Development and Evaluation of an Augmented Reality Maintenance Service for Canon Europe

Rutger Kraaijer BSc

Game and Media Technology MSc thesis ICA-6014518

> Supervisors prof. dr. Remco C. Veltkamp drs. Rob Heemels

Presented for the degree of Master of Science at:

UTRECHT UNIVERSITY Department of Information and Computing Sciences

July 6, 2018

Abstract

This thesis describes the creation of a prototype application for Canon Business & Information Services to provide information from their technical documentation by making use of Microsoft HoloLens. The prototype was positively received during an internal steering committee demonstration and forms the basis for future development regarding this subject. A small-scale experiment showed that the internal tracking capabilities of HoloLens suffered from severe drift issues for the intended use case. Additionally, users wearing the HoloLens made more mistakes and seemed to move around less freely, which could be attributed to the added size and weight of the device and the fear of damaging it. Regarding further development it is advised to look into solutions to ensure instructions to remain spatially aligned and continue evaluation.

Contents

		3
	•	. 3
•••		. 3
		. 4
		. 4
		. 5
		5
		. 5
	•	. 6
		8
		. 8
		. 9
		. 12
		. 13
		. 16
		. 17
		. 19
		. 20
	•	. 20
		22
		. 27
		27
		27 29

1 Introduction

Canon Business & Information Services (BIS) provides Technical Documentation Lifecycle Services. Within this service they create the technical documentation for Canon products. Service technicians are one of the major users of technical documentation and use this for installation and maintenance procedures. Given the current development in Augmented Reality (AR) technology, Canon BIS is interested in providing all relevant information to technicians through an augmented reality headset.

1.1 Possible advantages of AR

See-through Head Worn Displays (HWDs) allow presenting information anywhere around the user. Its content is not restricted to physically available locations that devices like tablets or laptops require. They also allow for hands free operation by making use of gesture and speech recognition.

HWDs provide the opportunity to direct a user's attention to relevant locations thanks to the 3D location tracking technology present on many devices of this type. This could make performing tasks easier and faster to complete and possibly reduce the amount of training required before technicians can start working with machines they are unfamiliar with. An example of this is locating parts and elements (e.g., screws, cable connectors, levers) that need to be manipulated.

Current instruction media, both portable screens like laptops and tablets as well as printed instructions, are also only able to provide a schematic view from a set perspective. These visual instructions need to be understood and processed by the user to translate into the physical location that they refer to.

Additionally, the advantage of replacing a laptop or tablet with a HWD immediately allows the use of other applications like Microsoft's Remote Assist, which is designed for situations where users require aid with a real-life activity, and seems to be of great use for technical and mechanical environments.

1.2 Thesis goals

For this project Canon BIS aimed to work with students to develop a prototype application for Microsoft HoloLens that was able to support technicians with all necessary information and guide them through maintenance procedures on the Océ VarioPrint i300 (Figure 1). An important milestone for this project was the demonstration of the prototype to their steering committee, to show the potential of Augmented Reality for their business. We have also performed an early experiment to evaluate the prototype and compared it to a situation that reflects the way in which service technicians currently perform their work.



Figure 1: The Canon Océ VarioPrint i300 inkjet press

1.3 Evaluating the prototype

1.3.1 Task performance

We are interested in the task performance of users completing maintenance procedures using the prototype, compared to the technical documentation that is currently used. Currently, this is a web page that can be accessed with a tablet. Our first research question concerns the topic:

Q1: Will maintenance task performance significantly differ between users of the prototype and of the normal technical documentation?

The prototype application will support the same information as the technical documentation does, but will be enhanced by additionally showing spatially aligned information. We think that this additional information could be able to enhance task performance and thus make technicians using the HoloLens complete their tasks in a shorter time than users who rely on "traditional" documentation. We expect that some tasks, like locating parts or objects, could benefit from spatially aligned instructions more than others.

1.3.2 Usability

We are interested in how usable the prototype is perceived by users compared to the technical documentation. For this we have formulated the second research question:

Q2: Will the user experience differ significantly between users of the prototype and of the normal technical documentation?

The HoloLens will provide a completely unique experience to the users, which we think could influence the perceived usability of the prototype. However, we also expect that the current documentation will already be perceived to be very useful, so we do not expect to find a significant difference.

1.3.3 Usefulness of spatial instructions

There are multiple options for showing information in an AR environment, as will be discussed in Section 2. We are interested in evaluating the spatial highlights that are shown in the prototype and compare it to the other instruction types that are available in both the prototype and the technical documentation. The corresponding research question is:

Q3: Are the spatial highlights perceived to be a useful addition to the textual and image information that is normally used?

We expect the spatial instructions to be considered very useful since they can directly point towards the real points of interest. When compared to text or text we expect spatial instructions to score at least comparatively.

1.4 Thesis structure

This thesis is structured as follows. In Section 2 we discuss other approaches where AR has been used to provide instructions in industry settings. Section 3 provides detail into the prototype's functionality, design, and implementation and the experiment that has been carried out afterwards. We present the results from the experiments in Section 4 and share our conclusions in Section 5. The thesis concludes with Section 6, which provides recommendations to Canon regarding the future of the project and possible future research subjects.

2 Literature

Maintenance seems to be a relatively popular goal for Augmented Reality applications. A study by Dini and Dalle Mura [3] evaluated 54 AR applications in Through-life Engineering Services and found that when distributing the applications over several service types (maintenance, inspection, training, safety, machine setup), 54% of these applications focused on maintenance.

2.1 Advantages of AR

Research on using augmented reality has been conducted for various types of industry and use cases. To provide an overview of what has been done we discuss some of the research that concerns instructions in AR and what advantages and disadvantages were found.

Tang et al. [17] performed an experiment for an abstract assembly task and compared four media types (print, laptop, AR with a virtual screen, spatially registered AR). Their results showed that printed media resulted in the longest completion times, but that there was no significant difference between the other media types. The error rate and mental workload were significantly lower for spatially registered AR.

Nilsson and Johansson [13] performed a usability study where professional users in a hospital used an instructional application for assembling medical equipment. The results from their experiments showed that the acceptance of an AR system was high for the user group and the application was found to be a supportive tool for learning how to assemble new technology.

Henderson and Feiner [8] describes a prototype AR application to support maintenance on a personnel carrier turret. They tested three conditions: A tracked AR HWD, an untracked AR HWD, and a fixed display. The tracked AR condition was found to have a significant advantage over the others regarding localization of task targets. They found the display condition to have a significant positive effect on task completion time. Their qualitative results could not show any significant preference between the three conditions.

Henderson and Feiner [9] created an AR application to perform maintenance on a Rolls-Royce Dart 510 turboprop engine. They compared an AR and LCD condition. The AR condition had a significantly positive effect on completion time, accuracy and user experience.

Concluding, these studies showed that AR has the potential to make performing tasks easier and more efficient when compared to more traditional interfaces.

2.2 Providing instructions in AR

We identified several methods of providing instructions using augmented reality. We will discuss these through how they convey instructions to the user (presentation) and how the user is able to interact with the application (interaction method).

Presentation

The main medium in which instructions are presented is through visual cues. Visual instructions can consist of several information elements, which can differ from each other in terms of how they are presented. We found that the main factor to distinguish between visual elements is by comparing which coordinate system they are aligned to. Elements can be aligned spatially in the user's surroundings, the local screen space of the device, and egocentrically relative to the user.

Spatially aligned Spatially aligned objects, also referred to as exocentrically aligned in other literature [5], have their position based on a user's surroundings. This can be both a static point in space or a point attached to a real object. This form of alignment requires a process to determine where the device is relative to its surroundings. This can be done by tracking markers, but markerless solutions also exist [14][18].

Spatially aligned information is typically used to show information at precisely the location where it is of most use. An example of this can be found in [6],

where notes could be attached to people and contain their contact information. Looking more at instructional applications, they tend to either show static or animated CAD-like models, or highlight locations of interest.

An example of (animated) CAD models can be found in Reiners et al. [16], where for the task of assembling a door lock into a car door simplified CAD models were animated to illustrate how to perform each action. Tang et al. [17] showed where to place Duplo blocks for an assembly task by showing 3d representations of blocks at the positions where they have to be placed. The AMRA system [2] also superimposed CAD models over a camera feed.

Contrary to the previous examples, the animated 3d models from Nilsson and Johansson [13] were spatially aligned, but not relative to the object itself. Instead, the animations were aligned to a ring that was worn by the user. This approach made sure that the instructions were always positioned close to the equipment, but it would not overlap. The equipment that needed to be assembled in their task was very small, which is probably the reason why they did not want to align directly to the object. Besides the animation, task instructions were also conveyed aurally.

Screen space When in screen space, elements can also be aligned to 2D coordinate space of a screen. This type of information space is often referred to as a Heads Up Display (HUD).

Screen space elements have only been found in combination with spatially aligned content. For example, De Crescenzio et al. [1] shows animated tools in world space and shows a progress bar and text in screen space. Similarly, Platonov et al. [14] provides the same combination of spatial and screen space information. Henderson and Feiner [9] concerns a task where objects need to be aligned. This is done by providing a 3D arrow whose color and size depends on how closely objects were aligned. The arrow disappeared completely when fully aligned. In screen space, text-based instructions are shown.

Egocentrically aligned Egocentric alignment has object positions expressed relative to the user. This is mostly used when the virtual objects needs to be available regardless of the user's position.

Examples using egocentric alignment in AR have not been found in the field of instructional applications, but where the use case of an egocentric user interface itself was of interest. Instances of this are a menu where shortcuts are spread over a virtual sphere surrounding the user [10] and an interface presented in front of the user with virtual windows [4].

Similar to the previously discussed methods, Henderson and Feiner [8] showed instruction text descriptions and a close-up view depicting a 3D virtual scene in screen space. On the target location there were text labels showing the location of the target component and animated 3D models of tools to show their correct movement. Additionally, because the space in which the procedures were done was all around the user, they used additional "attention-directing information" in the form of an arrow to guide the user to a target. The arrow

was located in the center of the screen and pointed to the point of interest. Once the user was oriented towards the point the arrow disappeared.

Interaction method

For interaction methods the most used system we found is the use of speech and voice commands. One of the reasons mentioned for this is the low resolution of the HWD screen [16], but no other research has been found providing a specific reason. Voice commands are generally used for advance steps [17] [13], but more general solutions also exist [7].

Alternatively, external devices like wrist-worn controllers have been used. Henderson and Feiner [8] used an Android smartphone as a wrist controller that served as the primary means of interaction. A 2D application interface showed forward and back buttons to navigate between tasks and when a task was supported by animations, additional buttons and a slider were shown to start, stop, and control the speed of animated sequences. De Crescenzio et al. [1] designed their application to support different types of input, like keyboard, a button device, and also voice recognition.

For step navigation the application can also automatically recognize when a step has been completed. The alignment task used by Henderson and Feiner [9] was tracked in high detail to provide instructions, which inherently enabled recognizing when a task was completed.

3 Approach

In this section we share the functionality, design and implementation details of the prototype that has been developed. This is followed by a description of the experiment that has been conducted using the prototype.

3.1 Development

The prototype has been developed using the Unity 3D engine. Scripts have been written in C#. We have made use of the MixedRealityToolkit-Unity (MRTK) from Microsoft [11], which contains basic scripts and components to get started more quickly with developing applications for HoloLens. The main asset used from MRTK were the scene layout to set up Unity's camera system properly and the gaze cursor to show where the user is looking at and respond to the selection gesture.

3.1.1 Functional requirements

The goal of the prototype was to provide an experience that allows users to see the concept's potential. The features that were found to fulfill this goal are the following. **Spatially aligned instructions** The prototype needed to provide spatially aligned instructions so that the user would not need to switch focus between an external information source and the real object. By making use of a HWD, the user also keeps their hands free to directly execute the provided steps.

Textual and media information The prototype needed to provide details about instructions, so that it is clear what needs to be done exactly. The prototype required the ability to display text, image and video materials for each step in a procedure.

Gesture and voice input The prototype had to provide a multimodal experience, allowing both for both gestures and voice commands to navigate through the interface.

Procedure selection Users had to be able to make a selection from a list of procedures that are possible to perform on a given part of a machine. Once the procedure is selected, the step-by-step instructions should be loaded and presented.

Read procedures from file It was expected that a final application would have procedure data stored on an external source like a remote server. This was however not deemed necessary for the prototype, because network connectivity is no integral part of demonstrating the potential of AR. This would complicate development and demonstrability because the device would then require a network connection when used. To provide a step in this direction, the procedures that are displayed by the prototype had to be loaded from a local file to already provide separation between the application and the data. The files should be formatted in accordance with the DITA standard [15], which is an XML data model for authoring and publishing and an industry standard for storing procedures.

3.1.2 Positioning information

Even though HWDs are capable of displaying 3D information, many applications that are available on these devices can still be described as being 2D applications that are oriented in a 3D environment. Since the text, image, and video information that needed to be available in the prototype are all 2D, we had to determine how to present them. To do so we highlight two different user interaction systems that are already present on HoloLens: The operating system window manager and the Remote Assist application.

HoloLens window manager Figure 2 shows an example of how 2D applications are managed by HoloLens' operating system. The applications are presented as rectangular windows and are placed against the walls of a demo living room. This interface requires windows to be fixed into place for them to become interactable. As the demonstration also showed, the user is assumed to divert their full attention to the applications and should generally remain stationary.



Figure 2: 2D apps running as spatially aligned panels shown during a HoloLens demonstration at Microsoft's Build event in 2015

Remote Assist The Remote Assist application for HoloLens, also developed by Microsoft, is designed for situations where the user is more active and requires aid with a real-life activity. The panel showing the video feed and buttons follow the user when it is moving around and can optionally be pinned into place. When pinned, the panel stays in exactly the same position and orientation it was when activated. As Figure 3 shows, this is useful for situations in which the user's attention is required somewhere else.

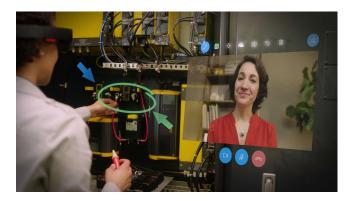


Figure 3: Marketing image of the Remote Assist application for HoloLens

The way in which the Remote Assist application manages the location of its contents can be useful for our prototype as well. However, while performing a procedure, users regularly move around and switch focus points. Even though Remote Assist's pin functionality prevents information from obscuring the concurrent real-life task, we expected that repeatedly pinning and unpinning when changing focus and moving around would quickly become cumbersome. At another extreme, the optimal position of the panel could be predefined for each step in a procedure, but predetermining which position is

optimal assumes that all users would have the same preferences, which was not expected to be the case.

Prototype panel positioning For a middle ground between completely manual and predefined positioning we reduce the number of possible positions by only allowing the panel to be placed onto a unit sphere around the user. Manipulation of the panel position is done similar to how the start menu on HoloLens and the video view in Remote Assist behave, which is that the panel follows user's gaze when the user looks away from the element, causing it to "drag" along. This way, the user is able to manipulate the positioning by only looking around. To stop this behavior, the spherical coordinates can be frozen, stopping the dragging behind the user's gaze. The panel will however stay relative to the user's position to ensure it stays close by when they move around.

To make the panel position itself relative to the user, its positional coordinate is calculated as the sum of the main camera position (the user's head) and an offset based on the spherical coordinate that can be changed by looking around. We use a spherical coordinate to express the position of the panel relative to the user because this allows for expressing the position with two angular parameters. Figure 4 shows a diagram of this positioning system.

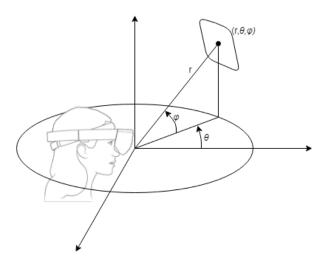


Figure 4: The panel is positioned using spherical coordinates (r, θ, ϕ) with the user at its origin. In the prototype *r* is fixed to 1.

When the panel is set to follow the user's gaze, it needs to be determined whether the user is looking at the panel. This is done by ray casting from the main camera forwards onto the 2D plane on which the panel resides. This point is then projected into the panel's local orientation to retrieve the local x and y coordinates of where the user is looking. Given the size of the plane, an optional padding argument, and the local x and y coordinates, we determine whether the panel is being looked at or not. If the user is not looking at the panel the angular difference between the edges of the panel and the user's current gaze is calculated. Using this angular difference, the spherical coordinates

 ϕ and θ change proportionally, resulting in a smooth movement of the panel towards the user's gaze.

3.1.3 Contents

The panel is constructed from a Unity Canvas element placed in the world coordinate system. It uses the Auto Layout system to place elements in nested groups such as horizontal and vertical layout groups, of which their size is made fully dependent on the layout elements within. This way the panel size changes with its content.

The panel consists of a title bar on top and a content space below. In the prototype, the content space can be populated with the Procedure view, which contains a list of procedures to pick from, and the Instruction view, which is the interface that displays step-by-step instructions.

Voice activatable buttons The HoloLens allows for voice recognition thanks to its array of four microphones and the provided speech API. Therefore, it was fairly easy to adhere to the "See It, Say It" model for voice input from Microsoft HoloLens [12]. This model requires functionality to be both reachable through a button and a corresponding voice command. To make the prototype adhere to this model, all buttons in the panel contain text that is both a description of its functionality and the key word for speech activation. This is achieved by creating a subclass of Unity's default UI Button to create the IconButton class. The IconButton consists of a large icon with a textual description below it. The script attached to the element registers the text on the button to the HoloLens Speech recognition service and triggers its OnClick event when the keyword is spoken. Optionally, the button can be made toggleable, which is used to create the **Play/Pause** button used in the video player, the **Unfollow/Follow** button in the title bar, and the **Video on/Video off** button in the instruction view.

Title bar The title bar has a navigation button on the left, a title centered, and a panel positioning button at the right.

The navigation button is either a **Close** or a **Back** button. Which button is visible depends on how the user reached the current content. If content was reached through navigation from another content view, the **Back** button is visible and can navigate the user back to the previous content. If there is no previous view, the **Close** button is available to close the panel completely.

The panel positioning button shows either a **Follow** or an **Unfollow** button. Which button is visible depends on the current positioning state, as discussed in Section 3.1.2.

Procedure view The procedure view can be summoned by clicking on a spatially aligned target in the user's surroundings. On tapping a target, the panel is activated if it was not already, and the procedure view is shown for the component that the target is associated with. Figure 5 shows the panel when the

target belonging to the Sentry Unit (a component of the Océ VarioPrint i300) is tapped.

Close	SentryUnit	• Unfollow
Remove	e the Sentry Belt unit	
Replace	e the Sentry belt	

Figure 5: The procedure view showing two procedures that are linked to the Sentry Unit

Instruction view The instruction view consists of pages that can be navigated between using the **Previous** and **Next** buttons. The first two pages are the "Context" and "Prerequisites" pages, which show information about the procedure and a list of tools and other necessities that are required to complete it. The subsequent pages are the step instructions.

The explanatory text is always visible. The visibility of the video player can be toggled using either **Video on** or **Video off**. By default, the video player is turned off, but the procedure data structure supports overriding the default which results in the video player showing up immediately when a given step is reached. Figure 6 shows a context page and a step instruction page.

When the final step is reached the **Next** button is replaced by a **Finish** button, which performs the same action as the **Back** button would.

3.1.4 Spatial instructions

The technical documentation currently shows steps as actions described in text with a supporting image that visualizes the points of interest. For example, in the step in Figure 7, a panel needs to be removed. The information is displayed by providing a description of the necessary actions ("Loosen screws (4x)" and "Remove plate") and is accompanied by a photo of the panel. The screws that need to be loosened and the panel itself are highlighted with a circle, and icons expressing the act of "loosening" and "removing" are displayed as well, connected to the corresponding circles. The icons are joined by an additional label displaying the number of points that it relates to. Besides circle highlights the documentation also uses arrows to represent sliding motions, as can be found in the step displayed in Figure 8.

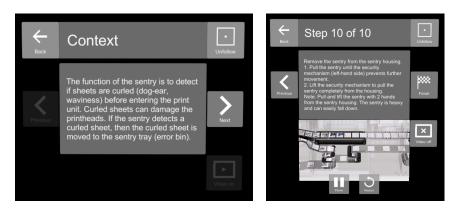


Figure 6: The procedure layout for the "Remove the Sentry belt unit" procedure. Left image: Context page of the instruction layout. Right image: Step 10 accompanied by a video. Since the video player is currently visible, there is a **Video off** button that allows turning off the player.

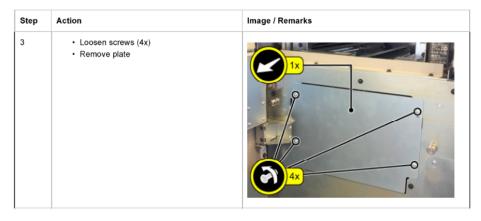


Figure 7: Step 3 of the "Remove the Sentry belt unit" task from Canon's documentation

Highlights In the prototype, highlights were placed onto real objects by using the model tracking capability from Visometry's visionLib [18]. When visionLib recognizes a given 3D object using camera input from the HoloLens, it places a given Unity GameObject into the same location as the origin of the matched 3D model. If the coordinates of the highlights are expressed in the same local space as the model they appear aligned in real life.

With scalability in mind we attempt to minimize the amount of additional information necessary to make instructions from the manual work in AR, since some information - especially 3D locations and orientations - can be time consuming and difficult to gather. Decreasing the efforts required to convert procedures into AR could increase the possibility of actually implementing such a system on a large scale.

Therefore, we analyzed the three factors that are used to place objects in a 3D environment (position, orientation, and size) and the required limitations it poses on spatial highlights when they are omitted. From this we have found

Step	Action	Image / Remarks
4	Lift the sentry belt from the sentry in the upper direction.	

Figure 8: Step 4 of the "Replace the Sentry belt" task from Canon's documentation

that we can show spatially aligned instructions that allow the prototype to convey the added benefits of AR by only specifying positional coordinates:

1 Position

Since object will have to be spatially aligned, a 3D position is considered essential. If the 3D model consists of a clear object hierarchy where each component is a separate child, the components that need to be highlighted can be referenced via e.g. a unique identifier. If there is no such hierarchy available, the position data needs to be specified manually by examining a 3D object and extracting coordinates by visually comparing it to reference images.

2 Orientation

Unlike positioning data, orientation (or rotation) can be omitted, but requires certain limitations on how the highlights can be presented. For example, if a highlight is a sphere that only has a uniform color, orientation has no influence on the object's visual appearance. This shape would however obstruct the object it is highlighting, making it a suboptimal method of highlighting. Another option we found is to automatically orient the highlight by making it always face the user. This way the object practically loses its third dimension, since it is always seen from the same side. However, this 2-dimensional shape can be made hollow.

3 Size

Similar to how orientation can be compared to position, size can be compared to orientation. Only if the size of the object should directly correspond with the size of the of the component that it highlights is it necessary that the size is specified. In most other cases a default size would suffice. However, with a fixed size the elements can overlap when they have to be placed close together. This can be prevented by evaluating the distance between the elements and scaling the size down accordingly.

The highlights in the prototype are two-dimensional rings that only require a 3D position to work. Positions of the points of interest were not available in the manual, nor was the provided CAD file structured in a way that could be referenced to. Therefore, the 3D coordinates were extracted manually by cross referencing the text and images from the manual with 3D CAD-models. The rings always face the user such that the object it highlights stays visible. If objects are close together the rings could overlap. To prevent this, they are scaled down accordingly. The rings can be seen highlighting a set of screws in Figure 9.

Positioning icons We found the combination of highlights with icons an interesting method of displaying information, since they make simple tasks (e.g., opening doors, fastening screws, disconnecting cables) require no additional explanation to understand what to do. This allows technicians who are familiar with the iconography to skip reading trivial step descriptions.

When converted into three-dimensional space, the icons that are placed in the corner of an image do not have an obvious equivalent. Since the procedure we worked with had the technician work with objects at eye level we automatically place the icons 10 cm above the top most highlight (z coordinate), and set the x and y coordinates to be the average of the x and y coordinates of the highlights that it supports. Given that highlights are generally placed close to each other, this placement method allows the icon to be close by without having to manually specify its location.



Figure 9: Step 5 of "Remove the Sentry belt unit" as seen from the perspective of the technician using the prototype. The four screws that need to be loosened are high-lighted and accompanied by the "remove screw" icon above it. The spatial highlights are slightly misaligned due to the HoloLens' screenshot feature.

3.1.5 Out of bounds indicator

Feedback from the steering committee demonstration (discussed in Section 4.1) showed that locating highlights that were instantiated out of view were difficult to locate. To remedy this, small arrow shaped indicators were added that

point to the highlights that are not currently in view. The out of bounds indicators are placed at the edge of the screen so that it is closest to its element and is also rotated such that the arrow points in the direction of the element. For the user this means that when they shift their gaze towards the arrow, which is further emphasized with the direction in which the arrow is pointing, their gaze gets closer to the element it highlights. Once an element enters the user's field of view, the arrow disappears. Figure 10 shows what the dot indicator looks like.

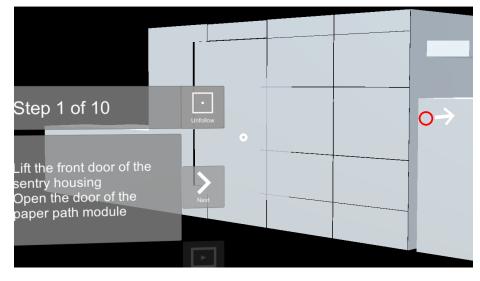


Figure 10: A screenshot of the development environment, showing a dot indicator at the right pointing to a highlight at the sentry housing door

The out of bounds indicators are placed onto a two-dimensional canvas that fills the HoloLens' display. To determine the 2D coordinate of the indicator based on the 3D position of an object, we transform the 3D world space coordinate of the object to the main camera's local coordinate system. This results in a vector that corresponds to the location of the object display's x and y coordinates. To place the indicator on the canvas, the angle between the position vector and a normal vector (1,0) is calculated. This angle is then used to get the location of the point that lies on the border of the canvas. The angle is also used to rotate the arrow sprite such that it points in the correct direction.

Due to the fact that HoloLens refreshes its orientation separately from Unity's frame update cycle, we added some padding to ensure the arrows do not clip out of the screen when the user looks around.

3.1.6 Procedure data structure

Procedures are saved as individual files in the XML format and expect files to conform to the DITA standard. To internally represent the procedures we use two structs, Procedure and Step, which contain information regarding the procedure itself and the specific step-by-step instructions. We use an enumeration called StepType to determine the appropriate icon to place above the augmented highlights. Figure 11 shows a class UML of Procedure, Step and Step-Type. Figure 12 shows an excerpt of the XML file for the "Remove the Sentry belt unit" procedure.

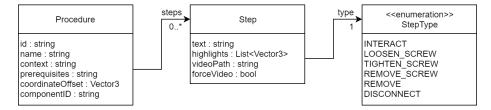


Figure 11: The class UML used for representing procedures

```
<task id="RemoveSentryBeltUnit" componentID="SentryUnit">
  <title>Remove the Sentry Belt Unit</title>
  <offset>
   <x>0.583857</x>
   <y>0</y>
    <z>-0.841318</z>
  </offset>
  <taskbody>
    <context>
     The function of the sentry is to detect if sheets are curled
      (dog-ear, waviness) before entering the print unit. Curled
      sheets can damage the printheads. If the sentry detects a
     curled sheet, then the curled sheet is moved to the sentry
      tray (error bin).
    </context>
    <prereq>
     screwdriver torx 10
        screwdriver torx 20
      </prereq>
    <steps>
      <step type="INTERACT">
       <cmd>Lift the front door of the sentry housing</cmd>
        <cmd>Open the door of the paper path module</cmd>
        <video forced="false">
         <source src="RemoveSentryBeltUnit_1"></source>
        </video>
        <highlights>
         <coordinate>
           <x>2.010999917984009</x>
           <y>0.4669999957084656</y>
            <z>-0.43700000643730166</z>
         </coordinate>
        </highlights>
      </step>
    </steps>
  </taskbody>
</task>
```

Figure 12: Fragment of XML structure containing the first step of the "Remove the Sentry belt unit" procedure

3.2 Experiment

Due to intellectual property issues between Canon and Océ it was not possible to get access to the required data (specifically CAD data) to create a procedure with enough steps to perform the experiment with. Therefore, we set up an experiment that did not require assets from Canon or Océ where students could participate in instead. Considering the actions that were described in the "Remove the Sentry Belt unit" procedure from Canon, we found that the task of assembling a computer was both mechanically similar and practically feasible to organize on the relatively short notice that was required to finish the project.

3.2.1 Preparation

The procedure for the experiment is called "Assemble a File Server". Participants were provided an empty computer server case and several components that would need to be installed. The screws were ordered and labeled to be easily identifiable. The procedure consists of 23 steps. The procedure guided the user through the installation of a motherboard, RAM, DVD drive, front panel I/O, power supply, and case fan. Each individual action (placing, securing with screws, connecting cables) was denoted as a separate step and had the appropriate text, image and spatial highlight(s). The images contained red rings around points of interest similar to the spatial highlight rings and their locations corresponded with each other. Figure 13 shows the computer case and the components used during the procedure.

Because there were no high detail 3D models available for the computer case, the model tracking calibration process was replaced by a manual process, which ensured that each participant started with the same precise alignment.



Figure 13: Starting situation of the experiment

3.2.2 Execution

Participants were placed in either the AR or control group. Group assignment was done in an alternating manner to keep the group sizes as equal as possible. The participants (n = 20) were students between 18 and 25 years old. In both groups 40% of the participants were female and 60% were male.

When starting the experiment, the participant was first explained that they would have to assemble a computer and that they would receive instructions either through an application on the HoloLens or through a laptop, based on the group they were assigned to.

Participants in the AR group then watched a live tutorial in which they could see the researcher's point of view through an external display. They were explained how to operate the HoloLens and how to interact with the prototype. A separate "Example procedure" was used for this purpose, which contained a single step that did not occur in the procedure used for the experiment and required participants to remove a small screwdriver from its holder from inside the case. After the demonstration the participants got to wear the HoloLens and were asked to perform the example procedure. Afterwards the participant was asked whether they were comfortable with the controls and the device. If this was not yet the case, they received help until they were.

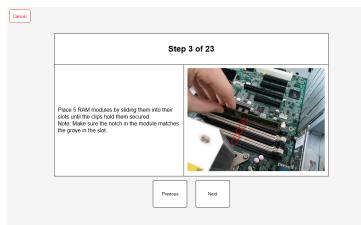


Figure 14: Step 3 from the "Assemble a File Server" in the documentation used by the Control group

The participants then performed the procedure. The AR group used the HoloLens to get instructions, and the control group had access to the technical documentation through a custom web interface that provided the same information as was available in the prototype, except for the spatially aligned highlights. A custom interface was used to allow for automatic time measurement of each individual step. Figure 14 shows a screenshot of the interface that was used by the control group.

We recorded the time that it took to complete each step of the maintenance procedure and after completing the procedure participants filled in a questionnaire containing the following questions:

- 1. Participant number
- 2. 5-point Likert scale rating of self perceived experience level of:
 - (a) Performing small mechanical and technical activities
 - (b) Assembling/building a computer
 - (c) Using Microsoft HoloLens before the session started
 - (d) Using gesture-based interfaces (e.g. Nintendo Wii remote, Virtual Reality controllers)



Figure 15: Photos of participants during the procedure. Left: AR group. Right: Control group. Photos are used with permission from participants.

- 3. 5-point Likert scale System Usability Scale questions
- 4. 5-point Likert scale question regarding perceived usefulness of information types
- 5. 5-point Likert scale Single Ease Question
- 6. Optional open answer feedback field

There were some small differences between the questionnaires for the two groups. The control group was not asked for questions 2c and 2d since the results from these questions would only be used in combination with other results that were specific to the AR group. The questionnaires for both groups can be found in Appendix A. The Likert-scale answer response anchors were based on the work of Wade [19].

4 Results

4.1 Steering committee demonstration

The demonstration for the steering committee took place on April 10th 2018. The version of the prototype that was presented was positively received and support was granted to allow for further efforts in researching the possibilities for AR. Feedback showed that spatial highlights were sometimes difficult to locate, especially when they were instantiated outside of the user's view. Before the experiments started a solution was implemented, as is described in Section 3.1.5.

4.2 The influence of experience

Figures 16 and 17 show the results of the experience related questions for the AR and Control group respectively. As can be seen, the distribution of expe-

rience is nearly equal between the groups. In both cases, participants most often rate their experience in performing mechanical activities to be average, and over half of the participants mention some experience regarding assembling computers. The largest portion of participants does not mention having any prior experience with the HoloLens, and experience with gesture-based interfaces is nearly evenly distributed over all categories.

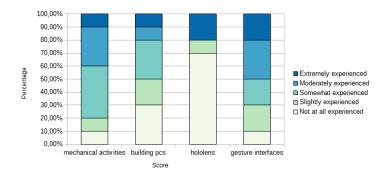


Figure 16: Results from the experience questions of the AR group

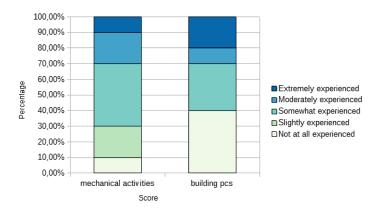


Figure 17: Results from the experience questions of the control group

Based on the participant's perceived level of experience of both mechanical activities and building computers, we can see whether a correlation can be found between these experience levels and both the completion time and the SEQ score. The correlation coefficients of these combinations can be found in Table 1.

Group	Experience	Time	SEQ
AR	Mechanical activities	0.106	-0.0810
	Building computers	0.412	-0.255
Control	Mechanical activities	-0.188	-0.745
	Building computers	0.0756	-0.500

Table 1: Correlation coefficients for the experience levels and the time and SEQ results

As can be seen in Table 1, a strong negative correlation was found between the perceived experience of both mechanical activities and building computers with the SEQ, which expressed the perceived task difficulty. This correlation means that participants who perceived their experience to be greater at these subjects generally seem to find the task easier. This was in line with expectations, since both of these data points are provided by the participants themselves, but nonetheless noteworthy.

A weak positive correlation was also found between computer-building experience and the completion time for the AR group, which indicates that participants who had higher experience of building computers seem to be completing their task slower. We do not know what might have caused this, since the opposite would be expected and we cannot think of an explanation why specifically more experienced participants would perform slower when using the HoloLens.

4.3 Tracking stability

During the experiments we identified an issue regarding the tracking capabilities of the HoloLens that did not occur during internal testing and other demonstrations. During the experiments all participants reported that even though initially all spatial information was aligned correctly, the highlights slowly drifted away until they were not aligned any more. This issue did not affect every participant at the same time and in the same intensity, but it occurred to every participant before reaching halfway through the procedure, in all cases reaching a drift of multiple centimeters.

To see whether the drift impacted task performance we have plotted step completion times for the two groups, which can be seen in Figure 18. The completion times for AR appear to be slightly higher for most steps. Additionally, we have performed a two-tailed unpaired t-test on the slopes of the linear regression lines of the completion times of both groups, which resulted in no significant differences (p = 0.525).

If users were dependent on the spatial highlights during the procedure, the drift is expected to negatively influence completion time. This would increase the difference in completion time since the Control group did not experience this issue. However, it seems that the AR group was not affected by the drift, since the difference in completion time did not change noticeably and the slope of the regression lines on the completion times did not significantly differ between the two groups.

4.4 Repeated errors

Once participants were finished with the procedure the end result was inspected for errors. Among other things these could have been incorrectly placed elements and missing or misplaced screws.

We identified one mistake that happened surprisingly frequently in step 15. In this step the power supply needed to be secured to the case. One of the

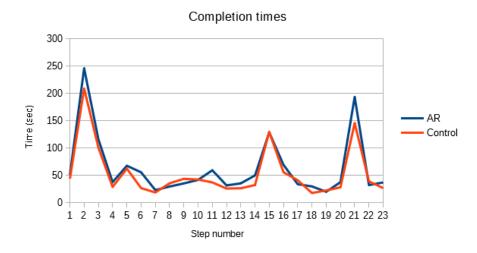


Figure 18: The average step completion times of both groups

four screws necessary to do so was regularly inserted into a hole roughly 2 cm above the one that is shown in the instructions, as can be seen in Figure 19. This error occurred five times in the AR group and once in the Control group. One of the five participants from the AR group noticed the error during the procedure and corrected their mistake, but for the other participants this went unnoticed.

4.5 Task performance

Figure 20 shows a plot of the performance recorded for the AR and control group. The average completion times were 1467 seconds for AR (24.45 min) and 1245 seconds for Control (20.76 min).

An issue during one of the experiment runs of the AR group required the procedure to restart, rendering the time results for the first two steps of one participant incorrect. Since this was the only occurrence of faulty data we have substituted these two times with the means of the other participants in the group.

We performed two-tailed unpaired t-tests on the performance results between the AR and Control group. The p-value for the total task times shows no significant difference between the two groups (p = 0.07). Three individual steps, 14, 16, and 23, do show significant p-values (p = 0.02, 0.03, 0.04 resp.). In these cases the AR group performs slower than the control group, having average times of 50.1, 69.6, and 37.4 seconds for AR and 32.6, 56, and 26.9 for Control. Steps 14 and 23 concern the placement of the power supply and the front panel. The AR group tended to initially place the power supply and front panel upside down. Step 16 concerned securing two screws that were partially obscured by cables coming from the power supply.



Figure 19: The left hole is not a screw hole but part of the ventilation grid of the power supply and was frequently confused with the screw hole on the right

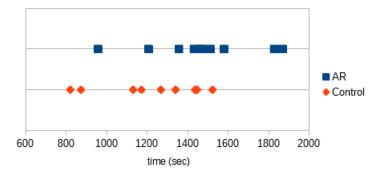


Figure 20: The total task performance times for both groups

4.6 Usability

To measure usability we used the System Usability Scale (SUS). Additionally, we measured the perceived task difficult with the Single Ease Question (SEQ). The SUS and SEQ scores can be found in Figures 21 and 22. The average SUS scores were 77.75 for AR and 83 for control. The average SEQ scores were 3.7 for AR and 4.4 for Control.

We performed two-tailed unpaired t-tests on the SUS and SEQ scores between the AR and Control group. The SUS scores were not significantly different (p = 0.10), but the SEQ scores of Control were significantly higher than AR (p < 0.01). The applications that were used were not found to be perceived significantly different, but the procedure was perceived to be significantly easier by the Control group.

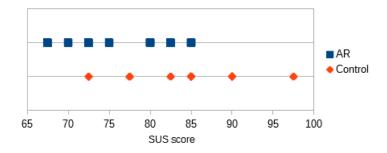


Figure 21: SUS scores for both groups

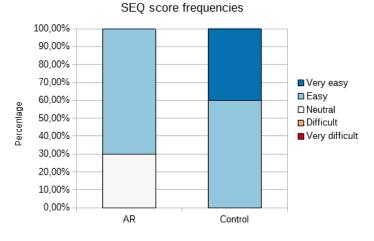


Figure 22: SEQ score frequencies for both groups

4.7 Usefulness of spatial instructions

Figures 23 and 24 show the results of the questions regarding the perceived usefulness of the available instruction types.

We performed an ANOVA on the usefulness scores within the AR group which did not show any significant difference between the information types (p = 0.540). Two-tailed unpaired t-test also did not show a significant difference between the perceived usefulness of the text and image information types between AR and Control (p = 0.774 and p = 0.660 resp.). The average usefulness scores were 3.8 (text), 4.2 (image), and 3.8 (spatial) for AR, and 3.9 (text) and 4.4 (image) for Control. The standard deviations of the usefulness scores were 0.92 (text) 1.0 (image) and 0,79 (spatial).

5 Conclusion

For this project we have built a prototype application that is able to guide users through maintenance procedures. Only positional coordinates were neces-

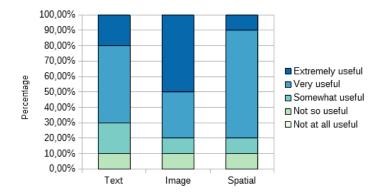


Figure 23: Perceived usefulness of instruction types by the AR group

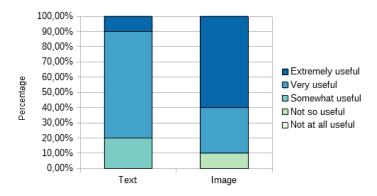


Figure 24: Perceived usefulness of instruction types by the Control group

sary to show highlights in our solution, minimizing the additional information needed to convert existing procedures into its AR counterpart.

The prototype was demonstrated to the steering committee and was positively received, granting support for further efforts into researching the integration of AR into Canon's Technical Documentation Lifecycle Services.

For interpreting significance we had initially set the p-value threshold at 0.05. Using this value we cannot confirm to have found a significant difference between task performance between the two groups (p = 0.07) as was expected for research question Q1, but we think a larger sample size could have resulted in a lower p-value. Regarding the usability question Q2, we found that the procedure was perceived to be significantly easier by the Control group than AR based on the SEQ score. The SUS score that directly measured the prototype's usability did however not reach a significant p-value (p = 0.10). These results suggest that the usability of the prototype was lower than the system used by the Control group. However, we believe the drifting issue could have affected the prototype's usability unfairly. Therefore no definite confirmation can be reported regarding the usability of the prototype. Similarly, we cannot conclude that the spatial information can be considered a useful addition to the textual and image information as was expected for question Q3. The perceived

usefulness of spatial instructions was neither significantly better or worse and the averages and standard deviations are very comparable between the different types, but we believe that participants did not grade the application itself but rather their initial impressions and the concept's potential.

The drifting issue and its evaluation in Section 4.3 show that it is highly probable that once the spatial highlights became uninformative the participants switched to using only text and image information, which is the same information that the Control group used. Still, the error reported in Section 4.4 occurred five times in the AR group and just once in the Control group, and the AR group also required more time to complete their task than Control. The usability test showed that the procedure was perceived to be significantly easier by the Control group than AR.

We believe that there is an underlying reason that explains these results, which is based on an observation that was made throughout the experiments. We found that users wearing the HoloLens did not move around as much and generally kept a greater distance from their hands and focus points compared to the Control group. Participants also did not seem to look around the computer case as freely as the Control group did. The size and weight of the device and possibly the fear of damaging it by bumping into objects or letting it slide of one's head when tilting are thought of to be possible causes for this behavior. Participants also more often initially tried inserting components in an incorrect orientation and seemed to have more trouble with connecting cables to the small connectors on the motherboard. This could be attributed the device limiting the user's view due to the combination of its stacked lenses and the protective outer shell. We acknowledge that our prototype might also be part of the cause, but we do not think this would be sufficient to explain the differences that were found.

6 Future work

Recommendations for Canon

Improving tracking stability Considering the results from our experiment it is clear that the biggest efforts for this project should be spent on increasing tracking stability. This could be made possible by continuously using the already present model tracker, but because of the large size of the machines that Canon works with more intricate solutions are expected to be necessary.

Gathering positional coordinates For large scale use cases, manually providing coordinates, as has been done for this prototype, is possible but not realistic. The process could be simplified by developing a tool that allows for easy editing, but a more streamlined way to do so would require CAD or other 3D model data to contain a clear hierarchical structure. Deciding which of these solutions is optimal requires an analysis and comparison of the required efforts of implementing them.

Intellectual property and further experiments Canon did not have direct access to Océ's data, due to intellectual property rights. Due to this, data like CAD models could not be shared with us, resulting in the decision to perform the evaluation experiment in a different setting. Previous research has shown that AR is capable of significantly shortening localization times, which our experiment would not have been likely to fully demonstrate since the computer case is relatively small. When the tracking stability is increased and the intellectual property rights issue is resolved, a similar experiment like ours could be performed on an Océ VarioPrint i300.

Future research

HoloLens' tracking precision on millimeter scale could be circumvented by using active tracking methods that continuously locate the object of interest, which would compensate with the drifting found during our experiment. However, we are doubtful that the cameras that are currently present in the HoloLens have a high enough resolution to do so. Research into what the minimum requirements are for sensors to do so would be interesting.

We only used position coordinates to present spatial highlights. Combined with images and video, we think this is a balanced method of providing spatial information without requiring much manual labor. Many solutions rely on three-dimensional spatially aligned instructions, but whether a measurable difference exists would be interesting.

When the HoloLens is close to an object it tends to lose track of its position. Since technicians are generally close to the machine they are working at, this could become a regular occurrence. This issue could be solved by letting the software recognize when it is in such a scenario and only use gyroscope and accelerometer data, or add sensors to the back of the device. If software is able to recognize and properly handle these situations, additional sensors are not required.

References

- [1] Francesca De Crescenzio et al. "Augmented reality for aircraft maintenance training and operations support". In: *IEEE Computer Graphics and Applications* 31.1 (2011), pp. 96–101.
- [2] Jean-Yves Didier et al. "AMRA: augmented reality assistance for train maintenance tasks". In: Workshop Industrial Augmented Reality, 4th ACM/IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2005). 2005, Elect–Proc.
- [3] G Dini and M Dalle Mura. "Application of augmented reality techniques in through-life engineering services". In: *Procedia CIRP* 38 (2015), pp. 14– 23.
- [4] Barrett M Ens, Rory Finnegan, and Pourang P Irani. "The personal cockpit: a spatial interface for effective task switching on head-worn displays". In: *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM. 2014, pp. 3171–3180.
- [5] Barrett Ens, Juan David Hincapié-Ramos, and Pourang Irani. "Ethereal planes: a design framework for 2D information space in 3D mixed reality environments". In: *Proceedings of the 2nd ACM symposium on Spatial user interaction*. ACM. 2014, pp. 2–12.
- [6] Steven Feiner et al. "Windows on the world: 2D windows for 3D augmented reality". In: Proceedings of the 6th annual ACM symposium on User interface software and technology. ACM. 1993, pp. 145–155.
- [7] Stuart Goose et al. "Speech-enabled augmented reality supporting mobile industrial maintenance". In: *IEEE Pervasive Computing* 2.1 (2003), pp. 65–70.
- [8] Steven J Henderson and Steven Feiner. "Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret". In: *Mixed and Augmented Reality*, 2009. *ISMAR* 2009. *8th IEEE International Symposium on*. IEEE. 2009, pp. 135–144.
- [9] Steven J Henderson and Steven K Feiner. "Augmented reality in the psychomotor phase of a procedural task". In: *Mixed and Augmented Reality (ISMAR)*, 2011 10th IEEE International Symposium on. IEEE. 2011, pp. 191– 200.
- [10] Frank Chun Yat Li, David Dearman, and Khai N Truong. "Virtual shelves: interactions with orientation aware devices". In: *Proceedings of the 22nd annual ACM symposium on User interface software and technology*. ACM. 2009, pp. 125–128.
- [11] Microsoft. *MixedRealityToolkit-Unity GitHub repository*. Jan. 2018. URL: https: //github.com/charlespwd/project-title.
- [12] Microsoft. Windows Dev Center Windows Mixed Reality. Dec. 2017. URL: https://developer.microsoft.com/en-us/windows/mixed-reality/.
- [13] Susanna Nilsson and Björn Johansson. "Fun and usable: augmented reality instructions in a hospital setting". In: *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining user interfaces*. ACM. 2007, pp. 123–130.
- [14] Juri Platonov et al. "A mobile markerless AR system for maintenance and repair". In: *Mixed and Augmented Reality, 2006. ISMAR 2006. IEEE/ACM International Symposium on.* IEEE. 2006, pp. 105–108.

- [15] Michael Priestley. "DITA XML: A reuse by reference architecture for technical documentation". In: Proceedings of the 19th annual international conference on Computer documentation. ACM. 2001, pp. 152–156.
- [16] Dirk Reiners et al. "Augmented reality for construction tasks: Doorlock assembly". In: *Proc. IEEE and ACM IWAR* 98.1 (1998), pp. 31–46.
- [17] Arthur Tang et al. "Comparative effectiveness of augmented reality in object assembly". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. 2003, pp. 73–80.
- [18] Visometry. *visionLib*. June 2018. URL: https://visionlib.com/.
- [19] M Vagias Wade. "Likert-type scale response anchors". In: *Clemson (SC): Clemson International Institute for Tourism and Research Development, Department of Parks, Recreation and Tourism Management* (2006).

A Questionnaire forms

AR group

1) Participant Number

2) Please rate your experience level of the following subjects:

	Not at experi				remely ienced
Performing small mechanical and technical activities	0	0	0	0	0
Assembling/building a computer	0	0	0	0	0
Using Microsoft Hololens before the session started	0	0	0	0	0
Using gesture based interfaces (e.g. Nintendo Wii remote, Virtual Reality controllers)	0	0	0	0	0

 Please rate the following statements regarding your experience of using the AR Maintenance prototype for guiding you through the procedure you just completed:

	Stron disagi			St	rongly agree
I think that I would like to use this system	0	0	0	0	0
I found the system unnecessarily complex	0	0	0	0	0
I thought the system was easy to use	0	0	0	0	0
I think that I would need the support of a technical person to be able to use this system	0	0	0	0	0
I found the various functions in the system were well integrated	0	0	0	0	0
I thought there was too much inconsistency in this system	0	0	0	0	0
I would imagine that most people would learn to use this system very quickly	0	0	0	0	0
I found the system very cumbersome to use	0	0	0	0	0
I felt very confident using the system	0	0	0	0	0
I needed to learn a lot of things before I could get going with this system	0	0	0	0	0

Please rate the following information types that were available in the prototype based on how useful they were for you:

		ot at eful	all			remely useful
Textual information		0	0	0	0	0
Imagery		0	0	0	0	0
Augmented reality highlight rings		0	0	0	0	0

5) Please answer the following question regarding the task you just completed:

	Very difficu	ılt			Very easy
Overall, this task was	0	0	0	0	0

6) If you have any additional feedback, please provide it here:

Control group

1) Participant Number

2) Please rate your experience level of following subjects:

	Not at experi				emely ienced
Performing small mechanical and technical activities	0	0	0	0	0
Assembling/building a computer	0	0	0	0	0

 Please rate the following statements regarding your experience of using the technical documentation for guiding you through the procedure you just completed:

	Stron disagi			St	rongly agree
I think that I would like to use this system	0	0	0	0	0
I found the system unnecessarily complex	0	0	0	0	0
I thought the system was easy to use	0	0	0	0	0
I think that I would need the support of a technical person to be able to use this system	0	0	0	0	0
I found the various functions in the system were well integrated	0	0	0	0	0
I thought there was too much inconsistency in this system	0	0	0	0	0
I would imagine that most people would learn to use this system very quickly	0	0	0	0	0
I found the system very cumbersome to use	0	0	0	0	0
I felt very confident using the system	0	0	0	0	0
I needed to learn a lot of things before I could get going with this system	0	0	0	0	0

 Please rate the following information types that were available in the technical documentation based on how useful they were for you:

	Not at useful				emely useful
Textual information	0	0	0	0	0
Imagery	0	0	0	0	0

			Very difficu	ılt			Ver ea
Overall, this task was			0	0	0	0	0
6) If you have any additional feedback, ple	ase provide i	t here:					
of it you have any additional recuback, pre	use provide i	t nere.					