Evaluation of the AR Explorer: A Study of Interaction Techniques for Handheld Augmented Augmented Reality

Mikkel Mogensen
Aalborg University Copenhagen

Raminta Stanulionyte
Aalborg University Copenhagen

January 11, 2018

Introduction

The renewed interest in Augmented Reality (AR) has seen a wave of applications for handheld devices such as smartphones. There have been talks of several issues plaguing AR since its inception i.e. hardware limitations, cost and so forth. But with the advancement of mobile hardware and manufacturing we are nearing the end of that chapter. In fact, many smartphones are already capable of supporting high-fidelity AR experiences when the software is made available. At present, the problems are first and foremost, our lack of understanding of perceptual and ergonomic issues. This has huge ramifications such as poor usability and user experience. Furthermore, inadequate ergonomic considerations may lead to fatigue and discomfort from prolonged hand or arm poses as well as discomfort related to unnatural gestures and/or improper handling of the device. Apart from the more obvious issues, we also have to ask ourselves who these types of applications benefit and how much they should be allowed to take over our lives. Certainly, consequences of augmentation are real points of concern, as they are capable of transforming the way we engage with the mundane world.

The main goal of this essay is to identify the remedies for some of these problems plaguing AR systems with the focus being on interaction. We seek to find answers to these problems based on a systematic litterature review, and conjectures based on our previous research that focused on Handheld Augmented Reality (HAR) for museums (Mogensen & Stanulionyte, 2017a). As our study was primarily centered around application within cultural heritage, the report will be based on such discourse. We define our problem statement as follows:

– *How can we create a useful and ergonomic HAR application for museums?*

The following structure will be used throughout the essay. In section 1 ("Summary of the AR Explorer") we give a brief rundown of our previous work. We then explain a human-centered approach in section 2 ("Human Centered Design") for designing products with focus on the end-users. Section 3 ("A Taxonomy of Mixed Reality") explain taxonomy that is imperative to the understanding of VR and AR. In section 4 ("Designing for interaction") describe how we can meet the needs and capabilities of the users based on cognition, perception and ergonomics. We then evaluate different interaction techniques in section 5 ("Interaction Techniques"). Each section is followed up with a discussion and finally our conclusion is drawn in section 6.

Summary of the AR Explorer

In the following section, we take a look at the 'AR Explorer' that was the outcome of the research and development presented in our paper "*How gamification motivates: an experimental study of the museum experience using handheld augmented reality*" (Mogensen & Stanulionyte, 2017b).

The resulting application allows users to select and explore features on the museum exhibit using a 3D cursor overlaid on the object. Information about the feature is then made visible through the interface. At the same time, a 3D representation of the museum object provides a detailed view of the selected feature. Furthermore, the application gives users the ability to freely explore; this is based on research which suggests that "*[...] diversion, curiosity and spontaneity are more characteristic of visitor intentions than structured learning*" Kotler & Kotler (2000). Users can explore features in any order and revisit features later on if desired.

Two different views were implemented in the application: E*xploration view*, F*eature view*. What follows is a description of each of these views, their purpose, features, and interfaces.

Exploration View

The exploration view is visible once the application starts. In the exploration view the user may view and select features on the museum object. Features that can be explored are shown using indicators overlaid on the surface of the object. Navigation between features is accomplished by pointing a 3D cursor on any of the indicators and tapping on the screen. Tapping triggers a smooth transition where a 3D representation of the museum object moves to frame the feature so it is displayed in the center of the screen.

Feature View

After selecting a feature users are transported to the feature view. The screen is divided into a *viewport* (top) framing the feature, and *interface* (bottom)

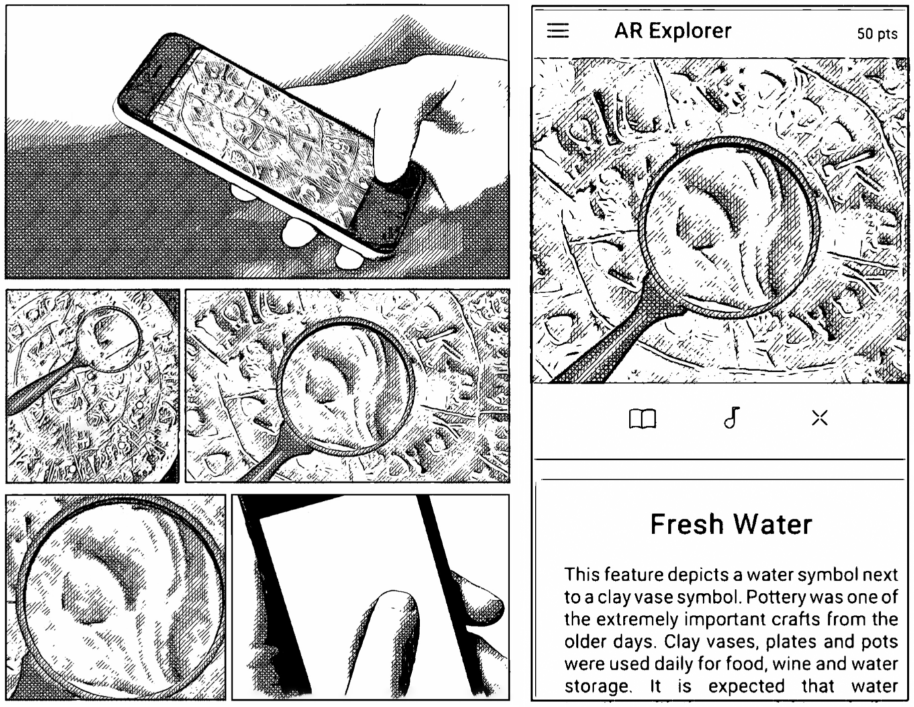

Figure 1: Storyboard of the *AR Explorer*. The user points the device on the object. A 3D cursor (magnifying glass) is shown at the location he is pointing. To explore a feature in further detail the user hovers the cursor over the feature and taps on the screen. The feature is framed in the screen (on the top half) and information is shown (on the lower half).

containing various GUI elements (text, images, buttons). The user is unable to navigate to other features while in the feature view. The layout of the interface is split into two parts: a top navigation bar and a content canvas containing a card. User can navigate the interface by tapping or sliding elements on the screen. On the lower half of the interface, in the content canvas, the user can scroll the text. In the navigation bar the user can select different options such as voice-over by tapping the note icon or exit the current activity by pressing the 'X' button.

Interaction

The interaction metaphor used for the *AR Explorer* is that of a magnifying glass. The user is able to inspect an object by pointing and magnifying features. This is done by pointing the device towards the point of interest (i.e. a symbol on the object) (refer to figure 2). A ray is cast to the object intersecting a 3D topology of the real object in world space. We then superimpose a 3D cursor shaped like a magnifying glass at the coordinate given by the ray. The magnifying glass displays a portion of a virtual copy of the physical object to the user. If the user taps while pointing at the feature this 3D representation of the object is interpolated to fit the screen space (refer to figure 3).

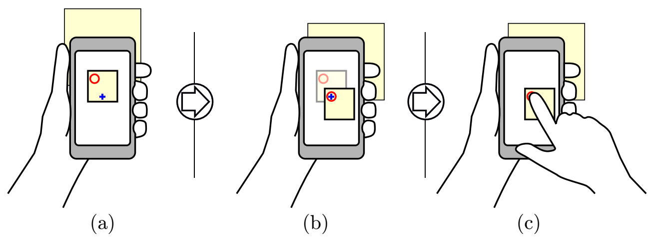

Figure 2: Illustration of interaction with the 'AR Explorer'. **(a)** The user points the phone on the museum exhibit. The red circle indicated a point of interest (POI) on the object (green box). **(b)** The user moves the view to position the POI to the centered crosshair that is marked by the blue cross. **(c)** The user taps anywhere on the screen to confirm the selection.

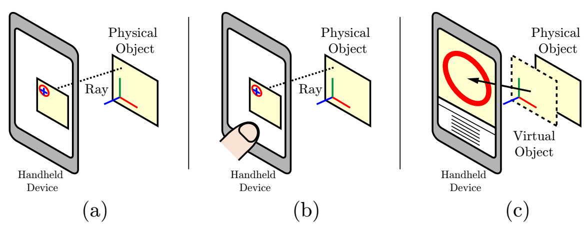

Figure 3: **(a)** A ray is cast onto the physical object and represented to the user on the screen. The ray that is isomorphically controlled by the device's orientation. **(b)** The user taps anywhere on the screen when the ray intersects a POI **(c)** a 3D representation of the physical object is interpolated from its origin to fit in the screen space; this frames the selected POI. Information about the POI is presented at the bottom half of the screen.

Human Centered Design

In this chapter, we give a brief rundown of the International Organization for Standardization (ISO) standards for User-Centered Design (UCD) and specification of usability. User centered design is crucial for designing artifacts for end-users that are useful and compliant. According to ISO 9241-210:2010 "Human-centered design for interactive systems", the approach enhances

"[...] *effectiveness and efficiency, improves human well-being, user satisfaction, accessibility and sustainability*". One should note that this approach involves not only the end-users but also stakeholders (the companies or recipients who order the applications or services) (Earthy et al., 2012a) for this reason some prefer to use the more general term, "human-centered" over "user-centered" throughout this chapter we will use the term Human Centered Design (HCD).

Guiding Principles

Interactive systems are present everywhere and we encounter them on a daily basis. Many typically follow an international standard provided by a committee of experts and the goal of such standards is to offer consistency and disciplined evaluation (Stewart 1991;Earthy et al. 2012b). Strict guidelines are in place and instated in ISO 9241-210:2010 Human-centered design for interactive systems. Among them are the following principles:

- The design is based upon an explicit understanding of users, tasks and environments
- Users are involved throughout design and development
- The design is driven and refined by user-centred evaluation
- The process is iterative
- The design addresses the whole user experience
- The design team includes multidisciplinary skills and perspectives

Knowing and understanding HCD should bring the designer one step closer towards the successful application that caters users needs, but one should be vary of the caveats of such framework. For example, it is a demanding procedure with an expense on resources as carefully planning should be implemented to aid each of the stages of development (Dvir et al., 2003). We can illustrate the stages of development as an iterative cycle (see figure 4). ISO 9241-210:2010 demands the process be repeated over several iterations as it helps progressively eliminate uncertainty. This implies that the design specifications at the beginning of the development are never sufficient as needs and expectations of users usually emerge in the course of development.

Discussion

ISO 9241-210:2010 "Human-centered design for interactive systems" provides consistent activities for development throughout the product life cycle. The standard is unambiguous and reviewed by a board of experts to ensure up-to-date paradigms based on evidence. Hence we find HCD to be highly suitable and worthwhile for the development of an AR application in cultural heritage, inasmuch as our focus is on providing satisfactory user experiences for museum visitors. In the next chapter ("A Taxonomy of Mixed Reality") the idea of the human centered approach will remain central for how we define problems

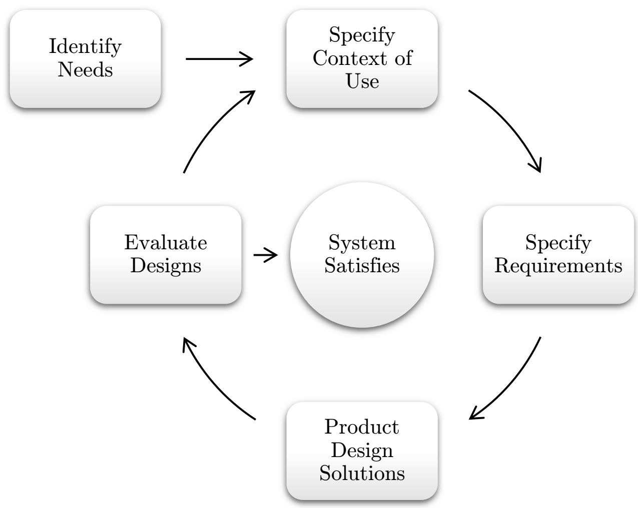

Figure 4: The general phases of the User Centred Design process (Usability.gov, 2015)

and seek solutions based on a classification of 3D interfaces for mixed reality applications.

A Taxonomy of Mixed Reality

What is virtual and what is real? Depending on who we ask, we might get a different answer. Some would say the projections of the alleged material reality is no different from actual material impressions, as they themselves are indirect representations. According to Ariso (2017), the definition very much depends on the importance we attach to the discourse. The relevance is given by the practical operations that result from that discourse, that is, the degree of sophistication and inclusion of the branches of science in our discourse yields the final answer. In this chapter we will introduce the notion of Mixed Reality (MR) and as our approach is human-centered, we leverage our discourse from interpretations of human senses and cognition. This will help us later to classify interfaces for interaction that align with our goals.

Definition of Taxonomy

Having a shared understanding is important as it bridges the communication gap between researchers and practitioners. Oftentimes taxonomies are used for that purpose. A taxonomy can be thought as a delimitation of an observable phenomenon into taxonomic units known as '*taxa*'. According to Covington &

Slaughter (1983), arriving at a proper scientific taxonomy is a process of decontextualization and abstraction. We can think of the former as the isolation of a phenomenon from its natural context i.e. the seventeenth century saw an interest in classifying plants. We see from the evolution of herbals, that more time was devoted to the plant structure (features, size, shape, color etc.) than recounting them (Covington & Slaughter, 1983). Technologies can likewise be decontextualized and classified (i.e. the classification of virtual and augmented reality) by virtue of the scientific 'botanic' taxonomy. We see that such a transposition is, necessary because it allows us to objectify such that we can analyze and arrange (or classify) the phenomenon. As Covington & Slaughter (1983), points out not only does decontextualization objectify, it also abstracts. This stems from the fact that any form of representing an object necessarily abstracts i.e. we can see photographs as abstractions of real objects, but photographs abstract less than other mediums such as a drawing. Fernandez & Eastman (1990)

But what makes a taxonomy relevant? Perhaps the logical answer is, that the relevance is determined by researchers and practitioners. 'Relevance' could then be said to refer to the consensus about the importance of contributions in a field. One way to measure this is by citation counts (Hargens & Schuman 1990; Tijssen et al. 2002), but whether the taxonomy is useful or not, we see as independent of such 'scientific balloting'. We believe this misconstrued overlooks valuable contributions by creating immutable power structures, a view shared by social constructivists Turnage (2010). With that in mind we allow ourselves to explore alternative discourses that that potentially allow us to circumvent dominant discourses. One of those dominant discourses is the one surrounding VR and AR, in particular 'Mixed Reality'.

Mixed Reality Definition

The notion of "Mixed Reality" (MR) can be traced back to an academic paper by Milgram and Kishino Milgram & Kishino (1994). At the time, Virtual Reality (VR) was used erroneously to describe environments to which certain qualia such as total immersion and complete synthesis did not necessarily pertain. The need for a definition eventually arose. In their paper Milgram and Kishino, therefore, formulate a taxonomy to describe a subclass of VR related technologies that involve the mixing of real and virtual elements.

According to Milgram & Kishino (1994), the merging of virtual and real MR aspects can be realized in a spectrum. As some systems are primarily video-based and enhanced by computer graphics and vice versa, it is practical to classify these systems in what the authors refer to as the "virtuality continuum" (see figure 5) . The definition asserts that anything between the two endpoints is to be considered MR. The MR system can be any proportion of the completely real environment (left) and completely virtual environment (right).

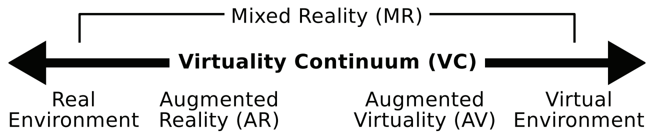

Figure 5: The Reality-Virtuality Continuum. Reprinted from "A Taxonomy of Mixed Reality Visual Displays" by Milgram & Kishino 1994 IEICE Transactions on Information Systems, Vol E77-D, No.12

Taxonomy of Mixed Reality

The taxonomy presented by Milgram & Kishino (1994) defines three dimensions of which we can classify any Mixed Reality application. These dimensions include Extent of World Knowledge (EWK), Reproduction Fidelity (RFI) and Extent of Presence Metaphor (EPM).

Extent of World Knowledge can be expressed in layman terms as the degree of knowledge (i.e. geometry, location, and attitude) the MR system posses about the world. It ranges from the system having no knowledge about the world (World Unmodelled) to the system having a complete model of the world (World Completely Modelled). It goes without saying that the system can be anywhere between (World Partially Modelled). The EWK dimension operates with three subcases: '*Where*', '*What*' and '*Where + What*'.

'*Where*' refers to cases in which quantitative data about location, orientation, and scale in the world is available, but it does not provide any information about objects in the world. To know what is there, the system needs information about the object's geometry. Supposed that a 3D scanning technique was used to capture the object. The MR system may use this 'What' information as a part of its registration procedure. But in order to superimpose information on the object both cases are needed (*Where + What*).

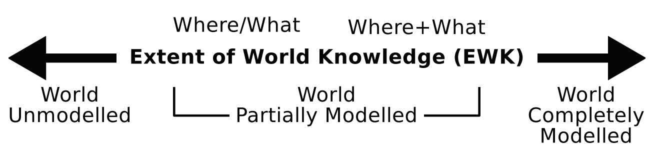

Figure 6: Extend of World Knowledge (EWK) indicates the degree of information about the (remote) world. Reprinted from "A Taxonomy of Mixed Reality Visual Displays" by Milgram & Kishino 1994 IEICE Transactions on Information Systems, Vol E77-D, No.12

Reproduction Fidelity addresses the quality with which the synthesizing display is capable of reproducing images of objects being displayed. This dimension pertains to the reproduction fidelity of both real and virtual objects. On one hand, we have to deal with the fidelity of display hardware (above the axis), because we interact with it directly. There are several things to consider such as resolution, color, stereoscopic video and so forth. Furthermore, as these displays facilitate the viewing experience, the RF dimension also deals with reproduction of the images (below the axis). It considers the graphics rendering ranging from simple wireframe models to the point where real and virtual objects cannot be distinguished by an observer.

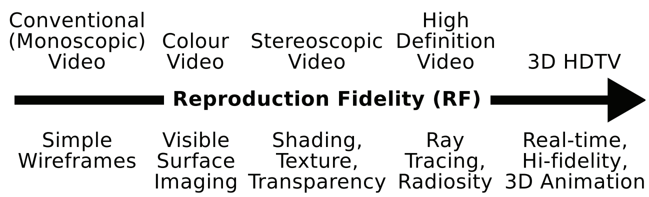

Figure 7: Reproduction Fidelity (RF) deals with the quality of reproduction of graphics rendering and the hardware that facilitates the MR experience. Reprinted from "A Taxonomy of Mixed Reality Visual Displays" by Milgram & Kishino 1994 IEICE Transactions on Information Systems, Vol E77-D, No.12

Extent of Presence Metaphor relates to the extent to which the feeling of presence is accommodated by the MR system. It ranges from a fixed monoscopic viewing of images to a metaphor of "real-time imaging" where images are indistinguishable from unmediated reality. "Monoscopic Imaging" as mentioned is fixed meaning, no tracking occurs at this stage. The adjacent definition "Multiscopic Imaging" on the other hand includes multiple viewports. Lateral movements of the user's head are tracked, but it is by no means completely immersive. This sort of display capability has been referred to as "fish tank virtual reality". Panoramic imaging is a further extension to this as it also tracks the orientation of the user's head. This can be realized either with a Large Screen where working volumes are restricted. HMDs are seen as a more inclusive instantiation of this class as these displays support the metaphor of being on the inside rather than on the outside looking in. "Surrogate travel" refers to the ability to move around within the world being viewed and "Realtime Imaging" refers to the solutions of temporally related issues such as updates rates of the display or simulations of dynamics.

The Stimuli-Framework Model

While the proposed taxonomy by Milgram & Kishino (1994) is certainly useful, there are cases of where the correct classification is debatable. According to Rosa et al. (2016) one example, is whether digitally captured and displayed elements are real or virtual. Rosa et al. (2016) also brings up an interesting discussion of whether an application in which real occurs in one modality and the virtual

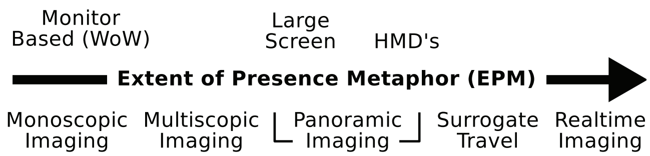

Figure 8: Extent of Presence Metaphor relates to the extent to which the feeling of presence is accomodated by the MR system. Reprinted from "A Taxonomy of Mixed Reality Visual Displays" by Milgram & Kishino 1994 IEICE Transactions on Information Systems, Vol E77-D, No.12

in another should also be considered AR. In this section we draw attention to points of critique and proposals that may guide the current discourse.

One issue that Milgram & Kishino (1994) has admitted to is that their classification does not differentiate between a remotely viewed video scene versus unmediated scenes i.e. looking at ones hand versus looking it through a video display is considered the same. The latter has been labeled *virtual* by some scholars Rosa et al. (2016), but such an erroneous use of the terminology is problematic as it leads to confusion, and we ought to find an agreeable definition. According Mann (2002) remotely viewed video scenes are distinct from unmediated scenes. Therefore, we should think of the environments on two spectrums: a real-mediated and a real-virtual spectrum. Furthermore, Müller (2015) classifies information in procedural tasks in AR into five conceptual layers: (i) the physical world, (ii) the mediated physical world, (iii) virtual objects that are spatially referenced and of spatial nature, (iv) virtual objects that are spatially referenced but not of spatial nature and (v) virtual objects that have no connection to the physical world. Milgram & Kishino (1994) and Müller (2015) fundamentally agree on their discourse i.e. both discuss real and virtual in terms of objects in AR. Mann (2002), however, focuses his analysis on the fact that AR is not only restricted to augmentation of the real environment by virtual objects but also allows for modification of real objects. we should instead consider those elements that make up the perception of our environment namely stimuli. Rosa et al. (2016) sees Mann's classification as the most adequate as it describes mediation as digitally modified stimulus and propose the following stimuli-framework:

> Real – the stimulus originates from the physical environment and is perceived without any form of mediation, by its intended modality.
>
> Mediated – the stimulus originates from the physical environment and is perceived through a digital sensor-display mediator, either by the intended modality (for which it may have been altered) or other modality(ies).

> Virtual – the stimulus originates from a computer-generated model, and is perceived through a digital display, either by the intended modality or other modality(ies).

According to Milgram & Kishino (1994) and Azuma et al. (2001) the basis of AR must be a 'real' environment whilst the augmentation is 'virtual'. The stimuli-framework, on the other hand, states that the basis does not need to be strictly real and augmentation does not need to be strictly virtual. There are instead two options of the basis *real* and *mediated* and for the augmentation which can be either *virtual, mediated* or both.

Discussion

The notion of mixed reality proved insufficient as it does not adequately distinguish between real and virtual. A redefinition of the term as proposed by Rosa et al. (2016) proved capable in this regard, by the introduction of the stimuli-framework and a basis-augmentation model. Further proposals by the same author extend the definition of multimodal AR, however, these are beyond the scope of this essay. Using the frameworks discussed in this chapter we find it easier to classify our museum application. According to the stimuli-framework we can classify the stimulus provided by our *AR Explorer* application as a mixture of mediated (the stimuli originates from the physical environment, and is displayed through a digital sensor mediator) and virtual (the stimuli is computer generated and displayed through a digital sensor mediator). We say the stimuli is mediated because the user interacts with the real museum, however, this interaction is facilitated and mediated by the smartphone display. Likewise, we say that the stimuli is virtual since information about the museum exhibit is generated i.e. the 3D cursor and markers indicating points of interest on the object. According to the basis-augmentation model, we can then classify the *AR Explorer* as Mediated-Virtual.

In the next chapter we will delve further into designing for interaction. Further on in the chapter ("3D interaction techniques") we will use our classification as a guide to find interfaces that align with our goals.

Designing for Interaction

In this section, we will be discussing the design principles of interaction tasks for 3D interfaces. To be more specific we will keep our focus on interaction within the field of handheld mobile AR devices. Since mobile AR technology is becoming accessible for users on their private devices, researchers deliver more and more papers on various issues and solutions within the field. Having a deep understanding of those issues and solutions could ease design and development processes.

Design Guidelines

During the design process of a 2D interface, developers' main concerns are how to design a system that would be understandable, intuitive and easy to use LaViola. et al. (2017). In addition, 3D interface designers need to consider the physical actions performed by users. While interacting with 3D interfaces users often have to perform tasks while wearing mobile devices on their bodies, for example, a head-mounted display for VR or AR, and/or holding devices in their hands such as controllers, phones, tablets etc. These factors ads additional constraints towards the actions that have to be performed during the use of an application. Furthermore, it contributes towards the user's fatigue. Moreover, not all of the mobile devices for 3D interfaces are wireless. Cables connected to such devices add further to the discomfort of using the application. As a solution to aforementioned issues, developers try to focus on a low weight, comfortable and wireless solutions that would not divert users attention from the preeminent task of interaction with the 3D interface LaViola. et al. (2017).

Since our research is highly focused on a mobile handheld AR system we are aware that the use of such system requires the whole body participation. Users not only have to hold the phone in their hands, in addition, they must walk around to perform observation tasks. The application should be designed in a manner that would not produce users' fatigue. In the book "*3D User Interfaces Theory and practice*" chapter 10 LaViola. et al. (2017) it is advised to design short sessions for the application use if interaction requires physical strain, such as holding the device at a certain height. Moreover, from the same book, we learn to keep the height of the hand's position close to the body to produce a comfortable user experience and reduce fatigue. We will be discussing the alternative solutions dealing with hand fatigue on extensive use of a mobile device in section "Human Factors".

When it comes to the user's body constraints we should not only focus on comfortable poses but as well understand that bimanual interaction (two-handed) is required to perform various tasks on a handheld AR device. Guiard's framework of bimanual manipulation explains the principals of various actions performed using two hands Xia et al. (2007). The bimanual movements of hands can be divided into two types: symmetric and asymmetric. Symmetric operations are those where both hands perform the same or similar movements and they are equally important, while in asymmetric operations both hands have different roles Xia et al. (2007). By human nature, the latter is more commonly used since people tend to purpose one hand, the Dominant Hand (DH), to acute and precise tasks while the other hand or Non-Dominant Hand (NDH) serves only as a support to the dominant hand (figure 9) Guiard (1987). Even though the bimanual operations are considered to be an instinctive way of completing tasks with minimum training required, the diversity of possible design solutions for such activities leaves an open question regarding the task distribution between DH and NDH Xia et al. (2007).

Hand	Role and Action
Non-Dominant (NDH)	- Leads the dominant hand - Sets the spatial frame of reference for the dominant hand - Performs coarse movements
Dominant (DH)	- Follows the non-dominant hand - Works within the established frame of reference set by the non-dominant hand - Performs fine movements

Figure 9: The roles of both hands in Guiard's Model. Reprinted from *Introduction: Toward a Multidisciplinary Science of Human-Computer Interaction* by Carroll, 2003, San Francisco: Morgan Kaufmann. Copyright by 2003 Elsevier Inc.

While delegating tasks between two hands it is important to determine which hand outdo in performing the selected task, and what role it plays in comparison to the inferior hand. To discover the roles of DH and NDH and task distribution between them, Xia et al. (2007) offers to follow Guiard's research on bimanual interaction and kinematic chain model (Guiard 1987, MacKenzie & Guiard 2001) as well as their own research.

The aforementioned information on bimanual controls for tasks found from literature should guide the design process while constructing interaction methods. This allows for smooth intuitive and effortless user experience avoiding disturbance in the flow of performing tasks (Neuburger & Egger, 2017). In addition to this, it is important to take into account other human factors that influence performance and user experience.

Human Factors

Human factors have a major impact on application usability and performance. They also play the leading role in the human centered design process. That being said one must have a good understanding of perceptive, cognitive and physical characteristics as well as limitations of human beings in order to design a comfortable, reliable and efficient system LaViola. et al. (2017). To dive deeper into human factors it is essential to acquire the general knowledge of how people process any given information and different stages of this process. Researchers Wickens & Carswell 2006 divides information processing into three different stages corresponding to three major factors: perception, cognition, physical ergonomics (see figure 10). Each of the stages analyzes the series of actions occurring between an event or stimuli and a user response to it.

The circle of information processing starts with an event or stimuli analysis using the perceptual mechanisms within the human body (see figure 10). According to Oxford Dictionary perception is "[...] *the ability to see, hear, or become aware of something through the senses*" *The Oxford English Dictionary*

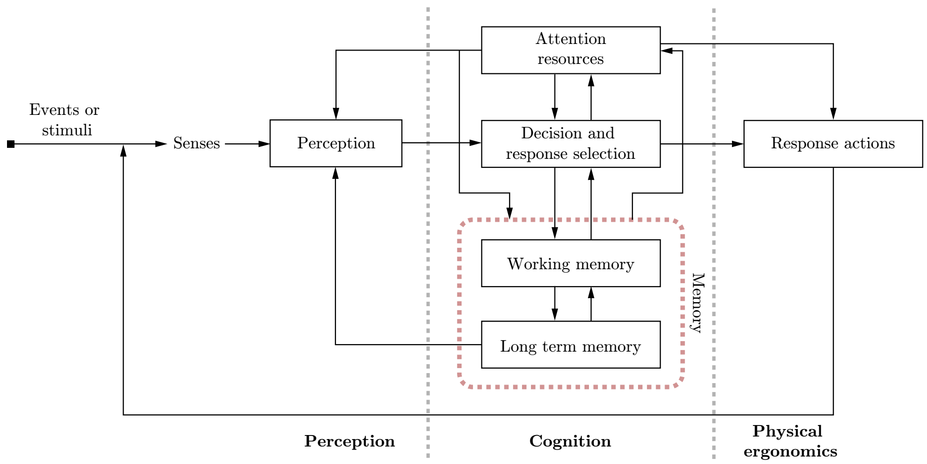

Figure 10: Information Processing diagram. Adapted from *3D User Interfaces: Theory and Practice 2nd Edition* (pp. 35-41) by LaViola. et al., 2017, Hoboken, NJ: Addison-Wesley. Copyright by 2017 Pearson Education, Inc.

Second Edition (VOLUME 11) (2001). While designing interaction for interactive 3D systems it is essential to follow the guidelines of perceptual cues to avoid visual and non-visual perceptive issues. Since perception is a long-studied subject guidelines for interaction design can be found in various research papers depending on the desired interface. Perceived information or stimuli is further evaluated by cognition (see figure 10). During cognitive evaluation phase users mainly face two issues: mental load (amount of cognitive work required to complete the task) and error (failing to perform the task) LaViola. et al. (2017). Exceeding the mental load barrier contributes towards the lower performance and mental fatigue. According to LaViola. et al. (2017) this could also be the cause of causes errors during the interaction. Errors may also occur due to faulty interaction design and ambiguous guidance (in regards to visual and non-visual cues). The designer must be well equipped with knowledge on cognitive processes to guarantee effective usability of the application. Once the cognitive evaluation is completed the user is ready to take an action.

Physical Ergonomics

It is a common tendency to dedicate a plethora of resources and time evaluating cognitive and perceptual issues while designing for 3D interfaces LaViola. et al. (2017). The physical ergonomics should not be left out when evaluating human factors, considering it is a major influencer to comfortable, effective and attractive interfaces. Physical ergonomics focus on the musculoskeletal system and physical activities related peculiarities: anatomical, anthropometric, physiological and biomechanical The International Ergonomics Association (2017). In the context of interface design and human-computer interaction, the physical ergonomics has a high dependency on anatomical capacities that determine how and how efficiently one can comply a specific task LaViola. et al. (2017) . When the human capacities such as load or pain tolerance are exceeded users start to feel fatigue. This prompts them to change pose i.e. lower arms, hand or alternate between the hand that holds the input device LaViola. et al. (2017). Analogous to fatigue comes user comfort. The term refers to the ease of use and effortless interaction. It includes various factors such as comfortable body posture (while interacting with the input device), the grip on the device and/or hand motions for interaction. Fatigue and user comfort are highly interconnected and they both have a large impact on users physical and psychological well being and overall performance LaViola. et al. (2017).

Extensive knowledge of human factors and physical ergonomics can help to avoid aforementioned issues and aid designers to create comfortable effective and user-friendly interaction techniques for 3D interfaces.

Discussion

While designing interaction for 3D interfaces there are many factors to consider and a plethora of guidelines to follow, however, there is no one true solution that fits all. Each application depending on the goal of the task as well as the device it is being designed for requires a personal evaluation in order to achieve a user-friendly ergonomic and efficient user experience. Human factors have a major impact on such design since they are directly linked to performance, user comfort and user well being. Understanding perceptive, cognitive and physical capacities of human beings work as a foundation guiding us towards successful application design for the 'AR Explorer' interface. With the design guidelines ready in the following chapter, we will shift our focus towards interaction techniques that would fit best to our AR application design.

Interaction Techniques

In this chapter, we look at interaction techniques that fit in the context of the *AR Explorer* and relate them to our approach. The goal is to to make a judgment about our design and how to improve it using solid interaction patterns.

Criteria for Interaction

According to the basis-augmentation model (refer to chapter) our application can be classified as *mediated-virtual*. We set out to find interaction techniques of similar classification. Specifically, since our basis is mediated by a video display and the augmentation is derived from a computer-generated model, we should be able to find compatible methods in this specific subdomain. Furthermore, we are interested in techniques related to pointing and selection, as these are the main interaction schemes used in the *AR Explorer*. We define the following criteria for our evaluation of interaction techniques:

- Ergonomics are prioritized
- Same classification (*mediated-virtual*)
- Chosen techniques should work in tandem with a smartphone or similar device.

Pointing Task

Pointing techniques are widely used as they are easy ways for users to interact with content. We are perhaps most familiar with the traditional mouse interface which can be used to position a cursor on a screen, and to manipulate containers or objects in the operating environment. To people who are familiar with this type of interface, it might seem natural, but think about the intuition (the control display relationship) of moving the mouse from left to right. This prompts a reciprocal and spatial congruent response from the cursor MacKenzie (2013). However, when moving the mouse forward and backward the cursor cannot respond accordingly, instead prompting a spatial transformation. By design the cursor moves up or down on the screen which can be difficult to some novice users Douglas & Mithal (1997). 2D interfaces usually operate with a 2 degrees of freedom (x, y), but more can be added such as rotation about a hidden axis. When designing pointing techniques in 3D we are potentially dealing with 6 degrees of freedom. Three for position (x, y, z) and three for rotation $(\theta x, \theta y, \theta z)$. This obviously makes it difficult to design for ergonomics as many muscles and joins are implicated. As we are dealing with a handheld AR system we apportion some of our discussion towards ergonomics relating to the upper extremities.

Techniques

Prior work on handheld AR has focused on optimization of target acquisition. Several techniques exist. One technique is *Navicam* (Rekimoto, 1997) that uses a interaction metaphor of a magnifying glass allowing the user to inspect objects in their environment using a palmtop interface. Objects are registered using computer vision and color-coded ID tags. *TouchProjector* Boring et al. (2010) is another approach that allows users to manipulate on distant displays by touching and dragging objects in live video. The user position the device with the non-dominant hand and while interacting with the dominant hand. This type of bi-manual interaction has shown to be better than dependent techniques (one hand influences the other) (Kabbash et al. 1994; Guimbretière et al. 2005). However, as stated by Boring et al. (2010) it is not a good solution in practice, because of unstable images created by minor hand movements. These movements are often a result of fatigue Grubert et al. (2012) or can be caused by neurological disorders. According to Nazri & Rambli (2015) a shaky viewpoint can make it difficult to interact but do not further elaborate on why. We find one explanation in Bai et al. (2012):

> "[...] pressing the touch screen while pointing the device at an AR tracking image can result in camera motion that has a serious negative effect on the AR tracking accuracy, and overall user experience."

However, this is only partly true i.e. issues with tracking are minor due to methods such as Visual SLAM (V-SLAM) Cadena et al. (2016) and Visual-inertial SLAM (VI-SLAM) Leutenegger et al. (2013) which enables robust tracking for indoors or outdoors. Furthermore, smartphone cameras have become more advanced over the years and some are able to optically or digitally stabilize the live video in real-time; shaking should therefore be considered an ergonomic issue rather than a technical one.

There are numerous ways to approach the shaking issue. Some researchers solve it by implementing a freeze-frame technique (Abawi et al. 2004a; Guven et al. 2006a; Lee et al. 2009). There are multiple way to use this technique i.e. Abawi et al. (2004b) utilizes the technique in calibration of real and virtual objects for an authoring tool for mixed reality called *AMIRE-ES*. Guven et al. (2006b) uses freezing to support annotating situated media such as multimedia and hypermedia embedded within the physical environment. Lee et al. (2009) proposed a three step method called 'Freeze-Set-Go'. During the first step, the user can observe content on live video. While observing the user can freeze the current frame tentatively. Once frozen the user can still manipulate virtual entities on the frame but is no longer required to point the device. This eliminates the issue of shaky viewpoint, however, the real-time tracking is interrupted in the process. Lee et al. (2009) also cautions that users can get confused once resuming the live video as the viewpoint changed. *TouchProjector* deals with the former by updating the video frame with the a copy of the remote screen the user is currently interacting with.

Alternative strategies to freeze frame interaction also exist. In Liao et al. (2010) a method called 'loose-registration' stabilize shakiness by superimposing a 3D copy of the viewed content instead of the live video. As a result persistent and precise alignment is not necessary and issues such as blurring (due to shakiness) does not occur. Again this should not be an issue on modern hardware, but studies that compares shakiness using different state-of-the-art tracking methods and camera refresh rates are hard to come by.

Another thing to consider is different touch-based gestures for pointing and selection tasks. Two evaluations of different pointing techniques was performed by researchers in the paper "*Precise pointing techniques for handheld Augmented Reality*" Vincent et al. (2013a). The researchers used four different interaction methods (refer to figure 11) and compared them: (1) *Direct Touch* on the live video (1:1 mapping with the screen), (2) *Shift&Freeze*: Shift combined with freeze-frame (3) *Screen-centered Crosshair* and (4) *Relative Pointing* with cursor stabilized on the physical object Vincent et al. (2013b). The authors based the *Shift&Freeze* technique on Vogel & Baudisch (2007) which describes a method that reveals occluded screen content in a callout above the user's finger. *Relative Pointing* works similar to a trackpad where the user swipes on the device screen to move the cursor on the physical object.

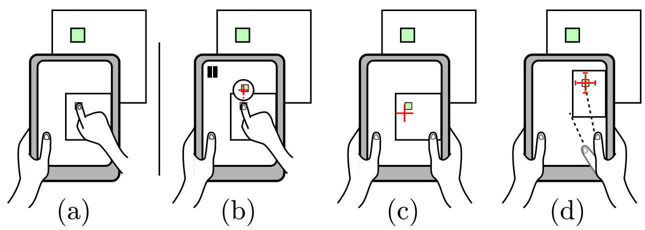

Figure 11: **(a)** Direct Touch (1:1 mapping with the screen), **(b)** Shift&Freeze **(c)** Screen-centered Crosshair **(d)** Relative Pointing. Reprinted from "Precise pointing techniques for handheld Augmented Reality" by Vincent et al. 2013b Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Pages 2441-2450.

The first evaluation sought to rate the overall user experience while the second evaluation rated performance. Results indicated that *Shift&Freeze* and *Relative Pointing* were preferred over *Direct Touch* and *Crosshair*. Another study by Olst (2012) also revealed that direct touch-based interfaces performed better in terms of speed of executing tasks as compared to crosshair interfaces.

Discussion

Several techniques have been proposed dealing with issues such as precision of target acquisition, and issues such as hand shakiness that may be brought on by bad ergonomics. In general research has shown that the *Crosshair* selection we use in the *AR Explorer* performs poorly in terms of target acquisition. However, one of the studies were conducted on tablets whereas *AR Explorer* uses a smartphone to facilitate interaction. Furthermore, non of the studies we examined have analyzed fatigue in detail i.e. providing statistics on optimal duration of use in scenarios requiring the user to hold hand poses for an extended duration. It is therefore unclear how big an issue ergonomics pose at this point in time. But solutions such as freeze-frame and loose-registration can be used to mitigate ergonomic issues if they indeed are problematic to the user experience.

Comparatively our method is similar to *NaviCam* in that the used metaphors are the same, the main difference being in our implementation i.e. the inclusion of ray-casting Mine (1995) for target acquisition. We also use 3D object recognition where *Navicam* uses image-processing to detect color-coded ID tags. *Touch-Projector* is fundamentally different from the crosshair selection that is used in our application, as it allows for bi-manual interaction. This is because we do not implement object manipulation. Users point the device on a point of interest and tap the screen which can be done using one hand. Finally, the 'loose registration' approach from Liao et al. (2010) is somewhat similar the *AR Explorer* since we overlay a digital copy of the object in place of the real (physical) object. This allows zooming in on the object without loss of quality.

Conclusion

Based on our evaluation we provide an answer to our problem statement *"How can we create a useful and ergonomic HAR application for museums"*. In view of our previous work and the gathered knowledge, we find that a feasible next step would be to reconsider the current interaction scheme (crosshair selection). But as there is currently no clear answer to the contribution to ergonomics of various interaction techniques, we must first examine ergonomics of different techniques in a separate study. Another aspect that should be considered in light of the recent advancements in technologies (hardware and software), is how they contribute to the user experience i.e. how do screen refresh rate affect cognition? Or how do registration issues (of a given tracking algorithm) affect the overall user experience? We also find that the human-centric approach can be used effectively in such studies to maintain focus on creating comfortable and effortless interaction design patterns. This could be achieved by keeping interaction methods within the capacities of cognitive, perceptive and physical ergonomics. Finally, the taxonomy of Mixed Reality has proven useful as a tool to classify techniques that are not directly related but, nonetheless, applicable for the *AR Explorer*. An extension of the taxonomy by means of multi-modality could be useful, especially as we are considering the possibilities of adding audio-visual cues to increase engagement of museum visitors.

References

Abawi, D. F., Dorner, R., Raller, M., & Zauner, J. (2004a, March). CVMP 2004 - efficient mixed reality application development. In *1st european conference on visual media production (cvmp) 2004* (p. 289-294).

Abawi, D. F., Dorner, R., Raller, M., & Zauner, J. (2004b, March). CVMP 2004 - efficient mixed reality application development. In *1st european conference on visual media production (cvmp) 2004* (p. 289-294).

Ariso, J. M. (2017). *Augmented Reality, Reflections on Its Contribution to Knowledge Formation.* Berlin, Boston: De Gruyter. (DOI: 10.1515/9783110497656)

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, *21*(6), 34–47.

Bai, H., Lee, G. A., & Billinghurst, M. (2012). Freeze view touch and finger gesture based interaction methods for handheld augmented reality interfaces. In *Proceedings of the 27th conference on image and vision computing new zealand - IVCNZ '12.* ACM Press.

Boring, S., Baur, D., Butz, A., Gustafson, S., & Baudisch, P. (2010). Touch projector. In *Proceedings of the 28th international conference on human factors in computing systems - CHI '10.* ACM Press.

Boring, S., Baur, D., Butz, A., Gustafson, S., & Baudisch, P. (2010). Touch projector. In *Proceedings of the 28th international conference on human factors in computing systems - CHI '10.* ACM Press.

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., et al. (2016, dec). Past Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Transactions on Robotics*, *32*(6), 1309–1332.

Carroll, J. M. (2003). Introduction: Toward a Multidisciplinary Science of Human-Computer Interaction. In *HCI models theories, and frameworks* (pp. 1–9). Elsevier.

Covington, M. A., & Slaughter, M. M. (1983, dec). Universal Languages and Scientific Taxonomy in the Seventeenth Century. *Language*, *59*(4), 920.

Douglas, S. A., & Mithal, A. K. (1997). *The Ergonomics of Computer Pointing Devices.* Springer London.

Dvir, D., Raz, T., & Shenhar, A. J. (2003, feb). An empirical analysis of the relationship between project planning and project success. *International Journal of Project Management*, *21*(2), 89–95.

Earthy, J., Jones, B. S., & Bevan, N. (2012a). ISO Standards for User-Centered Design and the Specification of Usability. In *Usability in government systems* (pp. 267–283). Elsevier.

Earthy, J., Jones, B. S., & Bevan, N. (2012b). ISO Standards for User-Centered Design and the Specification of Usability. In *Usability in government systems* (pp. 267–283). Elsevier.

Fernandez, M. J., & Eastman, C. M. (1990, oct). Basic Taxonomic Structures and Levels of Abstraction. *Advances in Classification Research Online*, *1* (1), 59–70.

Grubert, J., Grasset, R., & Reitmayr, G. (2012). Exploring the design of hybrid interfaces for augmented posters in public spaces. In *Proceedings of the 7th nordic conference on human-computer interaction making sense through design - NordiCHI '12.* ACM Press.

Guiard, Y. (1987, dec). Asymmetric Division of Labor in Human Skilled Bimanual Action. *Journal of Motor Behavior*, *19* (4), 486–517.

Guimbretière, F., Martin, A., & Winograd, T. (2005, sep). Benefits of merging command selection and direct manipulation. *ACM Transactions on Computer-Human Interaction*, *12* (3), 460–476.

Guven, S., Feiner, S., & Oda, O. (2006a, oct). Mobile augmented reality interaction techniques for authoring situated media on-site. In *2006 IEEE/ACM international symposium on mixed and augmented reality.* IEEE.

Guven, S., Feiner, S., & Oda, O. (2006b, oct). Mobile augmented reality interaction techniques for authoring situated media on-site. In *2006 IEEE/ACM international symposium on mixed and augmented reality.* IEEE.

Hargens, L. L., & Schuman, H. (1990, sep). Citation counts and social comparisons: Scientists' use and evaluation of citation index data. *Social Science Research*, *19* (3), 205–221.

Kabbash, P., Buxton, W., & Sellen, A. (1994). Two-handed input in a compound task. In *Conference companion on human factors in computing systems - CHI '94.* ACM Press.

Kotler, N., & Kotler, P. (2000, jan). Can Museums be All Things to All People?: Missions Goals, and Marketing's Role. *Museum Management and Curatorship*, *18* (3), 271–287.

LaViola., J. J., Kruijff, E., McMahan, R. P., Bowman, D., & Poupyrev, I. P. (2017). *3D User Interfaces: Theory and Practice (2nd Edition)* . Addison-Wesley Professional.

LaViola., J. J., Kruijff, E., McMahan, R. P., Bowman, D., & Poupyrev, I. P. (2017). *3D User Interfaces: Theory and Practice (2nd Edition) (Usability).* Addison-Wesley Professional.

Lee, G. A., Yang, U., Kim, Y., Jo, D., Kim, K.-H., Kim, J. H., et al. (2009). Freeze-Set-Go interaction method for handheld mobile augmented reality environments. In *Proceedings of the 16th ACM symposium on virtual reality software and technology - VRST '09.* ACM Press.

Lee, G. A., Yang, U., Kim, Y., Jo, D., Kim, K.-H., Kim, J. H., et al. (2009). Freeze-Set-Go interaction method for handheld mobile augmented reality environments. In *Proceedings of the 16th ACM symposium on virtual reality software and technology - VRST '09.* ACM Press.

Leutenegger, S., Furgale, P., Rabaud, V., Chli, M., Konolige, K., & Siegwart, R. (2013, jun). Keyframe-Based Visual-Inertial SLAM using Nonlinear Optimization. In *Robotics: Science and systems IX.* Robotics: Science and Systems Foundation.

Liao, C., Liu, Q., Liew, B., & Wilcox, L. (2010). Pacer : Finegrained interactive paper via camera-touch hybrid gestures on a cell phone. *Proceedings of CHI 2010ACM*, 2441-2450.

MacKenzie, I. S. (2013). *Human-Computer Interaction: An Empirical Research Perspective* (1st ed.). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

MacKenzie, I. S., & Guiard, Y. (2001). The two-handed desktop interface. In *CHI '01 extended abstracts on human factors in computing systems - CHI '01.* ACM Press.

Mann, S. (2002, aug 6). *Mediated Reality with implementations for everyday life.* http://wearcam.org/presence_connect/.

Milgram, P., & Kishino, F. (1994, dec). A Taxonomy of Mixed Reality Visual Displays. *IEICE Trans. Information Systems*, *vol. E77-D, no. 12*, 1321–1329.

Mine, M. R. (1995). *Virtual Environment Interaction Techniques* (Tech. Rep.). Chapel Hill, NC, USA.

Mogensen, M., & Stanulionyte, R. (2017a, dec). *How gamification motivates: an experimental study of the museum experience using handheld augmented reality.* (Aalborg University Copenhagen)

Mogensen, M., & Stanulionyte, R. (2017b, dec). *How gamification motivates: an experimental study of the museum experience using handheld augmented reality.* (Aalborg University Copenhagen)

Müller, T. (2015). Towards a Framework for Information Presentation in Augmented Reality for the Support of Procedural Tasks. In *Lecture notes in computer science* (pp. 490–497). Springer International Publishing.

Nazri, N. I. A. M., & Rambli, D. R. A. (2015). The roles of input and output modalities on user interaction in mobile augmented reality application. In *Proceedings of the asia pacific HCI and UX design symposium on - APCHIUX '15.* ACM Press.

Neuburger, L., & Egger, R. (2017, sep). Augmented Reality: Providing a Different Dimension for Museum Visitors. In *Augmented reality and virtual reality.* Springer International Publishing.

Olst, R. R. v. (2012, sep). *3D Manipulations in Handheld Augmented Reality Applications.*

The Oxford English Dictionary Second Edition (VOLUME 11). (2001). Clarendon Press.

Rekimoto, J. (1997, aug). NaviCam:A Magnifying Glass Approach to Augmented Reality. *Presence: Teleoperators and Virtual Environments*, *6*(4), 399–412.

Rosa, N., Werkhoven, P., & Hürst, W. (2016). (Re-)examination of multimodal augmented reality. In *Proceedings of the 2016 workshop on multimodal virtual and augmented reality - MVAR '16.* ACM Press.

Stewart, T. F. M. (1991). *Usability Now!: Directory of HCI standards.* Department of Trade and Industry.

The International Ergonomics Association. (2017). *Definition and Domains of Ergonomics.* http://www.iea.cc/whats/index.html. ((Accessed on 01/10/2018))

Tijssen, R. J. W., Visser, M. S., & Leeuwen, T. N. v. (2002, jul). Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? *Scientometrics*, *54*(3), 381–397.

Turnage, A. K. (2010). *Identification and disidentification in organizational discourse: A metaphor analysis of e-mail communication at Enron.* North Carolina State University.

Usability.gov. (2015, apr). *User Centered Design Basics.* U.S. Department of Health & Human Services.

Vincent, T., Nigay, L., & Kurata, T. (2013a). Precise Pointing Techniques for Handheld Augmented Reality. In *Human-computer interaction – INTERACT 2013* (pp. 122–139). Springer Berlin Heidelberg.

Vincent, T., Nigay, L., & Kurata, T. (2013b). Precise Pointing Techniques for Handheld Augmented Reality. In *Human-computer interaction – INTERACT 2013* (pp. 122–139). Springer Berlin Heidelberg.

Vogel, D., & Baudisch, P. (2007). Shift: A Technique for Operating Pen-Based Interfaces Using Touch. In *Proceedings of the SIGCHI conference on human factors in computing systems - CHI '07.* ACM Press.

Wickens, C. D., & Carswell, C. M. (2006, feb). Information Processing. In *Handbook of human factors and ergonomics* (pp. 111–149). John Wiley & Sons Inc.

Xia, X., Irani, P., & Wang, J. (2007). Evaluation of Guiard's Theory of Bimanual Control for Navigation and Selection. In *Ergonomics and health aspects of work with computers* (pp. 368–377). Springer Berlin Heidelberg.