a more brief way, as they expect to understand the main content only by glancing (Peeck 1993). More exact cues about what to observe in an image can thus assist more focused attention.

Images have been shown to have strong impact also on affective responses across different contexts. Liu and Stout (1987) have for instance indicated that either pictures and words or pictures together are more effective than words alone for eliciting positive attitudes in commercial contexts. Seeing portrait pictures of work colleagues during online conversation when getting acquainted with the working team elicits stronger positive attitude towards the team members than if the images aren’t there (Walther, Slovacek, and Tidwell 2001). Similarly, Pittman and Reich (2016) indicated in their recent study that image-based media such as Instagram led to positive affective responses such as happiness and that it lowered feelings of loneliness thanks to being an image-based media in comparison to Twitter that is dominated by text.

Moreover, Hong, Thong and Tam (2004) demonstrated the impact of visual information also on behavior intentions, more exactly in intentions to purchase when using e-commerce sites. The study by Lin, Lu and Wu (2012) demonstrated a similar impact of images on purchase intentions in the context of electronic word-of-mouth. The specificity of a task can, however, diminish an impact of visual material. Job seekers for example seem to respond more to the employment websites based on the amount and quality of textual information and are less reactive to online graphics and imagery (Allen et al. 2013).

From technological perspective, however, the behavior that is of the main interest is the repeated use of the application in the future. More research is required if such content
modalities lead to differences in that type of behavior, especially with immersive technologies.

There is growing knowledge on AR in education where researchers have for instance already shown that AR can foster collaborative learning (Dunleavy, Dede, and Mitchell 2009), conceptual understanding, spatial ability, practical skills (Cheng and Tsai 2013) and even foreign language learning (Perry 2015). In their experimental study, Chang et al. (2014) compared the effects of using an audio guide versus an AR application for students who were learning about artwork in a museum environment. Those who used the AR app showed superior learning outcomes related to analyzing artistic value of the displayed work. Given this established positive influence that AR can have on learning, there is a definite need for more knowledge about select and deployment of specific contents for AR. Could a variation in the type of augmented visualization enhance an exploration of an outdoors cultural site? Would some type of content also hinder such activity? Our first research question is thus:

(i) Do different AR content modalities impact consumer affective responses (attitude), cognitive responses (learning) and behavioral intentions (willingness to use the AR app again) when used outdoors?

2.3 Flow in AR

Another key concern for AR experience is to understand how involved and immersed users become while using the app. Flow has been considered a central measure of a person’s absorption in an activity (Csikszentmihalyi 1997). It refers to a combination of perceived sense of control, absorption, intrinsic interest and curiosity (Webster, Trevino, and Ray 1993) and indicates that a person is highly involved with the activity and enjoying it as well as paying lots of attention to it. If flow is achieved, it can impact user’s perceived
experiences, cognitive responses and behavior (van Noort, Voorveld, and van Reijmersdal 2012). Numerous research demonstrated that digital environment can greatly contribute to user’s immersion into flow (Mathwick and Rigdon 2004; Weibel et al. 2008; Hoffman and Novak 2009).

A study where online buying behavior was examined in a lab environment has demonstrated that AR had a significant effect on flow (Javornik 2016). In a cultural context, Chang et al. (2014) compared AR to a standard audio guide and found the use of AR resulted in a significant increase how immersed the users were in flow when appreciating the art in an indoor museum. Such studies are indicating that AR environment is successful in eliciting the state of flow in the user. Given that a combination of images and text often functions as the most appealing content type, we would expect for such combination to contribute to the state of flow. But would such state be realized in an outdoor context where interference from the environment might impact the flow? As people walk about outdoors, they encounter numerous surrounding elements, such as passers-by, buildings and street furniture. When needing to navigate between obstacles whilst walking, they feel more rushed and are less absorbed in searching for information (Harvey and Pointon 2017). Using AR outdoors would thus differ from AR experience in other contexts such as art gallery or online buying. Does outdoor environment still allow for a flow experience when using an AR app? Would using a combination of AR image and text enhance or detract from immersion when in such particular environment? This leads to our second research question:

(ii) Do explorations of outdoor historical sites vary in their level of flow depending on the type of AR content that is being used?
3 Methodology

We conducted an in-the-wild study in order to observe how participants take to an AR app designed for a mobile tablet when exploring an outdoor site. The study took place at a university campus in central London. We developed two AR apps for the study – one where AR content consisted of text information and one where the AR content included images in addition to the text. A control condition of audio only was compared to these two conditions. Details on the apps are given further below.

This methodology seemed particularly appropriate for gaining insights with high external validity. We were able to observe how these different contents function in an outdoor environment that was relatively busy and had external factors that are difficult to control such as noise and people passing by.

We used a questionnaire that asked about: (i) flow in the experience, (ii) their perception of how the site was augmented through the content, (iii) user’s attitude to the application and (iv) the intention to engage with the app again. The corresponding measurement scales together with the measures for Cronbach’s Alpha are displayed in Table 1.

A few notes on the measurements. Flow is a multidimensional construct (Hsu and Lu 2004), comprised of concentration, intrinsic interest, curiosity and control (Webster, Trevino, and Ryan 1993). We were interested in understanding the level of overall flow, but also in these specific dimensions – in particular the attention that the users paid to the activity of exploring and how much intrinsic interest and curiosity the different content types evoked in them.

Further on, perceived augmentation is a measurement that was developed for evaluating how much users perceive an AR app to virtually augment a physical entity (Javornik 2016). For that reason, we measured the perception of augmentation only for the two AR apps and not for the audio app, as the scale is not fit for that purpose.
Additionally, a multiple choice test was used to test the knowledge retention and to evaluate to which extent they remembered the information seen in the app or heard by the audio guide. Also, the researcher observed how the participants were using the app and how they were exploring the area. She asked the participants about their impressions about the app and the overall experience.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Measurement items (7-Likert scale)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow</td>
<td>I felt in control. (Control)</td>
</tr>
<tr>
<td></td>
<td>I was aware of distractions. (Attention focus)</td>
</tr>
<tr>
<td></td>
<td>I thought about other things. (Attention focus)</td>
</tr>
<tr>
<td></td>
<td>I was totally absorbed in my activity. (Attention focus)</td>
</tr>
<tr>
<td></td>
<td>I was focused. (Attention focus)</td>
</tr>
<tr>
<td></td>
<td>I felt curious. (Curiosity)</td>
</tr>
<tr>
<td></td>
<td>It was very interesting. (Intrinsic interest)</td>
</tr>
<tr>
<td></td>
<td>It was fun to explore the application. (Intrinsic interest)</td>
</tr>
<tr>
<td>Perceived augmentation</td>
<td>The application added virtual elements directly to the building.</td>
</tr>
<tr>
<td></td>
<td>The virtual elements existed as a part of the building when I used the application.</td>
</tr>
<tr>
<td></td>
<td>The virtual elements seemed to exist in real-time.</td>
</tr>
<tr>
<td>Behavioral intentions</td>
<td>I would like to use this application in the future.</td>
</tr>
<tr>
<td></td>
<td>I have the intention to talk about this application with friends and colleagues.</td>
</tr>
<tr>
<td>Application attitude</td>
<td>The application was…</td>
</tr>
<tr>
<td></td>
<td>Not entertaining - Entertaining</td>
</tr>
<tr>
<td></td>
<td>Not enjoyable – Enjoyable</td>
</tr>
<tr>
<td></td>
<td>Not boring – Boring</td>
</tr>
<tr>
<td></td>
<td>Not useful – Useful</td>
</tr>
<tr>
<td></td>
<td>Not interesting – Interesting</td>
</tr>
</tbody>
</table>

Table 1: Questionnaire used to measure flow, sense of augmentation, behavioral intentions and attitude to the application

4 Content design

The appearance of the digital content could be activated at selected sites by pointing the tablet in camera mode at the defined markers. The augmented content would then overlay the screen. Both versions were built using Unity3D³ and the Vuforia library⁴. This library provided tracking for each target, specifically whether that target was visible and where it

³ https://unity3d.com
⁴ https://developer.vuforia.com/downloads/sdk
was in relation to the image from the device camera. Using this information, the application calculated an approximation of the position and orientation of the device in relation to the target, which was, in-turn, used to inform a Unity3D ‘camera’. The camera was subsequently used to render viewpoints of a ‘scene’, which held information about the markers and the digital content that needed to appear overlaid on them (such as text and images). Digital content that was ‘attached’ to specific targets would thus appeared when the target was visible to the device camera and in positions and orientations that mirrored those of the virtual scene.

Both versions had the same textual information for each site, but they differed in that one version also had images appearing on the screen simultaneously with the text (Fig.1 and 2). Importantly, this permitted a systematic comparison between the two modes of content visualization (text vs. text + image).

![Figure 1: The two types of content that appeared on the screen in front of the Life Sciences building at the campus: text-only version (left) and text+image (right). The overlaid image is an enlarged visualisation of the entrance door handle.](image)

The textual information and images were selected from archives with historical and cultural content closely related to the sites. The process of designing the digital content in-
volved several steps: a) identifying the appropriate locations; b) selecting the relevant content for the particular location; c) choosing the amount of content to fit onto the screen. We chose the locations that would not detriment user’s safety (for instance being too close to a road), would allow users to see the augmented content without being interrupted by numerous passers-by and where the changes in outdoor lighting conditions would not greatly impact the marker recognition by too large an extent.

We selected three sites: a) a campus building named after Charles Darwin, b) the entrance to one of the university departments’ buildings, and c) the campus’s Art Museum. They were chosen based on their historical or cultural relevance (Fig. 3) and also on the diversity of the related content. The sites carried linkages to famous personas or events and were also significant due to the architectural developments and historic purposes they served. A team of three researchers chose the information that they perceived relevant for the locations, by choosing three pieces of information per each scene.

![Figure 2: A print screen of what a participant saw at one of the sites with the text+image (left) and the text-only (right) version.](image)

The aim was to choose content that was informative and site specific (i.e. not too general) and that it added an informative layer to the narrative of a selected location as a whole,
rather than fragmented details. Also, the content referred to the site’s most known characteristics, which is an important feature of AR outdoor systems (Yovcheva et al. 2014).

The text was fitted on the screen in the manner that it did not cover more than 20% of the view. For the versions that had both text and image, the image was overlaid in a way that all the content did not cover more than half of the screen when triggered from the defined markers. The researcher indicated to the participants the points where they should stand in order to trigger the AR content. The participants needed to point the tablet towards the selected building or site as part of the pre-defined route. The content would get bigger if the participant approached the markers or moved the tablet closer to the site. Interacting with the content was thus possible by pointing the device towards a site in order for the content to appear, flipping through different scenes when at the same site and then changing the size of the digital content by moving closer to or further from it. The content was directly linked to the markers and if the tablet was moved too left or right, the participant would lose the content of the screen.

Figure 3: Another selection of content that appeared on the screen in front of a department’s building at the campus
5 User Study

The study took place at a large university campus. The participants used an AR app to learn more about relevant historic and cultural content of the pre-defined sites or buildings. Each participant took part in only one condition to prevent transfer effects. 85 participants took part in the study, divided across the 3 conditions. 47 were female, 36 were male and for two we had missing data from the questionnaire. All were between 21 and 34 years old, except three participants who were older than 40. They were given a small monetary incentive to take part in the study.

The first group of participants (N=29) used a tablet with an app that overlaid only textual information on the sites when they pointed the tablet towards the pre-defined points. The second group of participants (N=28) used a tablet with an AR app that displayed the same textual information as the app in the first group. In addition, the app also displayed images that accompanied the textual information and thus provided visual complement to the written content. The third group of participants (N=28) used an audio guide (also a tablet). The audio contained the same information as the displayed textual information on the screen with both AR apps. The audio automatically set off on the same points where the AR apps delivered the content. Figure 4 shows a participant from each category exploring the campus with the assigned application.

5.1 Procedure

The participants were guided through the campus by a researcher, who collected their comments and made notes about their use of the app. The tour around the campus lasted between 15-20 minutes during which the researcher indicated the three sites that the participants could augment with the app or where they could listen to the audio guide (Fig.4). At the end, the participants were asked to fill out the questionnaire and multiple choice test.
While it is more common for an app to guide a user to the main sites for which the content has been created, we did not include such navigation features for two reasons. Firstly, as this study focused on the effects of the different types of content on the user experience, we tried to minimize the inclusion of other features that could interfere with this app feature. Secondly, the conditions needed to be as consistent as possible across all the three groups and adding navigation features to the AR apps create a substantially different experience in comparison to the audio guide group.

5.2 Findings

Overall, the participants in all the groups evaluated the different aspects of the experience positively (see Means and Standard deviations in Table 2). Fig.5 demonstrates a trend for all the variables to be evaluated the highest is the group that used the AR app with both text and images, while the group with audio guide gave the lowest scores. However, the mean values were grouped quite closely (between 4 and 6).

5.2.1 Immersion in the experience

Using both AR apps for outdoor exploration was found to be more immersive than listening to just an audio guide. An ANOVA analysis across the groups showed the reported flow by the participants in the audio guide group was significantly lower from the
groups that used the AR app with text only and those from the group that used the AR app with text and images. Higher immersion of the participants in flow suggests that the AR apps captured more of their attention, and made the exploration more interesting. Moreover, the analysis showed also significant difference between the two types of AR apps - the participants in the groups with AR text and image reported having experienced a higher flow after the tour than those that used AR app with text only. Having images as part of the AR content thus makes a difference in how absorbed the participants were in the exploration. There was not any significant difference between the two AR groups in terms of perceived augmentation. This signaled that both AR apps augmented the sites in a realistic manner and embedded the virtual content as part of the physical spaces. Differences in flow between the two types are thus not due to a different quality of augmentation, but rather to the differences in the type of content.

![Graph](image)

*Fig.5: Visualization of reported means for the variables (measurements on Likert-scale, 1-7, where 7 is the highest value)*

Our study demonstrated that the immersiveness of AR does not decline when outdoors, for both text and text+image AR overlays. This finding is supported by the fact that the participants also reported to have significantly more positive attitude towards both AR apps in comparison to the audio guide. The participants thought more highly of the two ex-
samples of AR technology in comparison to the audio - they found the AR apps to be more entertaining, enjoyable and useful.

<table>
<thead>
<tr>
<th>Flow</th>
<th>Audio (N=28)</th>
<th>AR text (N=29)</th>
<th>Audio (N=28)</th>
<th>AR text + image (N=28)</th>
<th>AR text (N=29)</th>
<th>AR text + image (N=28)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M(S.D)</td>
<td>4.50(0.42)</td>
<td>5.09(0.83)</td>
<td>4.50(0.42)</td>
<td>5.65(0.49)</td>
<td>5.09(0.83)</td>
<td>5.65(0.49)</td>
</tr>
<tr>
<td>ANOVA</td>
<td>F=7.353, p&lt;0.01</td>
<td>F=51.735, p=0.00</td>
<td>F=9.775, p&lt;0.01</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Perceived augmentation</th>
<th>M(S.D)</th>
<th>Audio</th>
<th>AR text</th>
<th>Audio</th>
<th>AR text + image</th>
<th>AR text</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td></td>
<td>F=5.874, p&lt;0.05</td>
<td>F=5.319, p&lt;0.05</td>
<td>F=0.035, p&lt;0.05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>App attitude</th>
<th>M(S.D)</th>
<th>Audio</th>
<th>AR text</th>
<th>Audio</th>
<th>AR text + image</th>
<th>AR text</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td></td>
<td>F=5.319, p&lt;0.05</td>
<td>F=0.035, p&gt;0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Behavioral intentions</th>
<th>M(S.D)</th>
<th>Audio</th>
<th>AR text</th>
<th>Audio</th>
<th>AR text + image</th>
<th>AR text</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td></td>
<td>F=725, p&lt;0.05</td>
<td>F=1.318, p&gt;0.05</td>
<td>F=0.039, p&gt;0.05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Means and standard deviations for the measures with ANOVA demonstrating the differences across the three groups

In terms of the participants’ ratings for their intentions to use such apps or guides again in the future, there was no significant difference across the three groups. While the AR apps, especially the ones with images, clearly created a more absorbing way of exploring the campus area, this did not translate into more pronounced intentions of using the AR apps more eagerly in the future in comparison to the audio guide.

Additional analysis was conducted about the different dimensions of flow (Table 3): control, attention focus, curiosity and intrinsic interest, aligned with measurement items validated by Webster et al. (1993) (as displayed in Table 1 above). The most important finding is that the only significant difference in terms of different dimensions of the flow between the two AR apps was reported in relation to the attention focus. The AR app with text + image made the participants significantly more focused than the text only. Displaying pictorial material makes a user more attentive to the exploration. The differences between the two AR apps in terms of perceived control or curiosity were negligible. Moreover, comparing the two AR apps with the audio guide, the AR app with text and image was reported to elicit significantly more perceived control and intrinsic interest than the audio guide, but such results weren’t reported for the AR app with text only. Overall, between the
two AR apps the one with added images stands out for its ability to call for more attention. In comparison to the audio guide, the text + image app performs significantly better on all but one flow dimensions.

<table>
<thead>
<tr>
<th>Control</th>
<th>M(SD)</th>
<th>Audio (N=28)</th>
<th>AR text (N=29)</th>
<th>Audio text + image (N=28)</th>
<th>AR text (N=29)</th>
<th>AR text + image (N=28)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>4.79(1.45)</td>
<td>5.45(1.64)</td>
<td>4.79(1.45)</td>
<td>5.45(1.64)</td>
<td>5.45(1.64)</td>
</tr>
<tr>
<td>ANOVA</td>
<td>F=2.608, p&lt;0.05</td>
<td>F=13.876, p&lt;0.01</td>
<td>F=2.216, p&gt;0.05</td>
<td>F=13.876, p&lt;0.01</td>
<td>F=2.216, p&gt;0.05</td>
<td>F=2.216, p&gt;0.05</td>
</tr>
<tr>
<td>Attention focus</td>
<td>M(SD)</td>
<td>3.84(1.86)</td>
<td>4.55(1.98)</td>
<td>3.84(1.86)</td>
<td>4.55(1.98)</td>
<td>4.55(1.98)</td>
</tr>
<tr>
<td>ANOVA</td>
<td>F=8.562, p&lt;0.01</td>
<td>F=48.680, p&lt;0.01</td>
<td>F=10.717, p&lt;0.01</td>
<td>F=48.680, p&lt;0.01</td>
<td>F=10.717, p&lt;0.01</td>
<td>F=10.717, p&lt;0.01</td>
</tr>
<tr>
<td>Curiosity</td>
<td>M(SD)</td>
<td>5.86(1.21)</td>
<td>5.97(1.18)</td>
<td>5.86(1.21)</td>
<td>5.97(1.18)</td>
<td>5.97(1.18)</td>
</tr>
<tr>
<td>ANOVA</td>
<td>F=117, p&gt;0.05</td>
<td>F=3.86, p&gt;0.05</td>
<td>F=0.62, p&gt;0.05</td>
<td>F=3.86, p&gt;0.05</td>
<td>F=0.62, p&gt;0.05</td>
<td>F=0.62, p&gt;0.05</td>
</tr>
<tr>
<td>Intrinsic interest</td>
<td>M(SD)</td>
<td>4.98(1.62)</td>
<td>5.53(1.27)</td>
<td>4.98(1.62)</td>
<td>5.53(1.27)</td>
<td>5.53(1.27)</td>
</tr>
<tr>
<td>ANOVA</td>
<td>F=2.057, p&gt;0.05</td>
<td>F=8.416, p&lt;0.01</td>
<td>F=2.681, p&gt;0.05</td>
<td>F=8.416, p&lt;0.01</td>
<td>F=2.681, p&gt;0.05</td>
<td>F=2.681, p&gt;0.05</td>
</tr>
</tbody>
</table>

Table 3: Analysis of the different flow dimensions per each group

5.2.2 Results of multiple choice test

A Z-test was conducted to examine the differences between the three apps in terms of knowledge retention (Table 4 and 5). For two of the five questions, the AR app with text and image resulted in significantly higher knowledge retention compared to when having only an audio guide. Specifically, these were:

Question 2: *In the early 19th century, the Gower street a) was situated in at the time important urban area; b) would give views even to some distant areas of London, c) was surrounded by forest or d) was an artistic and bohemian part of London*

Question 5 *Art Museum is home to: a) important collection of artwork dating back to 1200, b) Spanish artwork from the 16th century, c) work by Monet, Matisse and Leonardo da Vinci and d) collection that reflects British art education.*

   Question 2 referred to a historic outlook of the area that could be seen from the Darwin building. The AR app with text and image overlaid for that site an image of how that area used to look like in the past, which is the most likely explanation why those that used that app remembered the area better than those who only listened about its past out-
look. Interestingly, the AR app with only text also led to significantly higher number of responses in comparison to the audio guide only. However, as this was the single such example where an AR text-only app performed significantly better than the audio guide, further validation would be required to better understand the underlying reasons.

Furthermore, Question 5 referred to paintings, which are also very visual. The app with augmented text and image guide allowed participants to remember the content significantly better in comparison to the audio guide, as the images of the paintings offered strong visual cues.

<table>
<thead>
<tr>
<th></th>
<th>Number of participants per group who answered this question correctly (%)</th>
<th>Sum of total correct answers per group (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Question 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR app Text-only</td>
<td>17 (59%)</td>
<td></td>
</tr>
<tr>
<td>(N=29)</td>
<td></td>
<td>101 (70%)</td>
</tr>
<tr>
<td>Question 2</td>
<td>13 (45%)</td>
<td></td>
</tr>
<tr>
<td>Question 3</td>
<td>25 (86%)</td>
<td>95 (68%)</td>
</tr>
<tr>
<td>Question 4</td>
<td>26 (90%)</td>
<td></td>
</tr>
<tr>
<td>Question 5</td>
<td>20 (69%)</td>
<td></td>
</tr>
<tr>
<td>All questions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR app Text and image (N=28)</td>
<td>10 (36%)</td>
<td>14 (50%)</td>
</tr>
<tr>
<td>Audio guide (N=28)</td>
<td>19 (68%)</td>
<td>6 (21%)</td>
</tr>
</tbody>
</table>

Table 4: Overview of the right answers to the five multiple choice questions in the knowledge test across the three groups. For each question, the group with the highest number of correct answers is highlighted.

For Question 3 and 4, no significant differences were observed among the groups.

Those two questions asked about a name of a building or a person that was included in the app content. For example, Question 3 asked: Medawar building was embellished and improved by: a) Sir Albert Richardson; b) Japanese samurai; c) Charles Darwin; d) British artist Turner. In the instances when the exact information was required, the participants of AR apps did not outperform the audio guide users. That is a relevant finding related to learning outdoors with AR, showing this type of AR content does not lead to better retention of exact information such as specific names. AR can, however, significantly support knowledge retention of how something used to look like or particular objects, so in the cases when the visual cues give a better representation than narration or text. Previous litera-
ture in education and psychology has also shown that images embedded in text significantly facilitate retention of the textual information, provided that the images depict the textual information (Peeck 1993) and that the two are well integrated (Bodemer et al. 2004).

Interestingly, in Question 1, participants that used AR app with text and image performed significantly worse than in those that used AR app with text only or those that used the audio guide. The question asked: *In the Darwin building, Charles Darwin a) was preparing for a journey overseas; b) started writing his famous work on the Origins of the Species; c) lived with his family after his journey on HMS Beagle; d) conducted geological studies.* The correct answer was c) and large majority of all the wrong answers in the group that used AR app text+image indicated the answer b). One of the images that the AR app displayed on that site was the book cover of the Darwin’s most famous work. The possible explanation for these results can be that the image of the book cover remained salient in the memory and as Darwin’s name is most often associated with the Origins of the Species, that piece of information remained most prominent. That was not the case for the group that was not exposed to the image – in both the audio guide group and the text AR app group, about two thirds of participants indicated the right answer. It would be interesting to see if the

<table>
<thead>
<tr>
<th>Question</th>
<th>AR (Text) vs. AR (Text +image)</th>
<th>AR (Text) vs. Audio</th>
<th>AR(Text + image) vs. Audio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>z-score</td>
<td>p value</td>
<td>z-score</td>
</tr>
<tr>
<td>Question 1</td>
<td>-1.7315</td>
<td>&lt;.05</td>
<td>0.722</td>
</tr>
<tr>
<td>Question 2</td>
<td>-.391</td>
<td>&gt;.05</td>
<td>1.874</td>
</tr>
<tr>
<td>Question 3</td>
<td>.421</td>
<td>&gt;.05</td>
<td>.758</td>
</tr>
<tr>
<td>Question 4</td>
<td>-.427</td>
<td>&gt;.05</td>
<td>.816</td>
</tr>
<tr>
<td>Question 5</td>
<td>-.823</td>
<td>&gt;.05</td>
<td>1.194</td>
</tr>
<tr>
<td>All questions</td>
<td>0.3274</td>
<td>&gt;.05</td>
<td>1.5848</td>
</tr>
</tbody>
</table>

Table 5: Analysis of knowledge retention differences across the three groups. Significant different answers between the groups are highlighted.
knowledge retention would even out across the groups, had the image with book cover not been displayed.

In sum, the findings from the knowledge tests suggest that overall the participants in both AR groups performed better than those who used the audio content. A case-by-case analysis of the answers showed evidence that AR app with text and image leads to higher knowledge retention than audio guide. However, in the case where the AR image did not support the textual content, the AR images interfered with remembering the factual information. In the following section we discuss some of the findings that emerged from observing the participants using the app.

![Figure 6: Both content types (text+image on the left and text only on the right) for one of the scenes at Darwin building. The image of the cover remained salient in users’ memory to the extent that they misremembered the correct information about Darwin’s activity in this building.](image)

5.2.3 Observations during the campus tour

Participants appeared on the whole to be able to use the app when walking around the campus. In the image plus text condition, they held the tablet up at the three sites of interest and switched between the digital image and the real. For example, when approaching the Art Museum building at the third site, they focused on the paintings and kept approaching the
Museum while holding up the tablet – as if trying to step into the content or, more precisely, into the building where the paintings were displayed. Hence it seems that the AR condition with text and image delivered clearer cues to the participants what to focus on, acting as a reference point what to look at the three sites.

The researcher observed that some participants had difficulties with focusing on the overlaid content displayed on the tablet as they moved around the campus, especially as the campus was at times quite busy. The occasional flickering caused by tracking problems seemed particularly disturbing and some noted that the text was quite shaky when walking, remarking that it distracted from the pleasure of using the app.

Most of the participants in the two AR conditions were used to maximizing or enlarging the content on their smart devices by sliding two fingers on the screen in the opposite direction. This was not a function of the AR app, as they could maximize content only by physically moving closer to point of interest. Some did that, but still tried to enlarge the content with the two-finger pinch movement, as they wanted to get both the image and the text to be fitted in the frame at the point where they were standing. A couple of the participants commented on the low quality of the text resolution that became obvious when the content was enlarged by physically moving the tablet towards a site.

If users were moving away from the site and thus changing their relative distance to the real world site that was augmented, the augmented content was zoomed out and shrunk. They found it hard to relate the physical sites because of the minimized content, as one participant from the AR text group noted: “When reading from [a] distance, letters are too small, you want to enlarge them where you stand, without moving.” (P_{AR}1). A lack of features that could be triggered by touching the screen was perceived to be disturbing, as the participants are used to being able to scroll through the content. They remarked that the lack of navigation cues was confusing. In terms of content, this shows that clearly they expect to
manipulate AR like, say, a picture or a webpage and thus do not perceive it to be the physical environment, even though they interpret the augmented content in relation to the physical surrounding.

During the discussion after the campus tour, the participants came up with a number of comments about the two AR conditions. These are categorized in terms of (i) the role of the text, (ii) the role of the images and (iii) the effects of using it outdoors.

(i) Role of text

Participants noted that the text that appeared in the bubbles was of suitable length for them to read it on the screen. Another participant from the AR text group commented: “I like that there is not too much info and you get the most important.” (PAR2) They liked that the text was presented in bubbles so it was quick to read. Most of the participants in the AR conditions thought that having small bubbles of text was preferable compared with having one or two big bubbles of text – as it took up less of the screen.

The participants in both AR conditions emphasized that having additional information as an option should be available for those that wish to learn more. They expressed the desire for more control over the type of content that was displayed so that they could tailor it to their personal needs and preferences. As one participant from the AR text and image group commented: “Why is [the] info structured this way? I have general info to start but then I’m not sure if what you give is the most interesting for me. I have no choice to choose what I’m interested [sic] to see.” (PAR3)

(ii) The role of the images

Participants in the text and image condition noted that the images appeared as the dominant feature on their tablet screens with the bubble text playing a more supportive role. The im-
ages were considered successful at showing the details of a historic door handle (2nd site) and of paintings in the Art Museum (3rd site). When the visual door handle appeared on the screen, a participant asked: "Is this what I'm going to see if I go inside the building?" (P_{AR}4). They noted how the images enticed them to imagine more of what was hidden behind the walls of what they were able to see. However, it was not clear where the virtual object was in the real cultural artefact. For example, the handle was actually on the entrance door to the building, but some participants thought it was part of the interior design.

This kind of priming also led to more processing and their subsequent memory of the objects. As someone noted: “The most interesting is something that you can directly see, it’s more memorable...” (P_{AR}5). In contrast, some of the participants in the text only or audio conditions missed having visual material like images as a reference point, particularly on two occasions. One site where this occurred was when the audio talked about Darwin’s family and the other was at the Art Museum when the painters were mentioned. Participants mentioned wanting to see images of paintings in order to have a better understanding of what these artists produced. One from the audio guide group said, “They talk about paintings and I can't see them. Some visual clues?,” (P_{Audio}1) and participant from the AR text condition mentioned, “…maybe I want to know more about images, the paintings. An additional caption to the text would be good.” (P_{AR}6)

(iii) The effects of being in an outdoor context

In comparison to the AR apps, participants who used the audio guide mentioned how it was often difficult to relate the content to the particular sites and commented that listening to audio guide when outdoors was quite difficult. After having used the audio guide, a participant commented that: “There are other disturbances in outdoor space, it’s difficult to engage fully with content.” (P_{Audio}2). Another participant said that it was difficult
to remember what she heard: “…too many detailed things. I kept more peripheral info.” (P_{Audio3}). Interestingly, a participant in the text plus image condition pointed out that some awareness of the surroundings is crucial: “When in a busy environment you can get absorbed and you don’t want that ...you want to know what happens around you…” (P_{AR6}). The users want to keep the sense of control – not only with respect to the displayed content as the previous findings revealed, but also in relation to the surrounding when in a potentially overwhelming environment.

6 Discussion

The study has shown that the type of content in AR apps can have an effect on how it is used outdoors. Previous work (Chang et al. 2014) compared the impact of an AR app and audio guide in a museum. Our study extends those findings by contrasting two different types of AR apps and demonstrates how users take to them in an outdoor context. Guided by our research questions, we found that the experiences differ depending on which AR app the participants used. These differences showed in the level of flow, participants’ perception of how the app enhanced their visit of the difference sites and also knowledge retention.

AR apps with textual information and images created a more engaging experience and stronger immersion during the exploration of the sites. AR content acts as primers and reference points, suggesting what to look at, while audio guide does not permit such contextualization of the information. Moreover, having both images and bubble text in the same overlay appeared to work particularly well for coupling the background information with additional visual layers that referred to some characteristics of the sites. The role of the images was typically to support the content conveyed by the textual information. The text and image condition augmented the historical sites in a new way, which created a stronger im-
mersion and was able to go beyond simply delivering informational facts about a site. Instead, that particular AR app provided a sense of how the site used to look like, visualizing the important personas related to it, drawing attention to details such as a door handle and revealing more about what was inside the building.

However, such images seemed to occasionally provide cues that hinted beyond the intended content. While we did not provide the participants with instructions that the overlaid images necessarily reveal the interior of the buildings, they occasionally assumed that the images were displaying the insides of the buildings even when that was not the case.

The overall memory for the information was also better in the AR apps in comparison to the audio guide, suggesting that people retain the information better when they see it directly as part of a site. The AR app with text and image can lead to significantly better knowledge retention if the overlaid images visualize an artefact that is harder to depict with words, such as a painting. However, the images can also distract the participant in the sense that she pays less attention to the exact facts and thus risks misremembering them. Again, the images seem to carry more potential for ambiguity and different interpretations of how they are related to the physical environment. The selection of visual and textual content that complement each other well is key, however only linking well the visual and textual overlaid content is not sufficient for delivering a clear understanding of what the overlay refers to in the real environment. More precise design solutions are required that will provide intuitive understanding how the graphics complement the physical objects.

Despite our initial concerns about whether external factors would detract from the AR experience, the participants appeared both engaged with the app but were also able to navigate the outdoors relatively easily. Hence, combining digital images and text on the screen worked well as it enabled the participants to know what to look for and learn more about them. They were able to maintain the sense of the surrounding and actually explore it more
by moving around and approaching the points that AR enhanced. The audio however, appeared more problematic as surrounding noise could impede the audio narrative.

Moreover, because the AR overlays the physical environment in real time, it does not require of people to distance themselves from what they see. Rather, they often stepped close to the building as the overlaid content creates a more direct connection between the site and the information. In the audio condition, the participants have to look for what they are being told about – this can require more cognitive effort and the moment may be gone by further audio being played. There is less ability to pace and match what is being heard with what is being seen. AR augmentation, in contrast, enables people to approach an area or a site and hone in on what is being visually shown on the screen in a different way than if the information is being narrated.

7 Design recommendations

7.1 Level of interactivity

The apps were designed in the way that required or allowed very little interaction besides pointing the tablet towards a particular site. While the content could be enlarged or reduced in size simply by approaching or distancing oneself from a point, participants expressed a desire to have more control over the content. Being able to remove those information that are not of interest or enlarging the content by tapping or sliding fingertips on the screen might be a way of improving the design of similar apps. While the features should remain rather simple for outdoor environment, the control should not be completely taken away from the user. Rather, the app should support the interest that users have in knowing more about a particular area and should offer options for more in-depth learning.

The fact that outdoor environments are often to some extent stressful needs to be taken into account too. In our study, users were guided where to point the tablet in order to
view the AR content. Such assistance, however, cannot be always provided. Therefore, other ways need to be invented to signal to the users how to activate markers and access the content, but also to remain aware of the surrounding.

7.2 Combining information

People have different interests in terms of what they would like to see or learn about and providing more choice over the information or images that would be visualized could prove beneficial. But more importantly, users need to have an understanding or an intuition of what type of information is the app delivering – is it more focused on factual information or rather on historical development of buildings? Does it reveal insights of historic personas lived? Given that especially the pictorial cues can be interpreted in an unintentional way, designers need to specify with more clarity how the AR content complements the physical world.

Providing the right amount of information when using an AR app in situ outdoors is also crucial. In our study, we provided 3 – 5 facts about each site and participants’ feedback indicated that was an appropriate amount. The issue related to the “right amount of text” is a general concern in digital content design – especially when designing for mobile applications (Cutrell and Guan 2007; Kim et al. 2017). Cutrell and Guan (2007) for example, show that the length of the snippet should vary depending on the user task – navigation requires shorter snippets, while for informational task, longer snippets are more appropriate. An AR overlay combines both navigational (because it contextualizes information to a specific place or site) and also informational value. Valuable AR content needs to strike a balance between stating clearly at which particular physical object the overlaid content is aiming and also providing enough cues to give the user a sense of what the particular content repre-
sents. If interested to know more, the longer snippets can then appear when the user has tapped or triggered them.

8 Conclusions

Our in situ study showed that people have a positive experience with using AR apps when outdoors, especially when a combination of small amounts of text and images that relate to historical artefacts appear in front of them. AR can enhance the discovery and learning experience because it is able to provide connections between the physical and digital - compared with the audio that gives less of an orientation what the information relate to in the physical environment. A combination of both images and text seemed to enhance the user experience by being able to ‘guide’ the user to know what to look for and read interesting snippets related to that. The images open up the sites in a manner that immerses the participants deeper into an exploration, but they can also give way to ambiguous interpretations as to what they represent. More intuitive and interactive AR outdoor systems would lead to exciting new ways for users to get immersed in outdoor sites exploration.
Bibliography


