MASTER 2 in COMPUTER SCIENCE

HUMAN-COMPUTER INTERACTION PROGRAM

Enhancing the Web Browsing Experience on Mobile Devices with AR Glasses

Author:
Yang CHEN (Sophie)

Supervisor:
Emmanuel PIETRIGA
Caroline APPERT

Hosting lab/enterprise:
ILDA Team, Inria (Paris Saclay – Île-de-France)

1st Mar – 31st Aug 2022

Secrétariat - Tel: 01 69 15 66 36 - Fax: 01 69 15 42 72
Email: Murielle.Benard@u-psud.fr

Master Thesis

Contents

Summary

As Augmented Reality(AR) glasses become increasingly powerful, comfortable and affordable, we foresee a future where they will become common place and will be used in tandem with a mobile phone. Therefore, there is a strong motivation to investigate how an integrated system of AR glasses and smartphone could be used together to improve user experience. Such a system has a lot of potentials as their strengths and weaknesses complement each other. In this project, we investigate how AR display can be used to enhance user experience for web browsing on mobile devices. Through our exploration, we identify four main categories of use cases and three design factors. We implement an interactive prototype that allows users to offload different types of content from web pages to AR display. We explore different direct and mid-air interaction techniques that can offload content types effectively. Lastly, we present findings and insights from a 10-participant user study that provides empirical data supporting the practicality of our system, and discuss challenges and guidance for the future. In our study, the participants found the integrated system compelling and greatly enhanced the efficiency and comfort of certain web browsing tasks. More participants preferred the automatic offload of content to the manual mode and rated it higher in terms of comfort, ease of use and efficiency.

Keywords

Augmented Reality, Cross-Device Interactions, Mixed Reality Computing, Web Browsing

CHAPTER 1

Introduction

Augmented reality(AR) is a technology that injects virtual objects and elements into the physicality of the real world. It enhances the real physical world by overlaying additional information based on the user's environment and use context. Most of the time, the information is seamlessly interwoven with the physical world in such a way that it can be perceived as an immersive aspect of the real physical world. The virtual augmented elements are spatially registered in three dimensional (3D) space and are interactive in real time [2]. This allows users of AR systems to simultaneously interact with their physical and virtual environments.

Over the past few years, AR applications have been rolled out to millions of users for a multitude of use cases. For example, maps with an AR view are used to provide navigation guidance both in indoor and outdoor environments by augmenting the user's field of view with routing information. Applications such as Google Translate have an AR feature that automatically overlays translation of foreign words onto menus and instructions sheets. Popular mobile game Pokemon Go places interactive virtual characters in real world locations to encourage exploration of the physical environment.

Currently, the most common method of rendering AR experiences is through handheld devices. Most modern handheld devices like smartphones and tablets contain hardware components like camera, sensors and geolocation features as well as software capabilities like computer vision and object recognition that make them capable of supporting AR experiences. With handheld AR systems, users hold the devices in front of their eyes like a compact camera to get the augmented experience. However, holding the device for long periods would lead to fatigue (the "Gorilla arm effect"), making it unsuitable for continuous usage [10]. Hence, much effort has been focused in the development of Augmented Reality Head-Mounted Displays (ARHMD) that can be worn on the user's head, leaving the user's hands free to perform other tasks.

Currently, high performance ARHMDs are mostly developed and used in the enterprise space, in industries ranging from manufacturing to medical. They are often very costly - for example, a pair of Microsoft's Hololens 2, the ARHMD used in our project, cost around 4000 euro. Because of their rather limited battery life, they are often worn for short periods of time to assist with specialized tasks such as maintenance, assembly and training in enterprise settings.These and other factors such as limited field of view, bulky or uncomfortable form factor and poor contrast in certain lighting conditions make them unsuitable for continuous consumer use in daily life.

In recent years, there has been much effort in developing lower-cost ARHMD in the form factor of spectacles and even contact lenses. All the major tech companies including Google, Apple, Meta and Snapchat are working on developing consumer versions of AR glasses. As consumer AR glasses become increasingly affordable, compact and powerful in the near future, we foresee that more and more people will own a set of AR glasses and use it in tandem with a mobile device, just as smart watches are often used with a mobile phone today. AR experience is likely to transition from short term, single purpose usage to a continuous, pervasive experience. Grubert et al. term this continuous AR experience Pervasive Augmented Reality, in which an omnipresent and universal AR system provides continuous and pervasive digital information that is adaptive to the current changing requirements and constraints of the user's context [10]. There is therefore a strong motivation for investigating how an integrated system of AR display and mobile devices could work together to improve different aspects of people's lives in the near future.

In this project, we investigate how an AR display can be used to enhance user interaction for web browsing experience on mobile devices. Web browsing is one of the most common use cases for mobile usage across devices and demographics. However, despite being a common use case, the user experience for web browsing on mobile devices is not ideal due to limitations of the small display area. Interaction with web pages on small screens can be crude and tedious as the screens can typically only display a small portion of the entire page at a time. AR displays can help improve the user experience by displaying peripheral information in the air around the phone. The additional display space could be used to show an outline of the web page, preview content at the end of a hyperlink, or offload content from the page which can always remain visible regardless of scrolling.

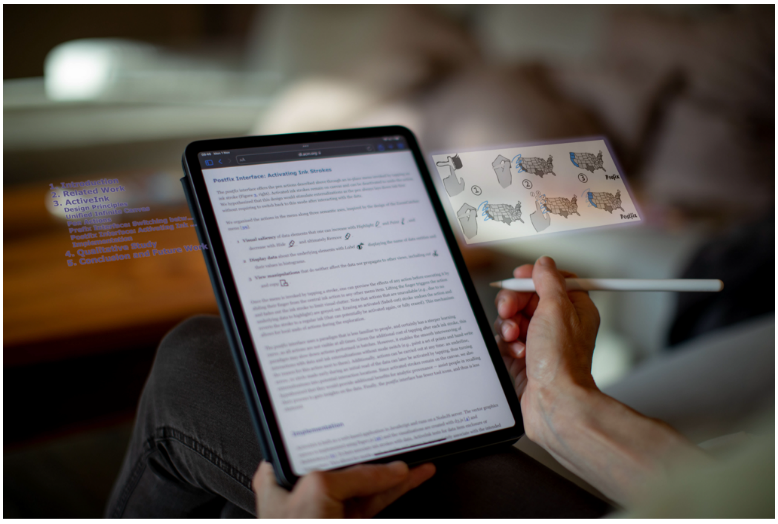


Figure 1.1: Concept art of how an AR display can be used to enhance web browsing

Our main contributions in this project are: 1) We explore use cases of how an integrated system of smartphone and AR display can be used to enhance web browsing and identify design factors to consider. 2) We implemented an interactive prototype that allows users to offload different types of content from web pages to AR display and explored novel interaction techniques. 3) We present findings and insights from a 10-person user study that provides some empirical data on the performance of using AR display to enhance web browsing activities and identifies challenges associated with it.

CHAPTER 2

Related Work

There has been extensive research done in the area of cross-device computing and interactions in the past few decades. In their recent paper, Brudy et al. presented an analysis and a taxonomy of a corpus of 510 papers in the cross-device computing domain [4]. In these multi-device environments, users can select the best device for a given task based on its strengths, but should also be able to use multiple devices in tandem as an integrated and seamless interactive system. Our related work focuses on recent projects which combine mobile devices with AR displays in order to leverage the strengths of both devices in order to enhance user interaction.

From these considerations, it can be seen that the properties of mobile devices and AR displays complement each other, as one device's strength is the other's weakness[Figure 2.1]. This provides additional motivation to design an interaction experience that combines the two which utilizes these contrasting capabilities.

2.1 Mobile Device and AR Display

Mobile devices provide a physical surface for the finger to interact on, enabling precise and familiar input, tactile feedback and reliable detection of touch events. Mobile devices have higher resolution displays, but are limited in screen size and screen real-estate. In addition, there is also a higher access cost in terms of time and effort taken to retrieve the device from where it's stored, such as a bag or pocket.

On the other hand, AR displays offer stereo displays that can render digital information in the user's spatial environment, although they have limited fields of view. The head mounted displays are worn by the users and can provide easier access to always-on digital information. ARHMDs often contain sensors for head, hand and gaze tracking, offering opportunities for

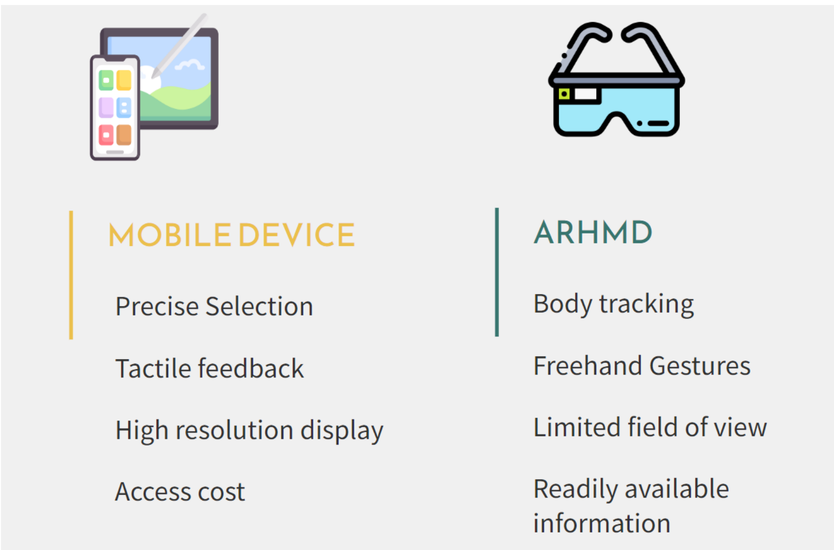


Figure 2.1: The properties of AR and smartphone complement each other

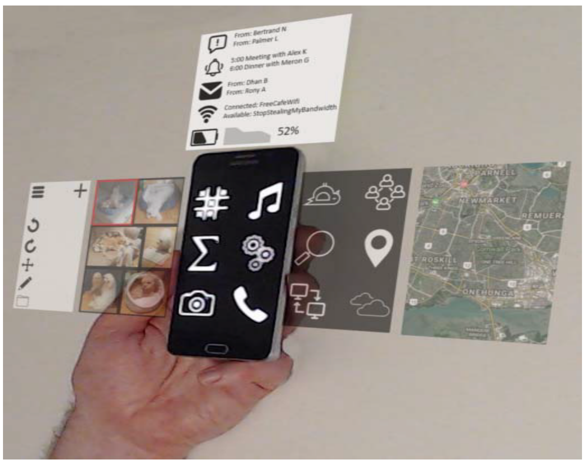
more intuitive and custom gestures. However, freehand gestures such as the “Bloom” gesture on the Hololens may be less familiar to users and can be prone to recognition errors. Mid-air interactions such as pointing in the air to provide input can also result in tremor and arm fatigue, a lack of tactile feedback, and a lower Fitts index of performance due to the lack of a physical surface to stabilize the finger [16].

Zhu and Grossman propose a broad categorization of work combining mobile devices with ARHMDs as being either mobile-centric, where HMDs are used to enhance mobile interactions, or HMD-centric, where mobile devices are used to enhance spatial interaction [17].

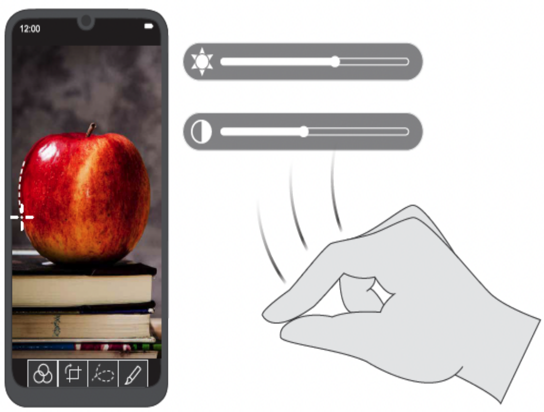
2.2 Mobile Centric Interactions

In mobile centric projects, AR display is often used to overcome the limited screen space available on mobile devices such as a phone, smart watch or tablets. Normand and McGuffin explored ways of extending the display of the phone by displaying 2D content on Virtually Extended Screen-Aligned Display(VESAD)[Figure 2.2a] around the phone that are coplanar and aligned with the phone [13]. They proposed guidelines for future VESAD design from their user studies, including avoiding placing input targets on VESAD far from the phone, and avoiding interactions or layouts that required users to cross their arms. These guidelines are taken into consideration in the design of our work.

Extending interactive surfaces to the air can be done by adding in either input, output capabilities or both. An example is MultiFi [9], which adapts user input and output capabilities based on the respective fidelity of the devices involved. For example, the high resolution screen of a smartphone can be used as the primary display while secondary information is displayed in AR. The phone’s touchscreen can also be used entirely for input as a full screen keyboard while text is displayed in AR. Similarly, in AR-enhanced widgets for Smartphone-centric Interaction, Brasier et al. (2021) explored a UI distribution strategy that offloaded widgets in the air around



(a) mock up of Virtually Extended Screen-Aligned Displays which helps to extend screen space (VESAD, 2018)



(b) Offloading UI widgets in the air can help to free up limited screen estate (Brasier et al., 2021)

Figure 2.2: Examples of Mobile Centric work

the phone[Figure 2.2b], and distributed input between direct touch and mid-air gestures [3]. Both distribution strategies help to free up limited screen real-estate and reduce occlusions, as users interact partly around the device.

In MultiFi, Grubert et al. (2015) distinguish between three possible alignment modes: body-aligned mode, where information space is spatially registered to the user's body, side-by-side mode, where the ARHMD and handheld devices are not spatially coordinated, and device-aligned mode, where the information space is spatially registered to the mobile device and moves with it [9]. In this mode, the AR display shows additional, peripheral information to the user at lower fidelity, thus extending the screen space of the touchscreen, yielding a focus plus context display. The designs we present in this project adhere mostly to the device-aligned mode.

2.3 AR Centric Interactions

AR centric research typically explores ways handheld devices can be used as input for interacting with content in virtual environments. Budhiraja et al. (2013) considered two ways a handheld device could be used for gesture input - by making gestures on its touch screen and by the movement of the phone itself [7]. In the first method, the user performs normal gestures such as pointing and moving on the handheld device and the movement is reflected on the AR display via a corresponding virtual cursor. In their Motion Pointing prototype, the user holds the device in one hand and moves the virtual cursor by tilting the device. The direction and speed of the cursor movement is controlled by the tilted angle and users can select an object under the cursor by touching the screen with their thumb. Similarly, in TrackCap [12], the mobile phone is attached to a plastic toy gun and used as a precise 6 Degree of Freedom input device for the head mounted display[Figure 2.3].

There has also been research on using the mobile device as a secondary output display to the AR display. Budhiraja et al. (2013) propose a multi-layered visualization where the handheld device can be used to provide a different representation of the 3D scene in AR, for example x-ray vision [6]. In this scenario, the user holds the mobile device in front of them, and views it through the AR display. The handheld device can show different versions of the objects in the AR scene, such as the interior of a building, or an older version of an object.

Figure 2.3: TrackCap(2019) uses the smartphone as a precise 6DOF input device for ARHMD

2.4 Web Browsing on Mobile Devices

There have been several studies done on the limitation of web browsing on mobile devices with small screen size. Research found that performing tasks on mobile devices with small screen size places heavy cognitive demands on the user's short term memory - the user will have difficulty initially acquiring the information before activating a mental model for interpreting the information and will have difficulty placing information within the existing mental model as they progress through the information [1]. This is likely due to the way webpages are displayed on a small screen compared to a desktop. Due to the narrow width of mobile devices, web pages on a mobile handheld are commonly displayed as one tall page and require extensive scrolling. On the desktop, users can usually get a general idea of the site content at a glance. There are usually explicit labels on web pages that make navigation through the site easy, including headers, breadcrumbs or menus. On the mobile device, these navigation cues are less accessible and usually hidden behind a collapsed menu in order to conserve space.

As such, Roto & Kaikkonen (2003) found that users had difficulties in getting the whole picture of the site and had to concentrate more in browsing the information in order to keep track of the current location on the page [14]. When they find content of interest in some particular spots on the webpage, they often have difficulties coming back to it later. Consequently, users of small screens were 50% less effective in completing retrieval tasks than users of larger screens [11]. Small screen users also made more mistakes while navigating the web pages and were less willing to browse deeply into the material [5].

CHAPTER 3

Design

Based on the previous work and the issues we identified from related literature, we brainstormed scenarios of how an integrated AR and handheld mobile system can be used to enhance users' web browsing activities. We term the integrated interactive system BrowseAR, and present four broad categories where BrowseAR can facilitate web browsing: navigation, offloading content, previewing and styling content.

3.1 Use Cases

Navigation

On mobile devices, content is usually displayed on long narrow web pages which require extensive scrolling from the user. In order to motivate users to increase their consumption time, many content providers even implement a type of endless scrolling, where new content is displayed whenever the user reaches the end of the page, or whenever a new update has been made such as in the case of social media applications like Twitter. As such, users often find it difficult to keep track of their reading location on the page or where a particular interesting piece of content is. It is also common for users to jump from one website to the other by clicking on the relevant links on the page, going down a 'rabbit hole' of links and forgetting how they ended up there.

BrowseAR can help users navigate both within pages and between pages. On websites displayed through desktop browsers, there are often navigation cues such as breadcrumbs and menus that aid the user in navigating the page. On mobile devices, these navigation cues are often hidden behind dotted menus to conserve the limited screen space and brought up only when needed. The system provides an opportunity for navigation aids to always be visible and accessible.

Navigating Within Pages

On websites with long content that requires extensive or endless scrolling, vertical markers can be used to mark the location of interesting content or a reading position that the user can return to later. A content outline can be automatically displayed alongside the page. This outline can provide users with an overview of the page at a glance, and allow them to easily navigate to specific sections of the page.

Web pages are structured documents written in hypertext mark-up language HTML, each HTML page gets parsed as a Document Object Model (DOM) with semantic tags that provides information about the structure, content and organization of the web page. This feature makes it possible to traverse through the DOM tree and select specific HTML elements to be offloaded in AR.

Navigating Between web pages

Breadcrumbs could be shown in AR to visualize a trail of links the users visited in the past to help users get to the current page, and to facilitate moving to specific pages of relevant content.

Users can create bookmarks of frequently visited webpages to access them easily and to save time typing them out each time.

Offloading Content

Offloading content is the use case that has been explored the most in previous literature. It provides two obvious benefits: displaying more information while freeing up the limited screen real-estate on mobile devices, and making relevant content persistent so that users can refer to it more easily.

BrowseAR can help to distribute input and output, with the wider field of view of ARHMD used for displaying content while the mobile device is used for inputting content [Figure 3.1] . For example, users can offload a web view of an email in AR so that they can refer to it easily while using the phone to type a reply. Similarly, users can watch a video and view a live stream of comments in AR, while using the phone as an input device to type real time comments.

In addition, the offloading of web-content can be bidirectional and cross-application. Users can use the system as a clipboard to transfer interesting content from different applications. For example, they can offload an interesting picture they found on a webpage to AR, and then later send that picture in an email, or share it on their social media clients. This is similar to the work presented in Gluey [15], which enables conteent to be transferred seamlessly between devices.

Previewing Content

BrowseAR help users preview content which allow them to access interested content easier and faster whilst reducing unnecessary and excessive scrolling. It can help users preview both internal content from the same page, or external content from another page [Figure 3.2].

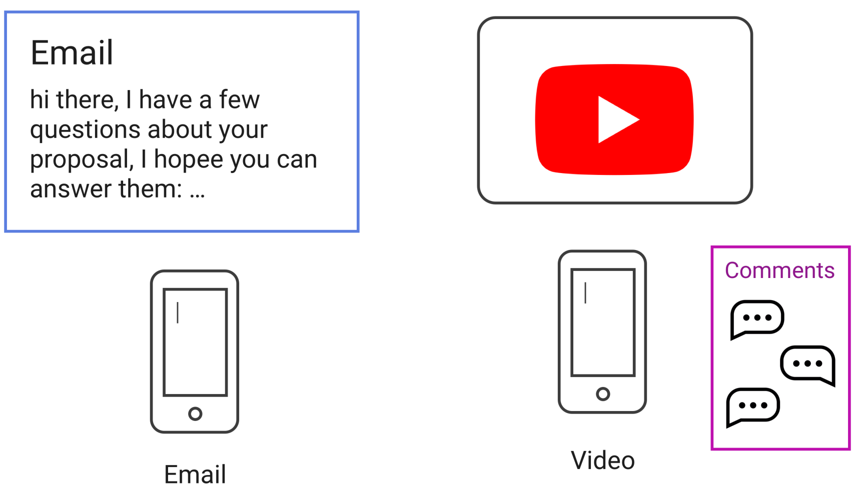


Figure 3.1: Different types of web content can be offloaded – User can offload an email in AR display while typing the reply on the phone (left). User can watch video and view a live comment stream in AR while typing comments on the phone (right)

Internal

When the user is reading an article or novel on their mobile device, the footnote or glossary of the article can be shown in AR alongside the primary content on the phone. This allows users to get a better understanding of the term easily without needing to scroll to the bottom of the page and disrupting their reading experience. Similarly, the previewing option can help researchers or students to look up bibliography references in scientific articles easily without needing to go back and forth.

External

The system can also help users preview content from external sites without leaving the current site. This is akin to using the preview feature on file explorer to get a quick glimpse of the files without having to open them up individually. Similarly, a user can get a preview of a website in a hyperlink to get an idea of the content referenced on the link without leaving the page. This could be useful also in the case of searching for information via a search engine. When the user clicks on an item in the search result list of hits, it could be very useful if the relevant section containing the search term is displayed in AR so the user does not have to click and search through the sites one by one. Another useful scenario would be on social media sites, where users can preview profile information of a Tweeter or LinkedIn account while going through their social media feed.

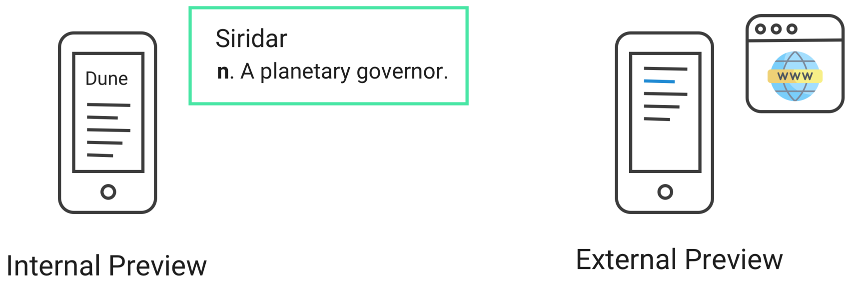


Figure 3.2: User can view glossary of a term next to the page (left). A preview of an external site can be displayed in AR to give users a general idea of the site (right)

Styling

Users can offload content in AR with the option of styling the content to their liking. This could be useful for accessibility purposes where users can take advantage of the larger display area of AR to show text in larger font sizes. Furthermore, such a system provides customization, allowing users to modify the Cascading Style Sheets(CSS) of web content so that they can view web content in dark mode or any color schemes of their preference.

3.2 Design Factors

We identify three factors in the design of an integrated mobile and AR Display system that can be taken into consideration: 1. Whether content is to be offloaded automatically or manually, 2.

How content is displayed and 3. The viewport of the content. We discuss each of the design factors in more details below.

Automatic vs Manual

In the previous section on use cases, we discussed how offloading content into AR can facilitate information gathering and retrieval tasks by letting users access information easily without having to scroll or navigate to another site. The offloading can be done automatically from the side of the web developer or content provider or manually from the side of the end user.

Automatic

We imagine there could be a specific AR tag in HTML for developers or content providers to specify whether an element should be offloaded in AR. This could be implemented in the form of a specific CSS class, or a new tag such as "ar-enabled" that could be added to the HTML. Content with the AR tag specified will appear automatically to the user in AR when the page loads or wherever their corresponding link text is visible in the viewport of the mobile device. For instance, the content of the footnote or a term in the glossary can automatically appear alongside an article as the user scrolls down the page.

Manual

On the other side, the end users can be given the option to manually offload a relevant piece of content whenever they desire. For example, when the user long presses on a link, an option will appear on the contextual menu that allows him to offload it. Alternatively, he can swipe on a picture on a website to offload it in AR in the direction of his choosing.

Content Display

There are several strategies to display content in AR [Figure 3.3]:

Free form

Content is displayed freely in the users' spatial environment. This provides the user with the flexibility and freedom to place content around him at a position and angle that is optimal for his viewing.

Fixed position

Content is displayed in fixed content frames around the phone. Whenever a new piece of content is offloaded, it replaces the existing content. Because of AR Display's limited field of view, this display strategy can help to ensure that the displayed content is more or less always visible.

Grid layout

Content can be displayed in structured collections around the phone. This allows some degree of freedom in offloading the content whilst keeping the content structured and organized so that it doesn't overwhelm the user.

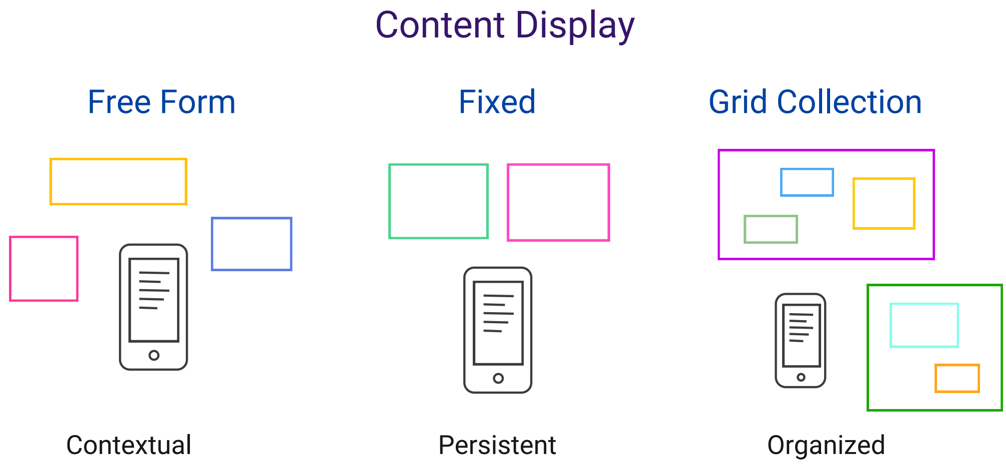


Figure 3.3: Different types of content display mode.

Content Viewport

Content viewport relates to the context in which offloaded content is viewed in AR

World – Content is placed at fixed 3D positions that are independent to the user's physical environment. This allows easy access to the content at a position that is convenient to access and comfortable to view.

Spatial – The content is displayed in the context of the user's environment - content can be pinned to walls and surfaces of the physical world. Therefore, it requires spatial awareness from the ARHMD to scan and track the user's physical environment. This can be useful in situations when a user wants to offload information that is relevant to a physical location, such as placing a recipe on the fridge door, or leaving some information around his working area that is useful and relevant to his work.

Gaze - Content follows the user's position and is placed in a position relative to the user's gaze. It can be displayed in a pre-defined vertical position so that it is only visible when the user glances up or down. This strategy ensures that the content remains easily accessible, but only visible when required.

Handheld Device - Content follows the position of the mobile device and is displayed around the phone. This allows for parallel viewing, users can easily relate the content offloaded in AR to the content on the screen of the mobile device.

CHAPTER 4

Implementation

We developed an interactive prototype that allows users to offload web content from a mobile device to AR display [Figure 4.1]. The hardware of our prototype consists of a Microsoft Hololens 2 and an Android smartphone. The web application used in our prototype is built using HTML, CSS, Javascript, the web application is served by NodeJS and deployed on Heroku. Users are able to access the web application on the mobile device through any web browser on the device. The Hololens runs a Unity program written in C# which is used to display web content in AR as well as to track hand gestures for mid-air input. Socket IO, an event driven web socket library, is used to achieve the two way communication between Hololens and the web client.

When the user selects a piece of content to offload on the mobile device by swiping on a picture for example, the web client sends the content to be offloaded to the web server. The web server then broadcasts the request to the web client and Hololens. When the Hololens receives the request and the web content from the server, the program formats the content and displays it in AR.

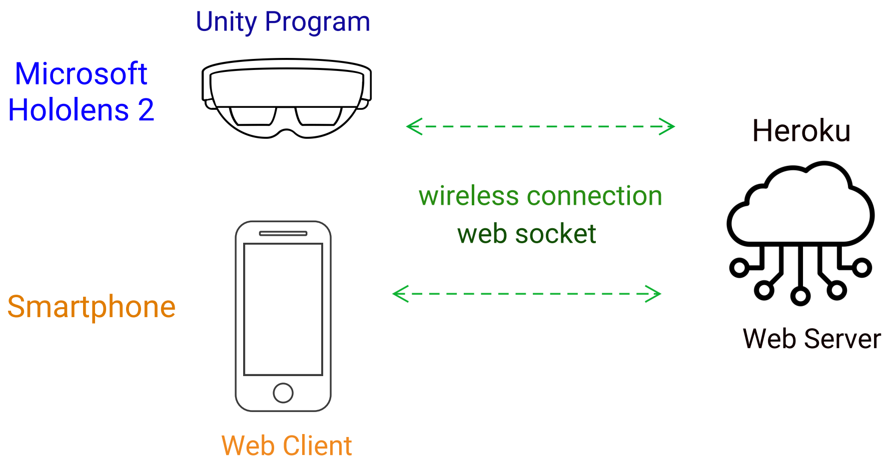


Figure 4.1: Infrastructure diagram of the prototype

4.1 Content Types

Our prototype supports the offloading of different web content types both within the page itself or from an external web site.

Internal Content

Text - An HTML web document is typically structured into content blocks by semantic tags (e.g. <p><section><img>). The user can choose to either offload an entire block of text, such as an entire paragraph, or a selection of text. When the user swipes on a paragraph or a selection, the touch action of the user is tracked and the element beneath the finger tip is

(a) The system can offload selected texts and images in AR

(b) 3D Webview (top) renders a preview of external links

Figure 4.2: Features of our prototype

selected. The text of the selected element is sent to the Unity program on Hololens which renders the text onto a content card and displays it in AR.

Picture - Users can offload pictures to AR by swiping on them [Figure 4.2a]. When a swipe is detected on the image, the web server sends the URL of the image to the Unity program running on Hololens. In the Unity program, the image is downloaded from the public URL and rendered as a material on a game object. The swipe action can be tracked either by the touch events (touchstart, touchmove and touchend) in javascript, or by Hammer.JS [1], a multi-touch gesture library.

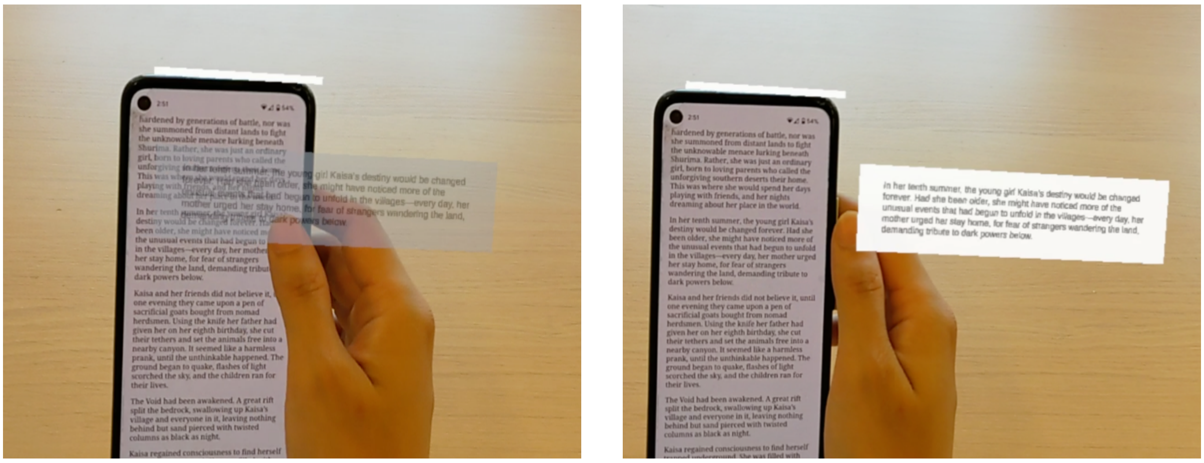
Internal Link - Users can preview and offload content referenced by anchor links within the page such as footnotes or references at the bottom of the web page by long pressing on the corresponding link text. When the user clicks on such a link, the web client obtains the ID of the anchor and finds the reference element by the ID. It then sends the text within the reference element to the web server which broadcasts it to Unity. The Unity program on the Hololens then renders the text onto a content card and displays it in AR.

External Content

External Link - Users can also get a preview of any external website that is linked on the web page. When the user long presses on an external link, the URL is sent to Hololens and opened up in a web browser within the Unity application. This makes use of 3D WebView, a cross platform web browser tool for Unity [Figure 4.2b] developed by Vuplex[2].

Since 3D WebView is essentially a web browser, it allows any type of web content to be displayed in AR, including but not restricted to text, images, formatted web content with links, and even videos. The size of the WebView window and the resolution of the content displayed can also be adjusted according to need.

[1] https://hammerjs.github.io/
[2] https://developer.vuplex.com/webview/overview



(a) As the user swipes directly on the phone, a preview of the selected content is generated.

(b) The user then continues swiping to offload the content to the side of the device

Figure 4.3: Direct swiping technique

The caveat of the tool is that it is quite computationally expensive and causes the Unity application to run slowly when there are multiple instances of WebView initialized. There are other similar Unity tools available that display webpages, such as Embedded Browser [3] and unity_browser [4]. However, they do not offer Hololens support. For this project, a trial version of 3D WebView was used to validate this as a viable option to render web content within AR.

Other than using the 3D WebView tool, another way of generating a preview of an external page is through taking a screen capture of the external page using javascript and showing that image in AR. A web example of this link preview is available at https://codepen.io/shaneapen/pen/bdrGRO. Due to the complexity involved in the steps of taking the screencap of an external link, storing the picture on the web server and then sending the picture URL to Hololens, we did not implement it in our system due to time constraints. However, it is a free alternative to 3D WebView that has less impact on the application performance since it is just rendering images versus instantiating a full-featured web browser.

4.2 Interaction Techniques

To investigate effective interactions between web content and AR display, we designed and implemented five interaction techniques that allow users to offload content blocks from the mobile device into AR - two on the device directly, two involving gestures in mid-air and a combination of both. The techniques are complementary and enable different types of offload operations, allowing the user to offload an entire HTML block or specific selection of text.

Direct Touch Techniques

The direct touch techniques involve direct interaction with the surface of the mobile device. They allow precise selection while providing tactile and haptic feedback to the user.

[3] https://assetstore.unity.com/packages/tools/gui/embedded-browser-55459
[4] https://github.com/tunerok/unity_browser

Swiping

Users can swipe an entire HTML block on the mobile device, such as a paragraph <p>or an image <img>block, to offload it in AR. For paragraphs, the text within the selected content block is sent to Hololens. In the case of images, the public URL of the image is sent to Hololens and is rendered onto a game object as a material.

Since vertical swiping interferes with the scrolling of the page, the swiping actions are restricted to horizontal swipes to the left and right of the device. There are event listeners on the web page that track the start and movement of the user's touch actions on the mobile device. When the horizontal movement of the finger from the touch start point passes a preview threshold of 60, a preview of the content block will appear in AR next to the user's finger [Figure 4.3a]. The preview block is semi-transparent and follows the user's finger as it continues swiping in the same direction. Once the horizontal movement exceeds the selection threshold of 170, the content is selected. The selected content block becomes opaque and will be fixed in a position that is proportional to the position of the link text on the device's viewport [Figure 4.3b].

Long Press

Users can offload content from internal and external links by long pressing on the link text. The user will receive haptic feedback from the phone when a long press is made. On Android devices, there is a known issue where the default contextual menu of the browser will also appear on the long press, and this pop-up menu cannot be prevented. However, contextual menus are almost always present on web pages, this is not a major issue. Furthermore, it provides visual feedback that a long press has been made.

Mid-Air Techniques

Mid-air techniques are freehand gestures made in front of the Hololens camera which are used as input for content selection. The main benefit of mid-air techniques is that they do not interfere with the touch actions on the web page which may already be busy with other actions implemented such as scrolling and panning. They also offer greater flexibility in the display of the offloaded content. Compared to the direct technique of swiping which can only offload

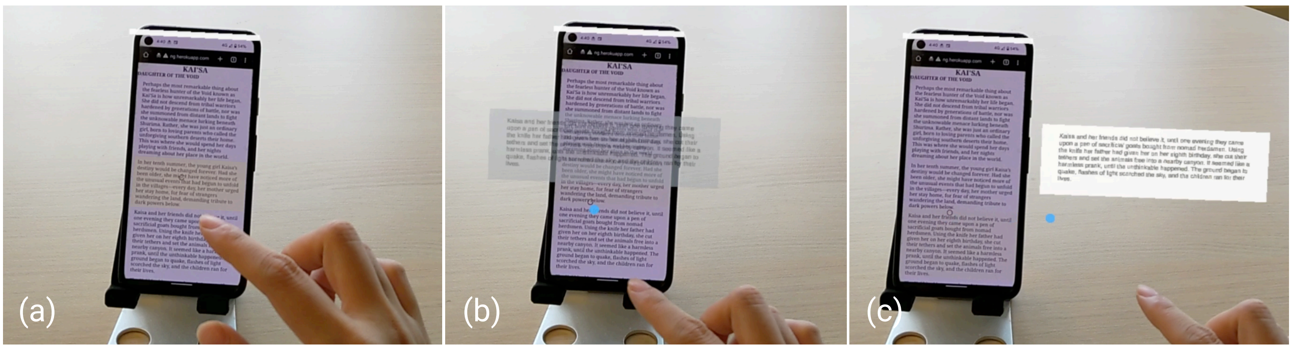


Figure 4.4: **Point and Swipe technique** - As user hovers his finger over the screen, a virtual cursor appears on the webpage and the corresponding element under it is highlighted (a). User approaches the screen to select the element and a preview of the selected content appears.(b) He then swipe to offload it in the direction of his choosing (c)

content to the left or right of the device, mid-air techniques allow content to also be offloaded to the top or bottom of the device since they do not interfere with the vertical scrolling of the web page. In the case of Air-tap and Swipe where the gesture can be made anywhere in front of the camera, it can help to reduce occlusion as the user's fingers do not need to be in front of the device's screen and thus will not block it.

Point and Swipe

When the user approaches the device with his hand and hovers his index finger over the screen, a virtual cursor will appear on the web page opened on the browser of the device [Figure 4.4]. The virtual cursor is controlled by the movement of the finger and the corresponding content block behind the cursor will be highlighted. To make a selection, the user moves his finger closer to the screen until the distance between the tip of the finger and the screen is less than 0.02. When a selection is made, the content in the block is sent to the Hololens and a preview of the selected content block appears in AR. The preview block is semi-transparent and follows the movement of the user's finger. The user can cancel the selection by moving his finger away from the screen, which destroys the preview. Otherwise, the user can then move his finger in any direction to offload the content in the direction of his choice. Once the movement of the finger exceeds 0.05, the content is offloaded onto the side of the device and becomes fully opaque.

Air Tap and Swipe

This is a technique based on the previous work of Brasier et al. (2021) [3]. The distance between the user's thumb tip and index finger tip is tracked. When the distance is less than a certain threshold (0.06), a virtual cursor will appear on the web content on the mobile device[Figure 4.5]. Similar to the previous technique, the virtual cursor follows the movement of the user's finger and the element behind the cursor position is highlighted. To make a selection, the user performs an "air tap" by touching his index finger and thumb against each other. When the selection is made, a preview of the selected content appears in AR. The selected content is semi-transparent and follows the movement of the user's thumb. The user can then drag the content preview in the direction of his preference. Once the drag distance exceeds a swipe threshold of 0.05, the content preview is offloaded onto the side of the device and becomes fully opaque. The user can also release the air tap by moving his thumb and index finger away from each other to cancel the selection.

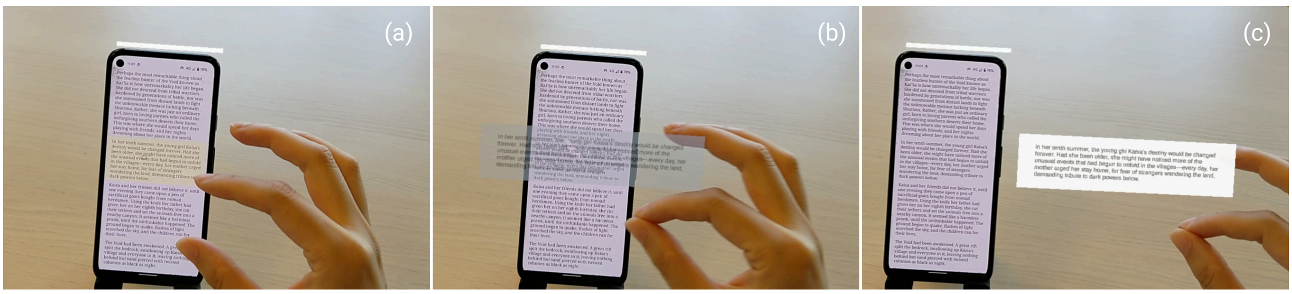


Figure 4.5: **Air-tap and Swipe technique** - As the user hovers over with the screen with a half air-tap gesture, a virtual cursor appears on the webpage and the corresponding element underneath it is highlighted(a). User performs an air-tap to make a selection (b) and drag the content preview to offload it in the desired direction (c).

Combination

To allow for precise selection within a block of text, we implement a technique that is a combination of direct touch technique and mid-air technique [Figure 4.6]. The user first makes the selection of text by first highlighting the text directly on the surface of the device with his finger. He then air-taps and swipes to offload the selected content in the direction of his choice. This technique gets the "best of both worlds" by enabling users to make precise selection while receiving tactile feedback, as well as allowing the content to be placed in any location around the phone.

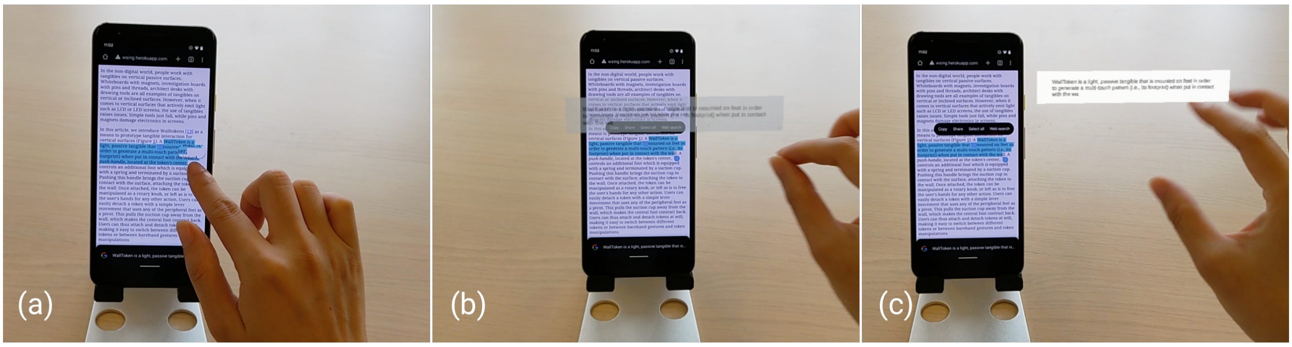


Figure 4.6: **Combination technique** - First, the user makes a text selection on the device(a). He then air-tap to preview the selection (b) and drag to offload it in the desired direction(c).

4.3 Offload Mode

In the previous section on Design, we discussed how the offloading of web content can be initiated automatically by the system or manually by the end user. We believe that most web pages will have a combination of elements that are automatically and manually offloaded. To demonstrate this, we prototyped a scenario of reading a research article on a mobile device [Figure 4.7]. When the user opens up the research article on the web browser, an outline of sections of the article is automatically displayed beside the phone when the page first loads. This outline can give the user a general impression of what the page is about and the different sections within the page. Behind the scene, the javascript on the web page traverses through the document tree of the HTML page and obtains a list of headers which it then sends to the Hololens to be displayed in AR.

As the user scrolls down the article, bibliography references appear automatically in AR whenever their corresponding link texts are visible in the viewport of the mobile device. This is done by putting an event listener on scroll action of the web page that obtains a list of all the bibliography links on the page and checks if the link text is visible in the viewport of the mobile device. The list of visible links is updated and sent to the Hololens everytime a scroll action is detected. Lastly, the user can swipe on any figure in the article to manually offload it. This sends the public URL of the figure to Hololens which renders it as a material on a Unity game object.

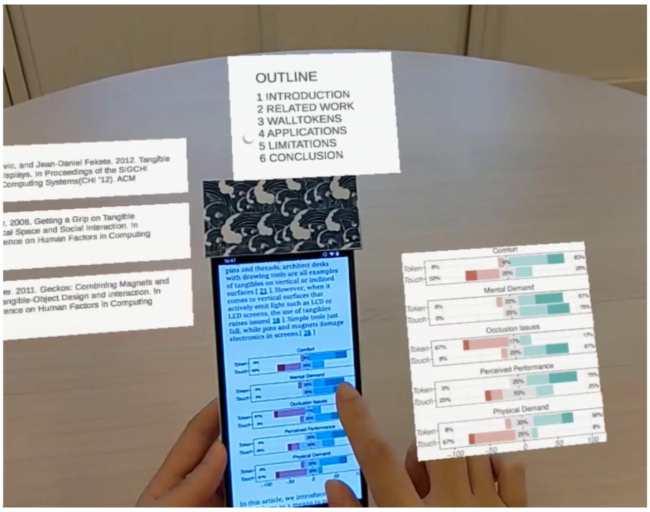


Figure 4.7: Prototype of a research scenario – When the page loads, an outline of the article is displayed. As the user scrolls down the page, references are automatically displayed on the side when the corresponding links are visible in the viewport of the device. Users can swipe to offload any figures. Video of the prototype is available at https://youtu.be/y5OXPFB19V8

.

4.4 Content Display

Our prototype follows a mobile-centric paradigm [17] and uses the device-aligned mode [9]. As such, the information space in AR is spatially registered to the mobile device and moves with it. Direct touch techniques like horizontal swiping on the surface of the device allows content to be placed only to the left or right of the device, while mid-air techniques allow content to be placed in all four directions relative to the phone. Collections are created to organize multiple content pieces within the space.

By default, the offloaded content follows the phone as it moves around. The position of the mobile device is tracked in real time using a Vuforia image tracker placed on top of the phone. When the tracking is detected, a spherical indicator is displayed on top of the image marker. When tracking is lost, the offloaded content stays fixed in the last tracked position until tracking is available again. The user can manually move the offloaded content anywhere in the physical 3D space or onto a clipboard, which is pinned onto a fixed 3D position by default. The clipboard can also be moved around and placed on top of surfaces in the physical world. It can also be unpinned, which enables it to follow the user at a predefined positional offset from the user's gaze. It can be placed in such a way that is only visible from a certain angle when the user glances up or down.

Chapter 5

Evaluation

In this section, we describe the evaluation method and procedure for assessing the performance of our AR enhanced web browsing system and the interaction techniques we designed. We separate our evaluation into two sessions: a preliminary study and the main evaluation session. The objective of the preliminary study is to familiarize the participants with our system and with interacting with content in Hololens 2, and to obtain qualitative feedback on the novel interactive techniques we have developed for offloading content to AR. In the main evaluation, our objective is to assess the effectiveness of AR at enhancing the web browsing experience through common scenarios that users encounter when browsing the web.

5.1 Preliminary Study

In the preliminary study, we show participants three interactive techniques they can use to offload an HTML content block to AR. We assess the effectiveness of each offloading technique on content blocks of different sizes and and at different viewport positions. During the trials, the participant is asked to offload a particular content block using one of the three techniques [Figure 5.1]. The particular content block is highlighted in blue.

Design

We follow a within-subject design with three factors - Technique ∈ {Touch and Swipe, Air - Tap and Swipe, Point and Swipe}, Content Size ∈ {Small, Medium, Large} and Viewport Position ∈ {Top, Center, Bottom}. In the touch and swipe technique, the user swipes on the screen directly to offload the content block towards the direction of his dominant hand. For air-tap and swipe, the user hovers over the screen with his fingers in a semi-air tap position to control

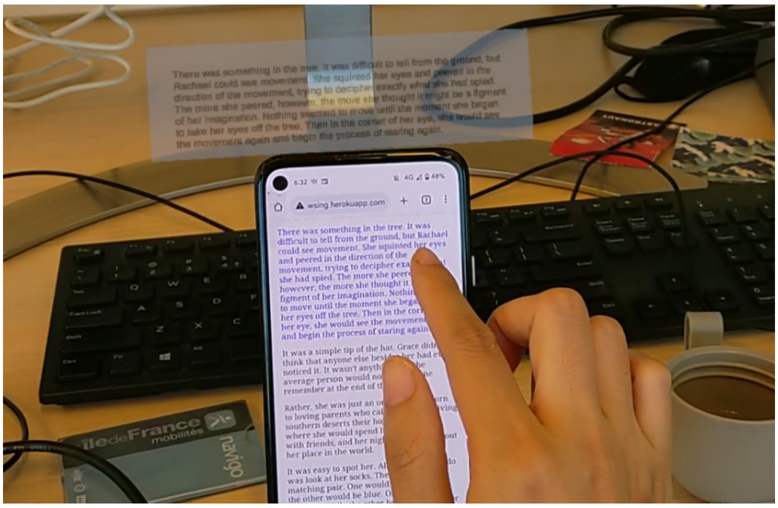


Figure 5.1: In the preliminary study, participants are asked to offload a particular content block (highlighted in blue) using a specific interaction technique.

a virtual cursor. He then air taps and swipes it to offload it in the direction of his dominant hand. For point and swipe technique, the user moves his finger close to the screen to select a content and then swipe to offload it.

Trials are blocked by Technique and counterbalanced across all the participants. Participants will perform 30 trials for each block. At the start of each block, the observer will demonstrate the technique used for the block and three practice trials will be given in order to get the participant familiarized with the technique. The main measures taken are *Completion Time* and *Error Count*.

Hypothesis

H1: User will prefer direct touch technique over mid-air techniques. We hypothesize that direct touch offload will be more performant than the mid-air techniques due to the tactile feedback afforded by the physical surface of the device. However, we are interested in the benefits that mid-air techniques could provide. We are also interested to find out the magnitude of the drop in performance of mid-air techniques over direct touch techniques.

H2: User will prefer Point and Swipe over Air-tap and Swipe technique Between the mid-air techniques, we hypothesize that users will prefer T3 over T2 as the hand movement of the preview and select gesture is closer to the natural form of the hand and is done in one smooth gesture. Holding the fingers in the mid-air tap position may also cause some finger fatigue.

Implementation

A Unity program containing all three interactive techniques is built for the purpose of the pilot study. It is possible to switch between the three techniques either by the participant from within the Unity program, or remotely by the observer view through the web client. The technique is set by the observer at the beginning of each block and the participant do not switch between the techniques during the trials. In our initial prototype, we used Vuforia to track the position of the phone so that we can offload content in the area around it. However, as Vuforia-based phone tracking introduces some jittering that affects the precision of the mid-air selection techniques, the phone is put on a stand and its position registered manually at the beginning of the experiment [Figure 5.2].

Two separate web views are created and deployed for the preliminary evaluation - the observer view and the participant view. The observer view is used by the person supervising the experiment. It acts as a controller for the trials and keeps track of the conditions for the individual trials from a CSV file containing the experiment design. The participant view is what the participant sees on the mobile device used for the experiment. When the supervisor starts the experiment from the observer view, a start button appears on the participant view. When the participant presses the start button, the participant view receives the conditions for that particular trial and renders the web page with the corresponding content size and viewport position. When the participant successfully offloads the highlighted content block, the participant view sends the completion time and error count to the observer view, and the observer view starts the next trial.

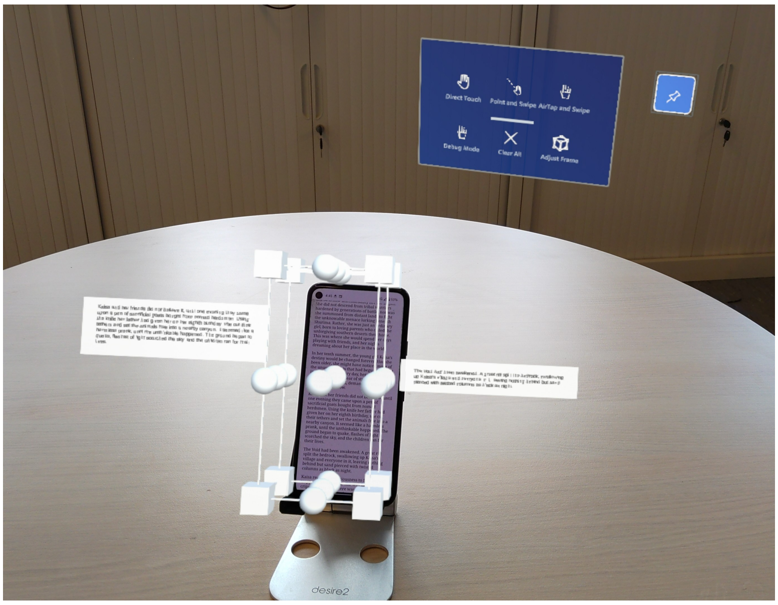


Figure 5.2: The position of the phone is set manually at the beginning of the experiment and an in-app menu allows users to switch between the interaction techniques

Pilot Study

We conducted a pilot study with internal members of our own team. Based on our testing, we found that the system is not robust enough for a wider trial and thus decided not to go ahead with the full study. The prototype transfers data between the mobile device and the hololens over wireless network. Hence, it is very susceptible to sporadic network latency and spikes which could significantly impact the selection time and affect the validity of the result. The prototype is also susceptible to inaccurate hand tracking which prevents users from making accurate selections. From our testing, we also foresee that participants who are inexperienced with interacting in AR with Hololens might require extensive training and will have difficulties completing the tasks particularly with the more novel mid-air interaction techniques.

Furthermore, to ensure consistency between the different techniques, features such as vertical swiping in the midair techniques are disabled for the pilot implementation. As such, we feel that the qualitative feedback from the study will be lacking as the interactive techniques in pilot do not showcase the full benefits of the mid-air techniques.

Based on trials conducted from our internal testing, we find that direct touch technique outperforms the mid-air techniques both in the selection time and the error rates. This is particularly true when the content block size is small or when the block is at the edge of the screen.

5.2 Main Study

In the main evaluation, our objective is to assess the potential impact of AR in facilitating regular web browsing experience. We have designed two common tasks associated with web browsing where there is a potential for AR to enhance the performance of the task. The experiment received ethical approval from the Comité d'Ethique pour la Recherche (CER) de l'Université Paris-Saclay (no 439).

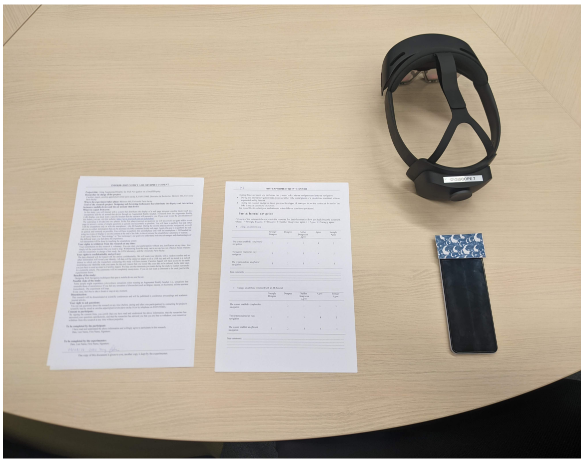
Participants

10 Participants, 6 male and 4 female, were recruited for the study [Figure 5.3a]. The average age of the participants is 26.9 and range from 22 to 46. Most of the participants are master's and PhD students who are working in the lab. 4 of the participants have perfect vision, while the rest wear vision-correction glasses for myopia and astigmatism. The participants gave their consent for participation by signing a consent form at the start of the experiment. No monetary compensation was given to the participants, however a snack was given to them at the end of the experiment to thank them for their participation.

Apparatus

We used Microsoft HoloLens 2 and Pixel 3 smartphone to run the experiment [Figure 5.3b]. The two devices communicate with each other over a wireless network using web sockets. A Vuforia image marker is placed on top of the phone to track its position and movement so that the phone can be used as a reference and content offloaded in AR is displayed around the phone.

(a) Participant performing the experiment



(b) Microsoft Hololens 2 and a Pixel3 phone were used for the experiment

Figure 5.3: Main Study

5.3 AR vs No AR

The objective of the first task is to investigate whether having an augmented AR display will help users relate distributed information on a web page easier. Participants are presented with a web page that contains several paragraphs of text and a fairly complex image. The paragraphs are filler text generated from a random paragraph generator and do not contain any important information. Three questions are randomly distributed between the paragraphs. Each question asks the participant for the number of a specific target item in the image. The participant is asked to find and answer all the questions to complete a trial.

Design and Procedure

The evaluation study follows a within-subject design with two factors: AR Display $\in$ {no AR, AR} and Document Length $\in$ {Short, Medium, Long}. For the smartphone-only condition, the paragraphs and the image are presented on the web page. The participant needs to scroll up and down the page in order to find the questions and to look at the image. For the AR Display condition, the participant will be asked to offload the picture to AR by swiping on the picture. The offloaded picture will always remain in view of the participant regardless of any scrolling so that the participant can answer the question while referring to the picture in AR [Figure 5.4] beside it. The main measures taken are *Completion Time* and *Error Count*.

Figure 5.4: In the With AR Display condition, participants can offload the image in AR so that they can view the image and question side-by-side.

To investigate the effect of back and forth scrolling, we vary the length of the text and hence the number of viewports needed to view the web page in its entirety. Participants are asked to perform the task on texts that have viewport lengths of 4, 8, 12 respectively.

Trials are blocked by AR Display and counterbalanced across all the participants. Participants will perform 15 trials for each block, and each block starts with a practice trial to get the participant familiarized with the task. At the start of the AR Display block, the participant will be given a short introduction on how to use the Hololens 2. An eye calibration is then performed on the Hololens 2 so that the AR content will be displayed at the right focal depth for the participant's eyes.

Implementation

Similar to the preliminary study, two web views are deployed for the experiment - the observer view and the participant view. The observer view acts as a controller for a trial and records the data from the trial, while the participant view is displayed on the smartphone of the participant and used by the participant to complete the trials. 10 images containing target counts of between 3 to 7 are generated from a Processing script. Each image has its corresponding set of three questions and are displayed according to the trial number (for example trial 0 will show image0.jpg). During a pilot conducted, the participant raised difficulty distinguishing between some colors, therefore the images are updated and more contrasting colors are used. Questions are validated by clicking on the 'Validate' button next to a text input field. When all three questions on the page have been answered and validated, it automatically moves to the next trial.

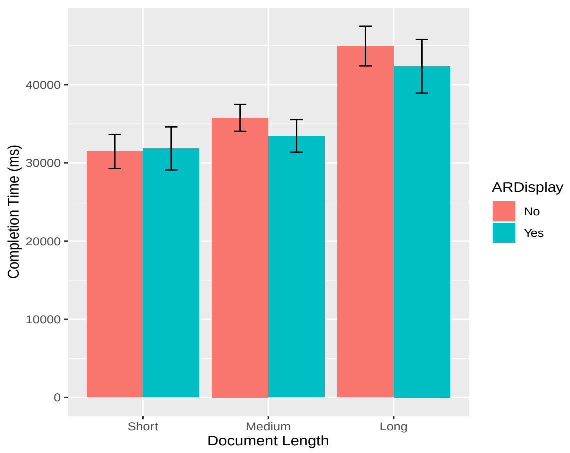
Hypothesis

H1: AR will facilitate information retrieval task We hypothesize that the AR condition will assist the user in completing the information gathering task faster as it will reduce the time needed for the user to scroll up and down the web page to find the relevant information. This will be especially true for longer web pages where more extensive scrolling is required.

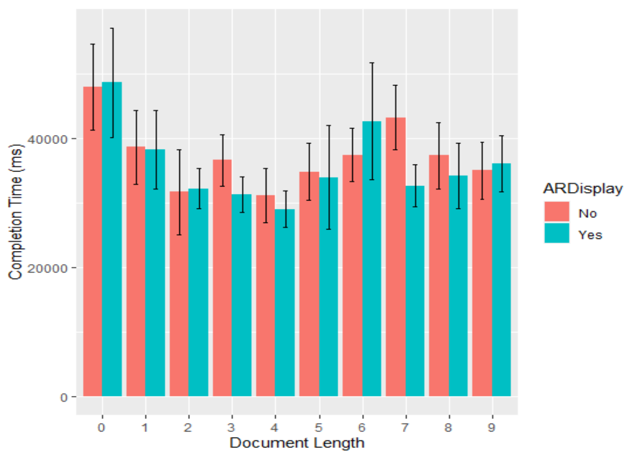
Results and Discussion

H1 is partially supported. We found some evidence supporting the hypothesis that AR Display will facilitate information retrieval tasks. On average, the participants using the AR display are able to complete the task faster on web pages with medium and long viewports [Figure 5.5a].

However, an analysis of variance did not find a significant effect of AR Display on completion time ($P = 0.29$). The ANOVA test revealed a significant simple effect on the number of errors



(a) Mean completion time (in ms) between the AR and No AR condition across three document lengths



(b) Mean completion time (in ms) by participants

Figure 5.5: Results of AR vs No AR

for AR Display (F(1,9) = 12, $p < 0.01$, $\eta^2 = 0.08$). The error count is higher for AR display across different content length. However, as the mean number of errors for each factor is low (from 0.02 to 0.20), we focus the analyses on the correct answers.

When we look at the completion time by participants, we found that there is a lot of inter-subject variability [Figure 5.5b]. Some participants (Participant 3, 7 and 8) seem to have really benefited from the AR condition, while the difference is not so pronounced with others. Some participants reported technical issues with the Hololens, such as difficulties seeing the image clearly in AR, slight delay in loading of the image, and sudden jarring movement of the image due to loss of Vuforia tracking during certain trials. In addition, previous work found that participants take longer time to complete visual search tasks in an integrated smartphone and AR display system due to the different focal distance the content is displayed at [8]. These factors may have affected the participant performance in AR and the task completion time. P3 said that "The system is not comfortable because of the technical limitation of the screen, low resolution, low FOV. I realized that during the first rounds, my eyes were lazy maybe because they have to move more. Despite this, this condition appears much more natural to me compared to without AR."

Despite the difference in completion time not being significant, the AR Display received generally favorable feedback from the participants. In the post experiment survey, 9 out of 10 participants preferred AR Display over smartphone only. The participants rated the AR Display system higher on comfort and efficiency compared to the smartphone-only system [Figure 5.6], giving a median and mode score of 5 for efficiency compared to 2 for the smartphone-only condition and a median and mode of 4 for comfort compared to 3 for the smartphone.

P0 said that "I really liked the fact that we could swipe on the image in order to keep it all the time with us on the right. AR is very practical in that regard." Other participants also liked being able to see the picture and the questions at the same time.

For the smartphone-only condition, participants reported that their eyes and fingers were tired from the extensive scrolling. They also found that completing the task with the smartphone

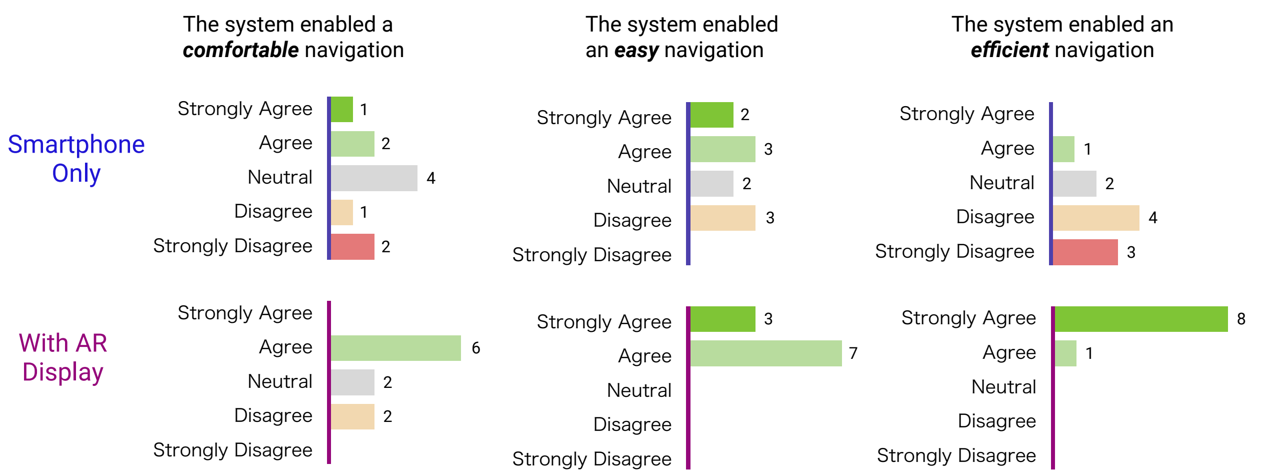


Figure 5.6: Participant rated the AR system higher on comfort, ease and efficiency compared to the smartphone only system

only requires more of a cognitive effort, as they have to keep track of where the questions are located in the paragraph, and remember the target count in order to answer the question. This is consistent with previous work that reported users had to concentrate on mobile web pages in order to keep track of locations on the page [14]. P7 said that “The back and forth between the questions and the picture is tiring physically (because of the thumb) and mentally (because of the remembering).” P1 mentioned that “My thumb hurts a bit from scrolling so much” and P4 also reported that “The amount of scrolling needed doubled with respect to the use of AR. It was very tiring for the eyes to deal with all the scrolling needed.”

The only participant who preferred the smartphone only condition found a trick that allows him to cut down on the scrolling required. “I found I could click on the input field to open the keyboard THEN navigate to the picture to get the information directly and skip navigation back to input. This made the task easier and diminished the mental load of remembering information”, reported P8. “Without this, I would have put 2 on comfortably for this technique.”

5.4 Automatic vs Manual

In Task 1, we compared the condition of smartphone display only against AR display to investigate whether having an augmented AR display will assist in information gathering tasks. For task 2, AR display is used for all conditions. The aim of this task is to investigate whether the automatic offloading of content in AR will facilitate information retrieval tasks compared to manual offloading of content by users. We are also interested in investigating whether the automatic offloading of too much information will disrupt the reading experience and distract the user.

Figure 5.7: Participants are asked to look at all the references linked to in a highlighted paragraph and count how many sources are from ACM. Sources from ACM will contain “ACM” as the last 3 characters.

Task and Procedure

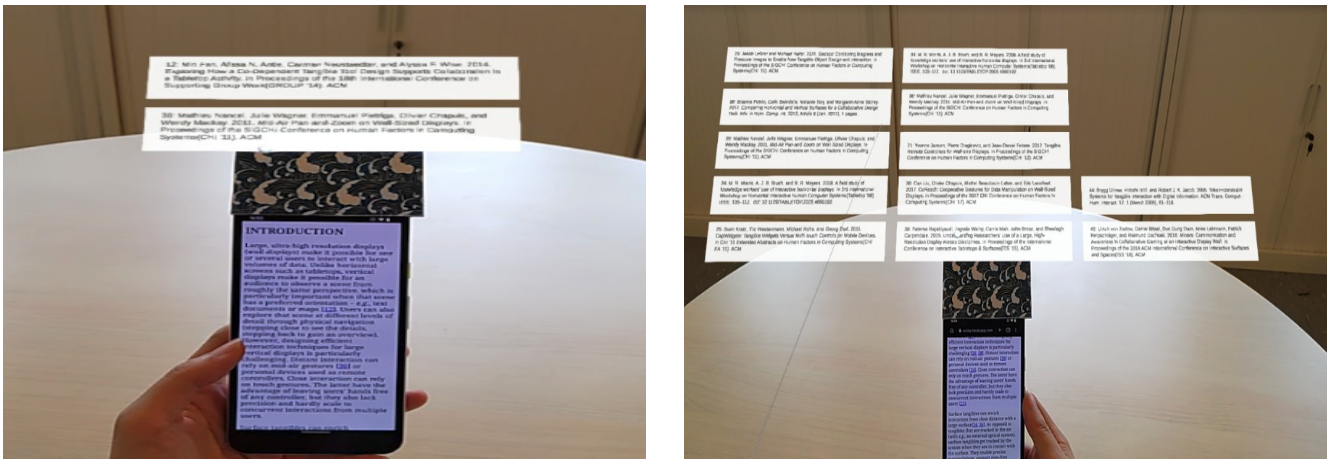
For this task, participants are asked to offload bibliography references from links in a research article. Users will be presented with a research article with a paragraph selected at random. They are asked to find the number of references from the paragraph that are from "ACM" [Figure 5.7]. To allow users to easily differentiate ACM references from non-ACM references, we have manually modified the references so that references from ACM will show "ACM" as the last characters. The experiment setup and apparatus was the same as before.

Design and Implementation

This experiment follows a within-subject design with two factors: Offload Mode ∈ {Manual, Automatic } and Link Density ∈ {Low, Medium, High}. In Manual Mode, the participant can long press on a bibliography link to offload the related citation in AR. In Automatic Mode, the bibliography references that are visible in the mobile device's viewport are automatically offloaded in AR as the user scrolls down the page. To investigate the impact of link density of the offloaded bibliography, we vary the number of bibliography references between trials. Content with low link density will show 1 - 4 references in the AR Display at a time [Figure 5.8a], content with medium link density 5 - 8 references, and content with high link density will display 9 - 12 references [Figure 5.8b]. Trials are blocked by Offload Mode and counterbalanced across all the participants. The participant will perform 15 trials for each block, and each block starts with a practice session to get participants familiarized with the offloading technique and the task. The main measures taken are *Completion Time* and *Error Count*.

Hypothesis

H2: Automatic offloading will facilitate information retrieval tasks. We hypothesize that automatic offload will facilitate the task of looking up bibliography by making the information more readily accessible to the user.



(a) Pages with low link density will display 2-4 references in automatic mode

(b) Pages with high link density will display 9 - 12 references in automatic mode

Figure 5.8: Web pages with different link densities

H3: Too much offloaded content will degrade the overall reading experience. When there is too much content offloaded, we hypothesize that the overload of information will disrupt the flow of the reader and his performance will decrease. The information may overwhelm the reader and make it more difficult to make connections between the individual link and the content relevant to it.

Results and Discussion

H2 is supported. We found strong evidence supporting the hypothesis that automatic offloading will facilitate information retrieval tasks. Participants using the automatic offloading technique had faster completion time across all link densities [Figure 5.9]. Analysis of variance revealed a significant simple effect on Completion Time for Offload Mode ($F(1,9) = 8$, $p < 0.01$, $\eta^2 = 0.29$) and Link Density ($F(2,18) = 24$, $p < 0.001$, $\eta^2 = 0.48$), and no significant effect on error count.

In the post experiment survey conducted, the participants rated the automatic technique higher in all three dimensions of comfort, ease of use and efficiency compared to the manual technique [Figure 5.10]. 7 out of 10 participants chose the automatic technique as their preferred technique as it took less effort for them to retrieve the bibliography information from the link. P0 said that "This was impressive, really quick and efficient! And on top of that you really can put minimal effort into it because it's so automatic and really well-made". P3 reported that "Here (referring to manual technique) you have to perform a lot of repetitions to obtain the information."

The participants who ranked the manual techniques higher said that even though this technique was more efficient, they still preferred the manual technique more as it gave them more control over the information that is offloaded. P5 said "I liked being in control even if it took more time for the tasks. It was less overwhelming".

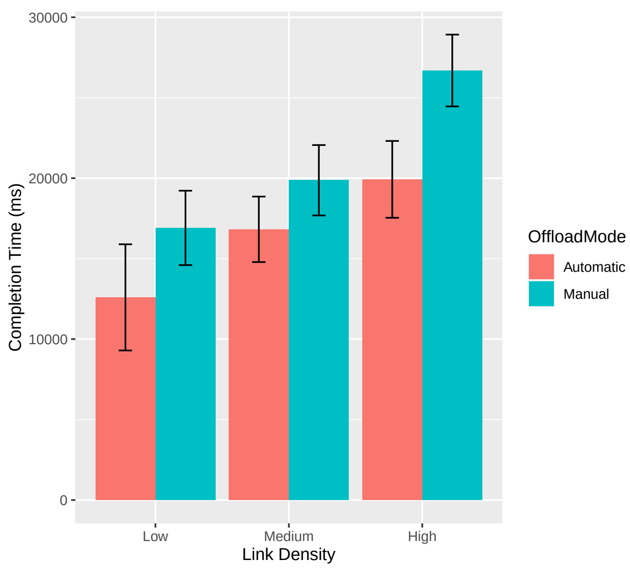


Figure 5.9: Mean completion time (in ms) for Automatic and Manual offload modes across different link densities

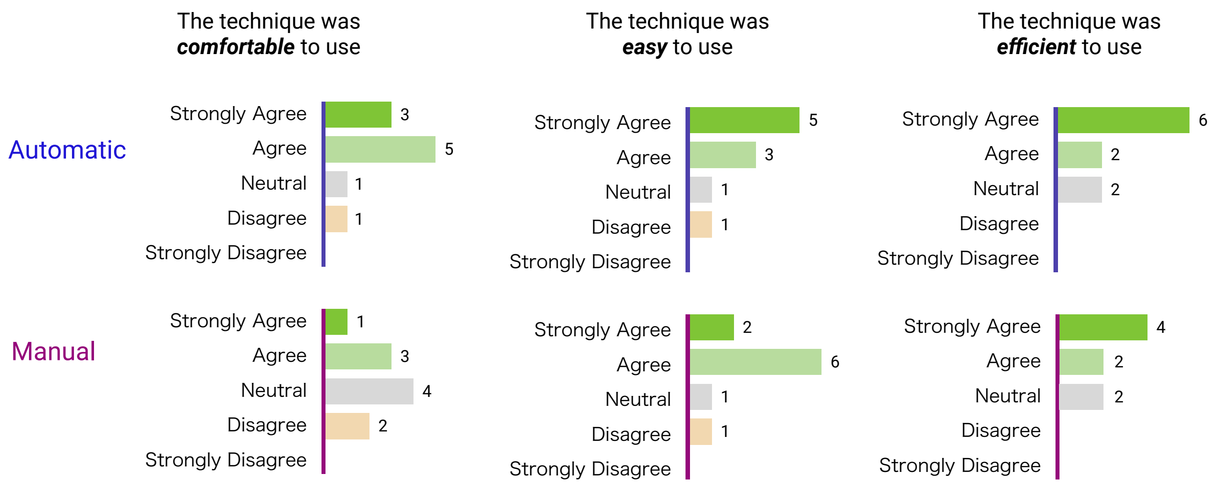


Figure 5.10: Participant rated the automatic technique as more comfortable, easier and more efficient compared to the manual technique

H3 is not supported. We did not find sufficient evidence supporting the hypothesis that too much offloaded content will disrupt the user and degrade the overall loading experience. Our initial hypothesis was when there were fewer content offloaded, the participant could easily make the link between the references and the offloaded content. When there is too much information offloaded, it will be more difficult for users to make the connections between the individual link and the content relevant to it. However, we did not observe this as the participants completed the task faster even for web pages with medium and high link density that have a lot more content being offloaded automatically. A surprising insight is that a participant who preferred the manual option liked the automatic option when there is higher link density, as she would otherwise have had to click on more links which would have required more effort.

In the post experiment survey conducted, most participants did not find the information distracting to the overall reading experience as they can just look away from it or position the content in a position and angle that is not visible if they did not want it. One of the participants reported that she prefers having more information offloaded automatically over less information, as the information is more readily accessible and requires less effort.

CHAPTER 6

Conclusion and Future Work

In this report, we present our exploration of how AR displays can be used to enhance the web browsing experience on mobile devices. Our exploration identified four main categories of use cases where an integrated AR and smartphone system can facilitate web browsing activities and three design factors to consider in such a system. This exploration is just preliminary and there is potential for many new use cases to be explored in the future.

Our implementation of an interactive prototype that can offload different types of web content to AR demonstrates the feasibility of such a system. We also prototyped different interaction techniques for content offloading that are suited for different use cases. A user evaluation validated the practicality of our system. In our study, the participants found the integrated system compelling and greatly enhanced the efficiency and comfort of certain web browsing tasks.

Since our prototype is a proof-of-concept rather than a polished system, there are many future improvements and new features we can make to this system. In terms of technical implementation, one improvement is to make the smartphone and ARHMD communicate directly with each other over a local network or short-range data exchange protocol à la Bluetooth. This would make the system more robust and remove the dependency on wireless network connections, as well as reduce latency that arise from connectivity issues. Another promising avenue to explore is to develop the AR application using web technologies such as AR.js and A-Frame. This provides a better integration with the web client application since they are written in the same language, and allows the AR application to be deployed on any devices that support WebXR.

During our evaluation, two participants found the automatic offloading flashing and blinking too much, which feels a bit jarring and causes the eyes to be tired after a while. More work could be done in the future to make the continuous display of information in AR smoother and more comfortable and pleasurable to the users' eyes. We would also like to explore other scenarios and use cases with our prototype, such as automatically offloading information from social media webpages, and bidirectional loading of web content between the ARHMD and smartphone.

Currently, there exist some technical constraints on ARHMDs including the bulky form factor, narrow field of view, color distortion and low resolution that affect the viewing experience in AR and make it unsuitable for extended usage. In addition, ARHMDs rely on certain environment settings such as lighting conditions for hand tracking and good network connectivity, which limits its context of use. We believe that as AR displays and smartphones continue to evolve and improve, there will be more opportunities to combine them and use them as a unified system. We hope that our contributions will serve as important groundwork for the future of pervasive AR, where AR displays will be used in tandem with smartphones to enhance different aspects of user activities like web browsing tasks.

Acknowledgments

I would like to thank my supervisors Emmanual Pietriga and Caroline Appert for giving me the opportunity to work on such an interesting project, and for their wisdom, patience, and invaluable guidance for every step of this project. They gave me so much incredible support during this project, and I'm lucky to have learnt so much from working with them. They gave me with challenges that pushed me outside of my comfort zone and helped me grow as a designer, developer and researcher. They are also always quick to provide me with the best equipments to make the project successful.

Special thanks to Raphaël James, Olivier Gladin, Emiel Harmsen, Xiang Wei, Eugenie Brasier and Francesco Riccardo di Gioia for their technical advice and help with troubleshooting. Thank you to Nate Hill, the founder of Vuplex, was very responsive to our queries and kindly provided us with a one month free trial of 3D Webview.

Thank you to my academic referent Huyen Nguyen for her guidance and assistance with the entire internship and thesis process. Thank you to Robby van Delden, my critical observer at University of Twente, for reviewing my thesis and providing me with useful advice and feedback. Thank you to Robby and Mariët Theune for their guidance on fulfilling the thesis requirement at my entry university.

Thank you to all the participants of my user evaluation for their time and for their patience during the experiment. Special thanks to Vincent Cavez for his help with participant recruitment.

Last but not least, thank you to my wonderful colleagues at the ILDA Team for making it a great environment to work in. Thank you to my fellow interns Xiang Wei, Johann Equilbec, Julien Berry, and Camille Dupré for being my companions on the same journey. I will always remember fondly our time spent together and the philosophical discussions over delicious meals at CESFO.

Bibliography

[1] M.J. Albers and L. Kim. User web browsing characteristics using palm handhelds for information retrieval. In *18th Annual Conference on Computer Documentation. ipcc sigdoc 2000. Technology and Teamwork. Proceedings. IEEE Professional Communication Society International Professional Communication Conference an*, pages 125–135, 2000.

[2] Ronald T Azuma. A survey of augmented reality. *Presence: teleoperators & virtual environments*, 6(4):355–385, 1997.

[3] Eugenie Brasier, Emmanuel Pietriga, and Caroline Appert. Ar-enhanced widgets for smartphone-centric interaction. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*, MobileHCI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[4] Frederik Brudy, Christian Holz, Roman Rädle, Chi-Jui Wu, Steven Houben, Clemens Nylandsted Klokmose, and Nicolai Marquardt. Cross-device taxonomy: Survey, opportunities and challenges of interactions spanning across multiple devices. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–28, New York, NY, USA, 2019. Association for Computing Machinery.

[5] George Buchanan, Sarah Farrant, Matt Jones, Harold Thimbleby, Gary Marsden, and Michael Pazzani. Improving mobile internet usability. In *Proceedings of the 10th International Conference on World Wide Web*, WWW '01, page 673–680, New York, NY, USA, 2001. Association for Computing Machinery.

[6] Rahul Budhiraja, Gun A. Lee, and Mark Billinghurst. Interaction techniques for hmd-hhd hybrid ar systems. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 243–244, 2013.

[7] Rahul Budhiraja, Gun A. Lee, and Mark Billinghurst. Using a hhd with a hmd for mobile ar interaction. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 1–6, 2013.

[8] Anna Eiberger, Per Ola Kristensson, Susanne Mayr, Matthias Kranz, and Jens Grubert. Effects of depth layer switching between an optical see-through head-mounted display and a body-proximate display. In *Symposium on Spatial User Interaction*, SUI '19, New York, NY, USA, 2019. Association for Computing Machinery.

[9] Jens Grubert, Matthias Heinisch, Aaron Quigley, and Dieter Schmalstieg. Multifi. pages 3933–3942, 04 2015.

[10] Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. Towards pervasive augmented reality: Context-awareness in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 23(6):1706–1724, 2017.

[11] Matt Jones, Gary Marsden, Norliza Mohd-Nasir, Kevin Boone, and George Buchanan. Improving web interaction on small displays. *Computer Networks*, 31(11):1129–1137, 1999.

[12] Peter Mohr, Markus Tatzgern, Tobias Langlotz, Andreas Lang, Dieter Schmalstieg, and Denis Kalkofen. Trackcap: Enabling smartphones for 3d interaction on mobile head-mounted displays. In *2019 CHI Conference*, pages 1–11, 04 2019.

[13] Erwan Normand and Michael McGuffin. Enlarging a smartphone with ar to create a handheld vesad (virtually extended screen-aligned display). pages 123–133, 10 2018.

[14] Virpi Roto and Anne Kaikkonen. A.: Perception of narrow web pages on a mobile phone. *Proceedings of the 19th International Symposium on Human Factors in Telecommunication (HFT 2003)*, 07 2008.

[15] Marcos Serrano, Barrett Ens, Xing-Dong Yang, and Pourang Irani. Gluey: Developing a head-worn display interface to unify the interaction experience in distributed display environments. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '15, page 161–171, New York, NY, USA, 2015. Association for Computing Machinery.

[16] Shumin Zhai, Paul Milgram, and William Buxton. The influence of muscle groups on performance of multiple degree-of-freedom input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '96, page 308–315, New York, NY, USA, 1996. Association for Computing Machinery.

[17] Fengyuan Zhu and Tovi Grossman. Bishare: Exploring bidirectional interactions between smartphones and head-mounted augmented reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–14, New York, NY, USA, 2020. Association for Computing Machinery.

Appendix

Statistical Analaysis

Experiment 1 - AR vs No AR

Legend: 1. AR Display 2. Document Length 3. ARDisplay:Document Length

Repeated measure 2-way ANOVA on Completion Time

Effect	Dn	Df	SSn	SSd	F	p	ges
1	1	9	33864397	241144604	1.263887	2.900062e-01	0.04269600
2	2	18	1565446932	338707745	41.596399	1.782814e-07*	0.67338755
3	2	18	26872330	179434713	1.347849	2.847946e-01	0.03418178

Repeated measure 2-way ANOVA on Error Count

Effect	Dn	Df	SSn	SSd	F	p	ges
1	1	9	0.08066667	0.05933333	12.2359551	0.006744777*	0.08055925
2	2	18	0.08133333	0.43866667	1.6686930	0.216362314	0.08117099
3	2	18	0.01733333	0.42266667	0.3690852	0.696478799	0.01847903

Experiment 2 – Manual vs Automatic

Legend: 1. Offload Mode 2. Document Length 3. OffloadMode:LinkDensity

Repeated measure 2-way ANOVA on Completion Time

Effect	Dn	Df	SSn	SSd	F	p	ges
1	1	9	332733241	355432794	8.425219	1.752040e-02*	0.29158826
2	2	18	738355578	271519617	24.474107	7.341783e-06*	0.47736590
3	2	18	35670418	181420804	1.769553	1.987941e-01	0.04226134

Repeated measure 2-way ANOVA on Completion Time

Effect	Dn	Df	SSn	SSd	F	p	ges
1	1	9	0.024	0.1893333	1.1408451	0.3132843	0.02812500
2	2	18	0.028	0.3853333	0.6539792	0.5318943	0.03265941
3	2	18	0.052	0.2546667	1.8376963	0.1878234	0.05900151

PARTICIPANT QUESTIONNAIRE

Participant ID:

Age:

Occupation:

Gender

F M Other/prefer not to disclose

Eye Vision

Perfect Vision Others (please specify):

On a scale of 1 - 5, how experienced are you with Augmented Reality and Head Mounted AR Displays like Hololens ?

No Experience at All				Very Experienced
1	2	3	4	5

NOTICE D'INFORMATION ET CONSENTEMENT ECLAIRE

Titre du projet : Using Augmented Reality for Web Navigation on a Small Display

Chercheur titulaire responsable scientifique du projet :

Caroline Appert, caroline.appert@universite-paris-saclay.fr, 0169153460, Directrice de Recherche, Bâtiment 660, Université Paris-Saclay

Lieu de la recherche : Bâtiment 660, Université Paris-Saclay

But du projet de recherche : Conception de techniques de navigation pour le Web qui distribuent l'affichage et l'interaction entre un dispositif mobile et l'air autour de ce dispositif

Ce que l'on attend de vous :

Au cours de cette étude, vous allez interagir avec un système qui distribue l'affichage d'une page Web entre un dispositif mobile de type smartphone et l'air autour de ce dispositif grâce à un casque de Réalité Augmentée. Pour bénéficier de l'affichage en Réalité Augmentée (AR), vous devez porter un casque spécifique que l'opérateur vous présentera. Si vous désirez voir les spécifications de ce casque, vous pouvez visiter ce site web : https://www.microsoft.com/fr-fr/hololens.

L'expérience est divisée en deux phases. Au cours de la première phase (*navigation interne*), nous vous demanderons de naviguer au sein d'une page Web pour répondre à une série de questions simples Web le plus rapidement et correctement possible. Vous devrez réaliser cette tâche soit uniquement avec le smartphone soit avec le smartphone + le casque d'AR. Au cours de la seconde phase (*navigation externe*), nous vous demanderons de collecter des informations qui sont accessibles via les liens contenus dans la page Web. Ici encore, le but est d'exécuter la tâche le plus rapidement et correctement possible. Vous devrez réaliser cette seconde phase uniquement avec le smartphone + le casque d'AR mais en utilisant deux types d'affichage pour voir le contenu au bout des liens dans l'air autour du smartphone : automatique ou manuelle.

Dans tous les cas, il n'y a pas de "meilleure stratégie" ou de "meilleure technique", notre but est de comprendre aussi bien les avantages que les désavantages des différentes vos impressions sur le déroulement de l'expérience.

Toutes les interactions se feront en touchant l'écran du smartphone.

Vos droits de vous retirer de la recherche à tout moment :

Votre contribution à cette recherche est volontaire. Vous pouvez arrêter votre participation sans justification nécessaire à tout moment. Il vous suffit d'indiquer à l'expérimentateur que vous souhaitez arrêter. Un arrêt de l'étude ne peut en aucun cas avoir un effet sur les relations futures avec les chercheurs en charge de cette étude, le laboratoire LISN, et l'Université Paris-Saclay.

Vos droits à la confidentialité et au respect de la vie privée :

Les données obtenues seront traitées avec la confidentialité la plus entière. Nous masquerons votre identité à l'aide d'un numéro aléatoire et aucun autre renseignement ne pourra révéler votre identité. Toutes les données seront stockées sur papier ou sur clé USB et seront stockées dans un tiroir fermé à clé auquel seuls les chercheurs menant cette étude auront accès. Caroline Appert aura accès la table associant votre identifiant à votre nom pour la seule raison où vous voudriez que vos données soient effacées. Dans ce dernier cas, il vous suffit d'envoyer un mail à Caroline Appert. Nous pourrons utiliser les commentaires que vous faites au cours de l'étude pour expliquer nos résultats dans un article scientifique. Les commentaires seront complètement anonymes. Si vous désirez qu'un commentaire ne soit pas utilisé, il vous suffit de le signaler à l'expérimentateur.

Bénéfices de l'étude :

Concevoir des techniques de navigation sur le Web qui utilisent l'air autour du téléphone.

Risques possibles de l'étude :

Certaines personnes peuvent éprouver des sensations de *cybersickness* lorsqu'elles portent un casque de réalité augmentée (c'est-à-dire des sensations qui ressemblent à celles du mal des transports). Si vous ressentez une quelconque sensation d'inconfort (comme de la fatigue, des nausées ou des vertiges), prévenez immédiatement l'opérateur. L'expérience s'arrêtera. Dans tous les cas, vous pouvez faire une pause ou vous arrêter à tout moment.

Diffusion :

Cette recherche sera diffusée dans des conférences scientifiques et elle sera publiée dans des actes de conférence et des articles de revues académiques.

Vos droits de poser des questions :

Vous pouvez poser des questions au sujet de la recherche en tout temps (avant, pendant et après votre participation) en communiquant avec le responsable scientifique du projet par courrier électronique à caroline.appert@universite-paris-saclay.fr (ou par téléphone au 0169153460).

Consentement à la participation :

En signant le formulaire de consentement, vous certifiez que vous avez lu et compris les renseignements ci-dessus, que le chercheur a répondu à vos questions de façon satisfaisante et qu'il vous a avisé que vous étiez libre d'annuler votre consentement ou de vous retirer de cette recherche à tout moment, sans préjudice.

A remplir par le participant :

J'ai lu et compris les renseignements ci-dessus et j'accepte de plein gré de participer à cette recherche.
Date, Nom, Prénom, Signature :

A remplir par l'expérimentateur :

Date, Nom, Prénom, Signature :

Un exemplaire de ce document vous est remis, un autre exemplaire est conservé par l'expérimentateur.

INFORMATION NOTICE AND INFORMED CONSENT

Project title: Using Augmented Reality for Web Navigation on a Small Display

Researcher in charge of the project:

Caroline Appert, caroline.appert@universite-paris-saclay.fr, 0169153460, Directrice de Recherche, Bâtiment 660, Université Paris-Saclay

Where the experiment takes place: Bâtiment 660, Université Paris-Saclay

Goal of the research project: Designing web browsing techniques that distribute the display and interaction between a mobile device and the air around that device

What we expect from you:

In this study, you will interact with a system that distributes the display of a web page between a mobile device such as a smartphone and the air around that device through an Augmented Reality headset. To benefit from the Augmented Reality (AR) display, you must wear a specific headset that the operator will present to you. If you want to see the specifications of this hedset, you can visit this website: https://www.microsoft.com/en-us/hololens.

The experience is divided into two phases. In the first phase (*internal navigation*), we will ask you to navigate within a web page to answer a series of simple questions as quickly and accurately as possible. You will have to perform this task either with the smartphone only or with the smartphone + the AR headset. During the second phase (*external navigation*), we will ask you to collect information that can be accessed via links contained in the web page. Again, the goal is to perform the task as quickly and correctly as possible. You will have to perform this second phase only with the smartphone + AR headset but using two types of display to see the content at the end of the links in the air around the smartphone: automatic or manual.

In all cases, there is no "best strategy" or "best technique", our goal is to understand both the advantages and disadvantages of the different ways you feel about the experience.

All interactions will be done by touching the smartphone screen.

Your rights to withdraw from the research at any time:

Your contribution to this research is voluntary. You can stop your participation without any justification at any time. You simply tell the experimenter that you want to stop. Withdrawing from the study can in no way have an effect on future relations with the researchers in charge of this study, the LISN laboratory, and the University Paris-Saclay.

Your rights to confidentiality and privacy:

The data obtained will be treated with the utmost confidentiality. We will mask your identity with a random number and no other information will reveal your identity. All data will be stored on paper or on a USB key and will be stored in a locked drawer to which only the researchers conducting this study will have access. Caroline Appert will have access to the table associating your identifier with your name for the only reason that you would like your data to be deleted. In the latter case, you just have to send an email to Caroline Appert. We may use the comments you make during the study to explain our results in a scientific article. The comments will be completely anonymous. If you do not want a comment to be used, just let the experimenter know.

Benefits of the study:

Designing Web Navigation techniques that span a mobile device and the air.

Possible risks of the study:

Some people might experience *cybersickness* sensations when wearing an Augmented Reality headset (i.e., sensations that resemble those of carsickness). If you feel any sensation of discomfort (such as fatigue, nausea, or dizziness), tell the operator immediately. The experiment will stop.

In any case, feel free to take a break or stop at any moment.

Dissemination:

This research will be disseminated at scientific conferences and will be published in conference proceedings and academic journal articles.

Your right to ask questions:

You can ask questions about the research at any time (before, during and after your participation) by contacting the project's scientific lead by email at caroline.appert@universite-paris-saclay.fr (or by telephone on 0169153460).

Consent to participate:

By signing the consent form, you certify that you have read and understood the above information, that the researcher has answered your questions satisfactorily, and that the researcher has advised you that you are free to withdraw your consent or withdraw from this research at any time without prejudice.

To be completed by the participant:

I have read and understand the above information and willingly agree to participate in this research.

Date, Last Name, First Name, Signature:

To be completed by the experimenter:

Date, Last Name, First Name, Signature:

One copy of this document is given to you, another copy is kept by the experimenter.

Au cours de cette expérience, vous avez exécuté deux types de tâches : navigation interne et navigation externe.

- Au cours des tâches de *navigation interne*, vous avez utilisé soit uniquement un smartphone soit un smartphone combiné à un casque de réalité augmentée.
- Au cours des tâches de *navigation externe*, vous avez utilisé deux types de stratégie pour voir le contenu au bout des liens dans les airs : automatique ou manuelle.

Nous aimerions collecter votre évaluation quant aux différentes conditions que vous avez testées.

Partie A : Navigation interne

Pour chacune des affirmations ci-dessous, entourez la réponse qui caractérise le mieux votre sentiment à l'égard de l'affirmation, où : 1 = Pas du tout d'accord, 2 = Pas d'accord, 3 = Ni pas d'accord ni d'accord, 4 = D'accord, 5 = Tout à fait d'accord.

- Avec un smartphone uniquement

	Pas du tout d'accord	Pas d'accord	Ni pas d'accord ni d'accord	D'accord	Tout à fait d'accord
Le système permettait une navigation *confortable*	1	2	3	4	5
Le système permettait une navigation *facile*	1	2	3	4	5
Le système permettait une navigation *efficace*	1	2	3	4	5

Commentaires libres: ....
....
....
....

- Avec un smartphone combiné à un casque de Réalité Augmentée

	Pas du tout d'accord	Pas d'accord	Ni pas d'accord ni d'accord	D'accord	Tout à fait d'accord
Le système permettait une navigation *confortable*	1	2	3	4	5
Le système permettait une navigation *facile*	1	2	3	4	5
Le système permettait une navigation *efficace*	1	2	3	4	5

Commentaires libres: ....
....
....
....

Classez les deux systèmes en fonction de votre préférence (1 : système préféré, 2 : deuxième système préféré) Si votre préférence est la même pour les deux systèmes, vous pouvez les classer "1" tous les deux.

Système	Rang
smartphone uniquement	
smartphone combiné à un casque de Réalité Augmentée	

Partie B : Navigation externe

Pour chacune des affirmations ci-dessous, entourez la réponse qui caractérise le mieux votre sentiment à l'égard de l'affirmation, où : 1 = Pas du tout d'accord, 2 = Pas d'accord, 3 = Ni pas d'accord ni d'accord, 4 = D'accord, 5 = Tout à fait d'accord.

- Technique ***automatique*** d'affichage des liens dans les airs

	Pas du tout d'accord	Pas d'accord	Ni pas d'accord ni d'accord	D'accord	Tout à fait d'accord
La technique pour collecter l'information disponible au bout des liens était *confortable* à utiliser	1	2	3	4	5
La technique pour collecter l'information disponible au bout des liens était *facile* à utiliser	1	2	3	4	5
La technique pour collecter l'information disponible au bout des liens était *efficace*.	1	2	3	4	5

Commentaires libres: ....
....
....
....

- Technique ***manuelle*** d'affichage des liens dans les airs

	Pas du tout d'accord	Pas d'accord	Ni pas d'accord ni d'accord	D'accord	Tout à fait d'accord
La technique pour collecter l'information disponible au bout des liens était *confortable* à utiliser	1	2	3	4	5
La technique pour collecter l'information disponible au bout des liens était *facile* à utiliser	1	2	3	4	5
La technique pour collecter l'information disponible au bout des liens était *efficace*.	1	2	3	4	5

Commentaires libres: ....
....
....
....

Classez les deux stratégies en fonction de votre préférence (1 : stratégie préférée, 2 : deuxième stratégie préférée) Si votre préférence est la même pour les deux stratégies, vous pouvez les classer "1" toutes les deux.

Stratégie	**Rang**
Stratégie automatique	
Stratégie manuelle	

During this experiment, you performed two types of tasks: internal navigation and external navigation.

- During the internal navigation tasks, you used either only a smartphone or a smartphone combined with an augmented reality headset.
- During the external navigation tasks, you used two types of strategies to see the content at the end of the links in the air: *automatic* or *manual*.

We would like to collect your evaluation as to the different conditions you tested.

Part A: Internal navigation

For each of the statements below, circle the response that best characterizes how you feel about the statement, where: 1 = Strongly disagree, 2 = Disagree, 3 = Neither disagree nor agree, 4 = Agree, 5 = Strongly agree.

- Using a smartphone only

	Strongly Disagree	Disagree	Neither Disagree or Agree	Agree	Strongly Agree
The system enabled a *comfortable* navigation	1	2	3	4	5
The system enabled an *easy* navigation	1	2	3	4	5
The system enabled an *efficient* navigation	1	2	3	4	5

Free comments: ....
....
....
....

- Using a smartphone combined with an AR headset

	Strongly Disagree	Disagree	Neither Disagree or Agree	Agree	Strongly Agree
The system enabled a *comfortable* navigation	1	2	3	4	5
The system enabled an *easy* navigation	1	2	3	4	5
The system enabled an *efficient* navigation	1	2	3	4	5

Free comments: ....
....
....
....

Rank the systems according to your preference (1: preferred system, 2: second preferred system) If your preference is the same for the two systems, you can rank them both "1".

System	Rank
Using a smartphone only	
Using a smartphone combined with an AR headset	

Part B: External navigation

1. For each of the statements below, circle the response that best characterizes how you feel about the statement, where: 1 = Strongly disagree, 2 = Disagree, 3 = Neither disagree nor agree, 4 = Agree, 5 = Strongly agree.

- ***Automatic*** strategy

	Strongly Disagree	Disagree	Neither Disagree or Agree	Agree	Strongly Agree
The technique for collecting the information available at the end of the links was *comfortable* to use	1	2	3	4	5
The technique for collecting the information available at the end of the links was *easy* to use	1	2	3	4	5
The technique for collecting the information available at the end of the links was *efficient* to use	1	2	3	4	5

Free comments: ....
....
....
....

- ***Manual*** strategy

	Strongly Disagree	Disagree	Neither Disagree or Agree	Agree	Strongly Agree
The technique for collecting the information available at the end of the links was *comfortable* to use	1	2	3	4	5
The technique for collecting the information available at the end of the links was *easy* to use	1	2	3	4	5
The technique for collecting the information available at the end of the links was *efficient* to use	1	2	3	4	5

Free comments: ....
....
....
....

2. Rank the two strategies according to your preference (1: preferred strategy, 2: second preferred strategy) If your preference is the same for both strategies, you can rank them both "1".

Strategy	**Rank**
Automatic	
Manual	