Towards Analytical Laboratories 4.0: Leveraging Augmented Reality

Teresa Figueiroa Teixeira de Abreu

Thesis to obtain the Master of Science Degree in

Mechanical Engineering

Supervisors: Prof. Susana Margarida da Silva Vieira
Eng. Andrea Costigliola

Examination Committee

Chairperson: Prof. Carlos Baptista Cardeira
Supervisor: Prof. Susana Margarida da Silva Vieira
Member of the Committee: Prof. Daniel Simões Lopes

September 2021

This work was created using LaTeX typesetting language
in the Overleaf environment (www.overleaf.com).

Acknowledgments

I would like to acknowledge my supervisors, Prof. Susana Vieira and Eng. Andrea Costigliola for all their support throughout the development of this dissertation.

I would like to thank Hovione Farmaciência S.A. for the opportunity to work in such a challenging and interesting project. I would also like to show my gratitude to Hovione's staff, to the Data Science & Digital Systems team, especially Eng. Rafel Lucas, to Sofia Silva and all the analysts that participated in the case study presented in this work.

A thank you to all my friends and family, for their support, friendship and encouragement. Especially to those that also contributed with their hands to this dissertation.

Lastly I would like to thank Instituto Superior Técnico and all the professors that crossed my path, for five years of learning and growth.

Abstract

The pharmaceutical industry has been facing challenges which have driven it to find new ways of increasing productivity and optimizing processes. To this end, pharmaceuticals are implementing the paradigm of Industry 4.0 with the name of Pharma 4.0. This strategy includes the implementation of technologies that aim at automating processes but also at increasing the worker's job performance. One of this technologies is Augmented Reality (AR), it augments the vision of the worker with information relevant to the task at hand. This thesis proposes an augmented reality application that aims at facilitating the worker's job while following the procedures in the analytical laboratory of a pharmaceutical company. The application shows procedures in a step-by-step guide which can be consulted using smart glasses (SG). Furthermore, in order to provide a hands-free solution, speech and gesture recognition were implemented successfully, providing a flexible way for the analysts to consult such procedures while performing the tasks. This thesis also presents a case study, following the technology acceptance model, in order to understand the value of such system in a pharmaceutical analytical laboratory. The results of the study show that such a technology can be very beneficial. The analysts reported that the tool was useful in their day to day lives. It was concluded that the application is a valuable addition to the laboratory, although many improvements can still be made.

Keywords

Pharma 4.0; Analytical laboratory; Augmented reality; Hands-free navigation; Technology acceptance model

Resumo

A indústria farmacêutica tem enfrentado desafios que a têm levado a encontrar novas formas de aumentar a produtividade e optimizar os processos. Para este fim, a indústria farmacêutica está a implementar o paradigma da Indústria 4.0 com o nome de *Pharma 4.0*. Esta estratégia inclui a implementação de tecnologias que visam automatizar processos, mas também aumentar o desempenho do trabalhador. Uma destas tecnologias é a Realidade Aumentada (RA), que aumenta a visão do trabalhador com informação relevante para a tarefa a ser executada. Esta tese propõe uma aplicação de realidade aumentada que visa facilitar o seguimento de procedimentos pelos trabalhadores no laboratório analítico de uma empresa farmacêutica. A aplicação mostra os procedimentos através de um guia passo-a-passo que pode ser consultado usando *smart glasses*. Além disso, para proporcionar uma solução que mantenha as mãos do trabalhador livres, foi implementado com sucesso reconhecimento de voz e de gestos para navegação da aplicação, proporcionando uma forma flexível de os analistas consultarem os procedimentos enquanto executam as tarefas. Esta tese apresenta também um estudo de caso, seguindo o modelo de aceitação da tecnologia, a fim de compreender o valor de tal sistema num laboratório analítico de uma farmacêutica. Os resultados mostram que esta tecnologia pode ser muito benéfica. Os analistas relataram que a ferramenta foi útil no seu dia-a-dia. Concluiu-se que a aplicação é uma adição valiosa ao laboratório, embora ainda possam ser feitas muitas melhorias.

Palavras Chave

Pharma 4.0; Laboratório analítico; Realidade aumentada; Navegação com mãos livres; Modelo de aceitação da tecnologia

Contents

List of Figures

List of Tables

Acronyms

AR Augmented Reality

ARSG Augmented Reality Smart Glasses

CM Confusion Matrix

EMA European Medicines Agency

FDA Food and Drug Administration

FOV Field of View

FP False Positive

FPN Feature Pyramid Network

GMP Good Manufacturing Practices

GLP Good Laboratory Practices

HMD Head Mounted Device

HHD Hand Held Device

I4.0 Industry 4.0

IAR Industrial Augmented Reality

ICH International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use

IND Investigational New Drug

ISPE International Society of Pharmaceutical Engineering

IoU Intersection over Union

KA	Knowledge Area
LRF	Local Receptive Field
mAP	Mean Average Precision
NDA	New Drug Application
PQS	Pharmaceutical Quality System
SG	Smart Glasses
SSD	Single Shot Detector
SSQ	Simulator Sickness Questionnaire
TAM	Technology Acceptance Model
TP	True Positive
VR	Virtual Reality

Chapter 1

Introduction

Recently, industry has been implementing new technologies in order to accommodate the changes brought by Industry 4.0 (I4.0). Several of these technologies have allowed for highly autonomous manufacturing systems, providing a greater productivity capacity and reducing human errors. Pharmaceutical companies have also been adopting this new model, under the name of Pharma 4.0, with the objective of increasing productivity and reducing the time-to-market of new drugs to overcome the challenges that have been affecting the industry in the past years.

Although Industry 4.0 technologies aim at making processes autonomous, there are still tasks that are impractical, or even impossible, to complete without human intervention. Many of the tasks in the pharmaceutical laboratories fall into this category, where although the workers are assisted by the machine, they still play a crucial role in completing necessary tasks. It is for this reason that one of the 9 pillars of the new Industrial Revolution is augmented reality (AR) [1]. Industry 5.0 complements the existing I4.0, but changes the main focus from digitalization for increased flexibility and efficiency, to sustainability and human-centered technologies. It aims at an increased collaboration between humans and smart systems, making augmented reality a core technology of this new paradigm [2]. Augmented Reality has the capability of closing the gap between the worker and the machine by bringing together the physical and virtual worlds. AR augments the senses of the worker by providing digital information relevant to the task at hand and showing it by superimposing virtual objects onto the real world, helping the worker to make better decisions, increase productivity, reduce errors and reduce the cognitive load of performing each task.

In this chapter, sections 1.1 and 1.2 will introduce the pharmaceutical industry and the Pharma 4.0 operating model, respectively. In section 1.3 the motivation, objectives and contributions of this work will be presented. Section 1.4 contains the outline of the dissertation.

1.1 The Pharmaceutical Industry

The pharmaceutical industry aims at discovering, developing, producing and marketing pharmaceutical drugs. Due to the nature of its product, the pharmaceutical industry is one of the most regulated industries today. Entities like the Food and Drug Administration (FDA), in the United States, European Medicines Agency (EMA), in Europe, and Infarmed, in Portugal, ensure that the drugs produced by these companies meet the necessary standards through the Good Manufacturing Practices (GMP) and Good Laboratory Practices (GLP) regulations. The pharmaceutical industry is thus driven by high standards of quality and safety in the research and manufacturing processes, in order to comply with the regulations and ensure the health of both consumers and workers. The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) is an organization that brings together the pharmaceutical companies and regulatory authorities to promote public health by discussing scientific and technical aspects of pharmaceuticals, and developing guidelines and requirements for pharmaceutical product registration and thus achieving a greater harmonization between this entities.

1.1.1 GMP and GLP regulations

The GMP and GLP ensure that every pharmaceutical company meets quality and safety measures systematically. In order to accomplish this, the product life cycle must be constantly monitored and is inspected by the regulatory entities on a regular basis. This two regulations are the minimum requirement for pharmaceutical companies to be able to market their products.

The GMP regulations ensure that the product follows pre-defined manufacturing criteria. This covers all the manufacturing processes, including equipments and facilities, governing every stage of the drug development life cycle.

The GLP regulations are designed to assure scientific data integrity and validity, they inspect the quality and reliability of laboratory tests.

1.1.2 Drug Development Stages

Before marketing, the drug undergoes several processes to be developed, research is conducted to ensure the drugs effectiveness in the cure of the targeted disease, as well as the consumers safety. Drug development is the process of creating a new pharmaceutical drug, from the discovery of a new compound up to the marketing of the drug. According to the FDA, there are 5 stages in this process [3].

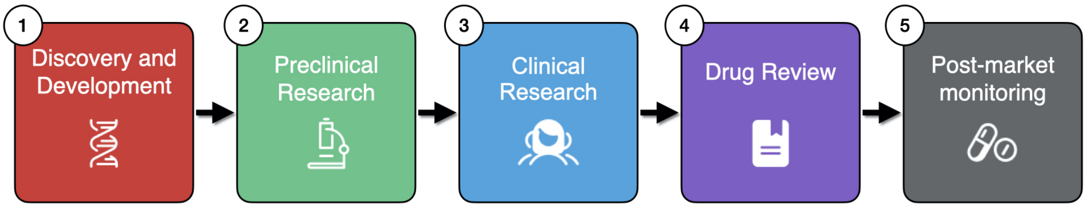

Figure 1.1: Drug development stages

1 **Discovery and Development:** The researchers identify new promising compounds and conduct detailed studies in order to evaluate benefits, mechanisms of action, dosage, effectiveness, etc.

2 **Preclinical Research:** Researchers must assess the toxicity of the drug, evaluating if the drug is safe for human testing, and decide whether the drug moves to the next stage of development. At this point the GLP regulations must be followed. An Investigational New Drug (IND) application must be sent and approved by regulatory authorities before beginning clinical trials.

3 **Clinical Research:** Clinical research refers to the trials that are done in humans. It has four phases, each has a bigger population than the previous one. Research questions and objectives for each phase must be outlined and strict protocols must be followed. If the clinical trial is successful and the drug shows good results for the purpose intended a New Drug Application (NDA) is submitted.

4 **Drug Review:** FDA reviews the NDA and gives the approval or rejection to the manufacturer.

5 **Post-market Safety Monitoring:** The product keeps being evaluated while in the market and changes to dosage and user information may be applied based on feedback received from the manufacturers, health professionals and consumers.

The drug development process is considered as the critical factor in determining the success of a pharmaceutical company, it is a process that can take up to 16 years and has an average cost of 1 billion USD [4]. Many of the potential compounds that are tested in stage 1 never make it to the market. Pharmaceuticals are trying to optimize this process in order to reduce costs and the time-to-market of new drugs.

1.2 Pharma 4.0

Pharmaceutical companies have been changing in the past few years due to increased competitiveness in emerging markets, shortage of patent lives and drug pricing laws. The necessity to increase productivity and optimize processes has led to the adoption of new strategies, namely the adoption of

Pharma 4.0. The International Society of Pharmaceutical Engineering (ISPE) describes Pharma 4.0 as an operating model that joins digitalization and ICH Q10 guidelines.

ICH Q10 is a quality guideline targeting the Pharmaceutical Quality System (PQS), applicable across the life cycle of the product, which complements the existing GMPs. It encourages the improvement of manufacturing processes by applying technical innovation, continual improvement policies, data monitoring and preventive action culture. Digitalization allows for a fully integrated value chain, where data can be continuously gathered in order to improve processes, help workers on decision making, and have autonomous systems. The combinations of these two concepts creates a powerful strategy that pharmaceuticals can follow in order to gain a competitive advantage in the market.

A study containing a survey to pharmaceuticals revealed that the principal areas of focus of companies when implementing Industry 4.0 included optimizing processes, monitoring plant performance, ensuring regulatory compliance, and minimize downtime. It also stated that one of the top Industry 4.0 elements being used is Augmented Reality, which can be a valuable asset in tackling any of the focus areas mentioned [4]. AR brings a new type of human-machine interface that allows the workforce to have a better performance in the new smart factory, providing information, assistance and guidance in a flexible way.

1.3 Motivation, Objectives and Contributions

The adoption of Augmented Reality systems in industry has been growing, especially in recent years [5]. The technology can help the worker to adapt in changing environments and to make the users job easier [6]. AR has been proving to be a valuable technology, not only in terms of productivity but also by improving the workers well being. Industry implementations include different sectors: automotive [7], aerospace, [8], shipyard [9], although most implementations in literature about the topic have been carried out in laboratory settings [5]. Reduction of the number of errors, decreasing of the execution time of tasks, increased learnability, facilitate decision making, and reduced cognitive workload are a few of the cited advantages of AR [5, 10–12].

1.3.1 Motivation

This dissertation is the product of a partnership between Instituto Superior Técnico and Hovione Farmaciência, S.A.. Pharmaceutical companies, such as Hovione, have many areas where procedures need to be carefully followed. Information is accessed using different tools such as paper, computers or tablets; and different formats such as manuals, documents, flowcharts etc. In Hovione's analytical laboratory, to perform a certain task, the analysts must follow a set of instructions represented by a flowchart, which is consulted using a small computer that the analysts must carry around. Some tasks are performed in workbenches while others are performed standing, in a fume hood, or using a specific machine, where there is no space available to put down objects. The workers in the lab require both hands to perform the necessary tasks. The way that analysts are currently accessing the flowcharts has a set of disadvantages:

- They have to switch their attention between the task being performed and the flowchart on the computer, that might even be far from the analysts if there is no workbench space
- Every time the analysts take their eyes off the flowchart and come back they have to find where they left off, which is cognitively expensive
- Since the analysts must interact with the computer using their hands, there is the need to stop the current task and put down any objects when an action on the computer is needed. While using the computer there is also the risk of cross contamination

These disadvantages manifest the need for improvement of these processes. Augmented reality can substitute the current way the flowcharts are consulted and provide a more convenient, fast, productive and overall easy way of going about the procedures in the laboratory. Therefore, this thesis proposes an Augmented Reality system that will be used to transfer these flowcharts from the workbench to the analyst camp of vision in a step-by-step guided application running on smart glasses (SG) that can be navigated hands-free, eliminating the problems stated above.

1.3.2 Objectives

The objectives of this thesis are:

- create an Augmented Reality application to revamp Hovione's flowchart procedure consultation in the analytical laboratory
- evaluate said application in order to assess if this type of system helps the analysts in their tasks, improving their productivity and facilitating their work, and is therefore a valuable asset

1.3.3 Contributions

This thesis contributions are:

- an augmented reality application that converts Hovione's flowchart procedures into step-by-step guides. The application retrieves the information directly from the company database, being automatically updated. It also provides the user some other auxiliary features that can further help with the completion of the procedure being performed

- the implementation and testing of speech recognition inside the application in order to provide hands-free navigation using an existing library and applying it to the problem at hand.
- the implementation and testing of gesture recognition to further improve the hands-free feature, applying object detection models to a new dataset.
- the evaluation of user acceptance of the augmented reality system by conducting a case study featuring Hovione's analysts.

The application developed in the context of this thesis will be implemented in Hovione's laboratories as a permanent tool to which all workers will have access.

1.4 Thesis Outline

Chapter 1 briefly introduces the Pharmaceutical Industry and Pharma 4.0 operating model, and describes the motivation, objective and contributions of this thesis. In **chapter** 2 the Augmented Reality technology, its benefits, challenges, applications in industry, and navigation solutions are described. The technology acceptance model applied to AR is also introduced. In **chapter** 3 the development of the application is presented, detailing all its components and features. Also the smart glasses used are briefly described and compared with other solutions in the market. **Chapter** 4 contains the description and results of the case study conducted to evaluate the users acceptance. Finally in **chapter** 5 conclusions and future work are presented.

Chapter 2

Augmented Reality in Industry

2.1 Definition and Brief History of Augmented Reality

Augmented reality (AR) is an emerging technology that connects the physical and virtual worlds. It enables the user to easily access relevant information in real-time while still being aware of the real environment. Unlike Virtual Reality (VR), that fully replaces the real world with a virtual one, AR can be used in settings where the real environment is relevant, or when not seeing the surroundings can even be dangerous. In the Reality-Virtuality Continuum displayed in figure 2.1, Virtual Reality is positioned at the end of the spectrum, demonstrating the fully virtual nature of the technology. Augmented Reality, however, is positioned more towards the middle of the spectrum indicating that this concept combines both virtual and real components. The fact that it is presented more toward the left of the spectrum means that it is based on the real environment and augments it with virtual objects. Augmented Virtuality, on the other hand, has a virtual base and is augmented with some amount of "reality" [13]. AR uses an electronic device, such as a smartphone, or head-mounted display (HMD), to superimpose virtual objects on the users field of view (FOV). This way of collecting information is more complete than simply seeing the information on a screen or piece of paper, since it is shown in the real-world context that it is related to.

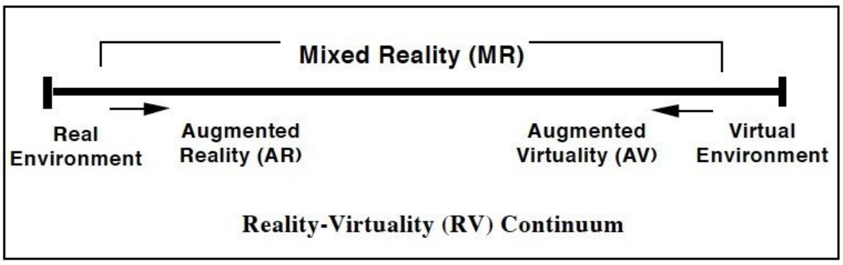

Figure 2.1: Reality-Virtuality Continuum [13]

The history of Augmented Reality goes back to 1968, when Ivan Sutherland created the first head-mounted display [14]. 10 years later, in 1978, Gavan Lintern and Stenley Roscoe used a head-up display (HUD) for flight training showing the benefits the technology could bring for the first time [15]. But it was only in 1992 that the term “Augmented Reality” appeared, coined by Caudell & Mizell. They described AR as a technology that “is used to augment the visual field of the user with information necessary in the performance of the current task” [16]. That same year, Louis Rosenberg introduced the first fully functioning AR system, *Virtual Fixtures*: “Like the ruler guiding the pencil, virtual fixtures overlaid on top of a remote workspace could act to reduce mental processing required to perform the task, reduce the work load of certain sensory modalities, and most of all allow precision and performance to exceed natural human abilities.” [17]. These last 3 works showed the potential of AR for the industry sector, advocating that it would be a technology that would improve the workers’ performance.

Although the concept has been around for some time, AR has only began to be more greatly implemented in the past few years, enabled by the leaps in miniaturization. The appearance of smartphones had a great impact in the technology since they allow for considerable computational power in anyone’s pocket. The game *Pokemon Go* is a good example of the mainstreaming of AR due to the usage of smartphones, it was one of the greatest hits of AR, being one of the first applications to become known worldwide. The advances in technology also allowed for the recent commercialization of different head-mounted displays that further allow for the implementation of the technology in industry.

2.2 Types of Augmented Reality

Augmented Reality classification can follow under one of the two categories: view-based and triggered. Triggered AR uses a tracking system to trigger an action, view-based AR is more simple, referring to the augmentation of static views or augmentation of views that do not interact with the real environment [18]. Figure 2.2 shows the different types of AR.

Under view-based AR, indirect augmentation refers to the augmentation of a static image, for example, taking a picture of a room and then change the color of the wall (figure 2.3). Non-specific digital augmentation allows the user to interact with the augmentation, but the augmentation does not interact with the real-world, the application developed in the context of this thesis falls under this category.

The tracking system in AR allows the application to intake the real world in order to interact with it, the different types of triggered AR are divided by the tracking system.

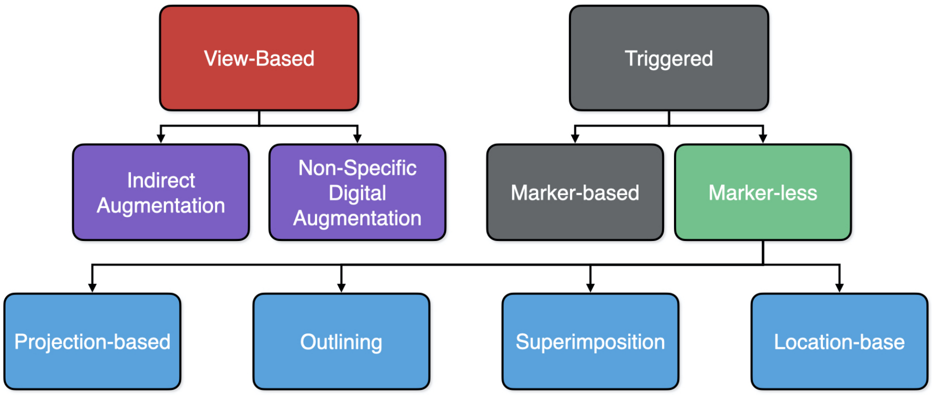

Figure 2.2: Different types of Augmented Reality

Marker-based AR reads a QR code or bar code and reacts accordingly (figure 2.4). Location-based AR uses GPS data to connect to the real world. The already mentioned game *Pokemon Go* uses this type of tracking system - each type of *Pokemon* appears in different parts of a the world map (figure 2.5). Projection-based AR works as a type of hologram, the virtual objects are projected onto a surface and the user can then interact with them (for example projected keyboards - figure 2.6). Superimposition AR recognizes certain items of the real world. An example of this type of AR is to study anatomy, the system recognizes a body part and helps the user identify its components (figure 2.7). Outlining utilizes shapes and lines, it can be used to highlight the road lines when driving with low light (figure 2.8).

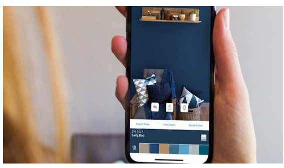

Figure 2.3: Indirect AR

Figure 2.4: Marker-based AR

Figure 2.5: Location-based AR

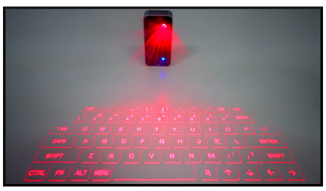

Figure 2.6: Projection-based AR

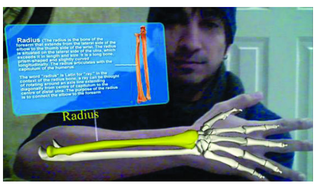

Figure 2.7: Superimposition AR

Figure 2.8: Outlining AR

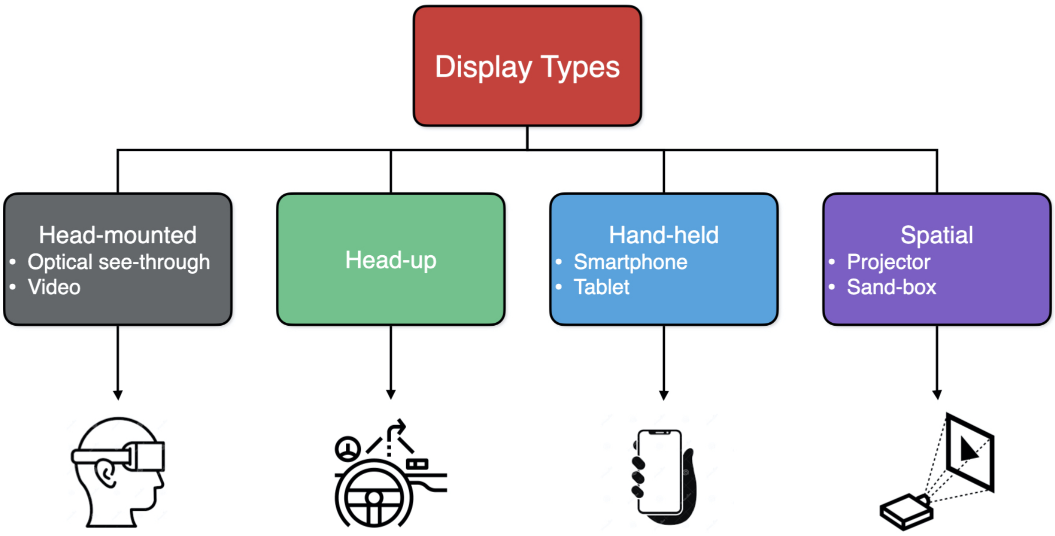

Figure 2.9: Different types of AR displays

Augmented Reality can also be categorized by the type of display that is used. These are presented in figure 2.9. The most commonly used type is Hand-held AR, because it is easily accessible to anyone. A person with a smartphone or tablet can download an AR application and start using it straight away. In industry, however, the preference is for Head-mounted devices, since these allow for hands-free interaction and a more immersive experience [10]. There are also some applications in industry that use spatial AR, but these are not as common since this type of device is usually fixed to a certain place, making it less flexible [19]. Head-up AR is commonly used in vehicle windshields and airplane navigation.

2.3 Advantages and Challenges of Augmented Reality in Industry

In order to better understand the potential of AR technology in industry and also why it is still not being largely adopted, some benefits and challenges are presented in this section, these where gathered from different sources ([10], [5], [12]):

- Decrease in number of errors
- Improved quality
- Reduction of execution time
- Increased learnability
- Improved flexibility

- Facilitate decision making

All of the stated advantages can bring reduction of costs, associated with the fact that they increase productivity. Some of these advantages are correlated: by reducing the number of errors a worker makes while executing a task, the execution time of that same task will be reduced and the quality increased. Furthermore, reducing the number of errors may lead to less losses of parts and/or products and, consequently, savings in material costs. AR is also commonly used for training purposes, because it enables autonomous training systems, where there is no need for a supervisor, and also it can provide a more interactive and immersive way of presenting content, which leads to increased learnability. Other less common advantages include increased health, by the reduction of the cognitive load the worker has while performing the task, and increased long term memory retention [20], which further advocates for AR as a training tool.

On the other hand, AR is a relatively new technology and still needs some improvements. The following challenges were found in literature ([10], [21], [19]):

- Projection quality, accuracy and interaction
- Hardware
- Tracking methods
- User's health and acceptance
- Development complexity
- Creating natural applications and tools

Problems related to AR hardware are amongst the most common. Although smartphones already have a good computation capacity, for some applications it is still not enough. The same applies regarding head-mounted displays, a balance has to be reached between the hardware quality and the comfort of the user. Headsets can not be too heavy, or they might cause injuries when used for long periods of time, the display must be seen clearly and the navigation should be natural in order to provide a good user experience. All these aspects need to be improved in order to lead to a higher user acceptance and consequent mainstreaming of AR in industry as well as in other sectors.

2.4 Applications of Augmented Reality in Industry

Augmented Reality is currently used for different industry applications, the most common ones being maintenance, assembly and training [22]. Since AR is still a growing technology, most implementations in literature about the topic have been carried out in laboratory settings [5]. In this section some of these applications are described as well as their conclusions.

Maintenance is a common field in AR applications since it can greatly benefit from the technology. Some motivations for the implementation of AR in the maintenance field are:

1. help reducing human-induced errors while performing maintenance tasks, especially on aircrafts and other machines where failure can bring dire consequences;
2. the need to have a flexible and mobile support, since maintenance tasks are mobile activities;
3. the need to reduce costs associated with maintenance tasks

One of the ways AR is being applied in the maintenance field is by supporting remote maintenance. Remote maintenance is the performance of maintenance tasks when the expert is not present on-site. Using AR, the expert can guide the operator in the procedure. Mourtzis et al. [21] proposed a remote maintenance system featuring AR and video conference, where the expert would guide the operator in the procedure through a computer. The operator would use an HMD or mobile device to stream his/her view and get instructions. The work concluded that this type of maintenance approach would lead to reduction of costs and decrease in the machines downtime, thereby increasing productivity. A similar approach was proposed by Vorrabet et al. [23], this work presented a comparison between remote assistance using a simple phone-call and using an HMD with the capability of interchanging video and annotations between the two parts (operator and maintenance expert). The result of the experiment concluded that the AR approach decreased the completion time of the tasks and improved efficiency.

Assembly/disassembly tasks can benefit greatly from AR, these tasks usually involve procedures that can be converted into step-by-step guides. Hou et al. [24] conducted a study where an AR system was tested to guide workers in a piping assembly, results showed that the completion time of the task were reduced when compared to conventional methods, a decrease in the number of errors and in the cognitive workload were also observed. Mura [7] developed an AR guide for the support of panel alignment in car body assembly. By comparing the AR system with the conventional method, it was noted that the assembly task, when performed with the AR tool, was completed almost four times faster. Makris [25] proposed a process to automatically generate assembly instructions for Augmented Reality using CAD. They observed that by applying this method, both operator and engineers would benefit from reduced cognitive load, also the time between the product design and its production would be shortened.

Companies have to spend time and resources providing training sessions for their workers. AR has the capability of making this process more efficient. Training AR applications in industry are often related to maintenance and/or assembly tasks. Macchiarella [20] conducted a study where four types of training methods, applied to a maintenance task, were compared: a video-based presentation, a print-based presentation, an augmented reality presentation and an interactive augmented reality presentation. It was concluded that the augmented reality based instructions lead to an increased long term memory retention. Gavish et al. [26] evaluated virtual and augmented reality platforms for training of industrial maintenance and assembly (IMA) tasks. The study showed that AR helps the worker reduce the number of errors while performing a task, although it also stated that the participants in the experiment took longer to train with the AR and VR platforms than with conventional methods, which can be due to the lack of exposure to these technologies. Hahn et al. [27] proposed a system to teach operators the assembly process of printed circuit boards. The study resulted in an errorless performance of all participants, many of which stated that the system helped them obtain such results and would appreciate a permanent deployment of the system in the production.

Augmented reality applications in the pharmaceutical industry are very scarce in literature, one application was found presented by Forrest et al. [28], which proposed a remote assistance system using Microsoft HoloLens headset similar to the remote maintenance described at the beginning of this section. The authors propose to use the same type of video conference to transfer experimental methods across multiple pharmaceutical laboratories, eliminating the sharing of this type of information through written protocols that many times lead to error and resulted in inefficient information transferring. The solution could reduce the travel time of the scientists between laboratories and associated costs, and reduce the overall drug development time.

Having into account that pharmaceuticals have a unique way of operating, solutions that fit other industries may not be feasible for a pharmaceutical application, therefore it is important to increase research on this topic to fill the current gap in literature regarding AR in the pharmaceutical industry.

2.5 Navigation for Augmented Reality Applications

The most used AR devices in industry applications are head-mounted displays (or smart glasses). This type of device allow the user to be hands-free while performing tasks in the shop-floor, depending on the type of navigation they use. Two common options for hands-free navigation are speech recognition and gesture recognition. These types of navigation will be introduced in this section.

2.5.1 Speech Recognition

The most common navigation control strategy for augmented reality applications is speech recognition (SR) [29]. Most AR devices have this feature included. For devices that do not include this feature, it can be implemented. There are libraries that provide speech recognition models that can be used to integrate SR in AR applications.

This type of navigation is natural to use and easy to implement using the existing libraries. Nevertheless it can be impractical to use in environments where background noise is present.

2.5.2 Gesture Recognition

Implementing hand-gesture control on smart glasses can provide an easy and flexible interaction approach [30]. Smart glasses have an integrated camera that can be used to implement vision-based gesture recognition using computer vision techniques. Other implementations of gesture recognition include using external objects, such as a glove, to capture the gestures using sensors (these are not in the scope of this work).

Computer vision is a field of computer science that aims at processing images and videos in order to extract relevant data. A common application of computer vision is recognition, which analyzes if a certain image contains some specific objects. Convolutional neural networks (CNN) are a type of artificial neural networks that are commonly applied to image classification and recognition problems. In this section the basis of CNNs functioning is explained.

Articial Neural Networks

An artificial neural network is an algorithm that attempts at imitating the functioning of the human brain. These algorithms are composed of interconnected processing elements, called neurons or units. ANNs have the ability to ”learn” by ”seeing” data. Unlike traditional algorithms, where the user sets the rules and inputs the data in order to obtain results, ANNs take in labeled data, this is the input and the output, in order to learn the rules.

The **neuron**, represented in figure 2.10 is the ”unit cell” of an ANN.

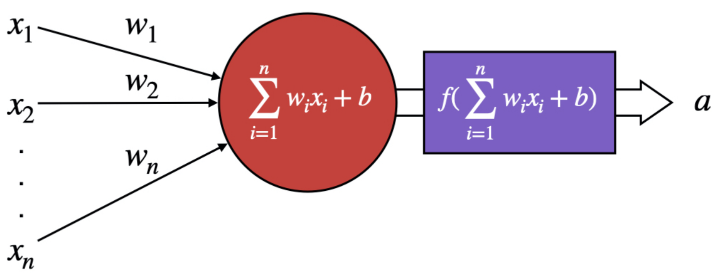

Figure 2.10: Artificial neural network neuron

The neuron takes in the inputs $[x_1, x_2, ..., x_n]$ and produces an output a by applying a function f to the weighted sum of the inputs plus a bias b. Function f is called an activation function, which is typically non-linear, since most real-world data is non-linear.

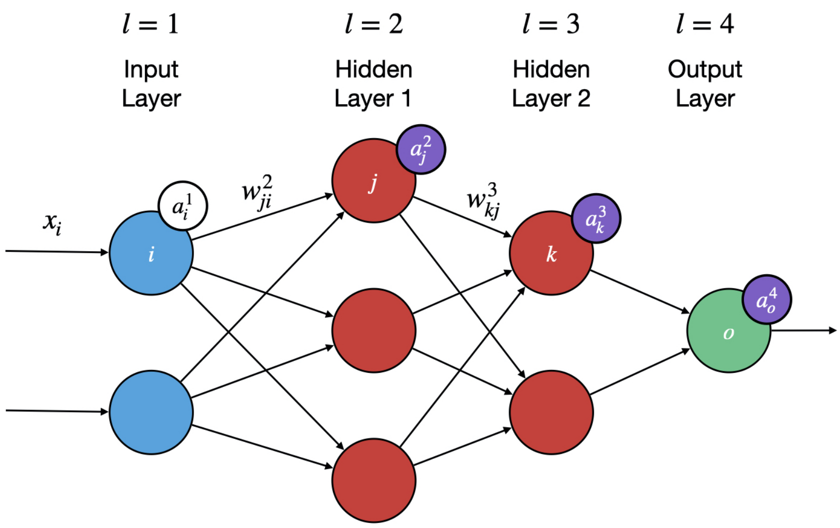

Figure 2.11: Example of an artificial neural network

Figure 2.11 represents an example of an artificial neural network with 4 layers: one input layer, two hidden layers, and an output layer. Neural networks can have many different architectures. In a ANN, the outputs of one layer are fed as inputs to the next layer until a final output is given by the final layer. This process is called forward propagation. For each neuron equation 2.1 is applied, except for the neurons in the first layer, $l = 1$, which simply represent the input: $a_i^l = x_i$.

$$a_j^l = f\left(\sum_i w_{ji}^l a_i^{l-1} + b_j^l\right) \tag{2.1}$$

where,

- a_j^l is the activation of the j^{th} neuron in the l^{th} layer
- w_{ji}^l is the weight that goes from the i^{th} neuron in the $(l-1)^{th}$ layer, to the j^{th} neuron in the l^{th} layer.
- a_i^{l-1} is the activation of the i^{th} neuron in the $(l-1)^{th}$ layer
- b_j^l is the bias of the j^{th} neuron in the l^{th} layer

The **training**, or learning, of a neural network consists of finding the right weights and biases in order for the network to compute the correct output for a given input. The process of training involves an optimization algorithm, gradient descent, and the application of backpropagation.

Gradient descent is an optimizer that minimizes a cost (or loss) function. The cost function is a measure of ”how bad” the network is categorizing the data. A commonly used cost function is called cross-entropy, given in equation 2.2.

$$C = -\frac{1}{n}\sum_x [y \ln a + (1-y)\ln(1-a)] \tag{2.2}$$

Where

- y is the desired output for input x
- a is the ANN output for input x
- n is the total number of training data

The goal in training is to find the weights and biases that minimize the cost function. These are found applying the gradient descent algorithm, which is to apply a small change to the weights and biases by adding a multiple of the negative of the gradient, as shown in equations 2.3 and 2.4

$$w^l_{ji} \rightarrow w^{l\prime}_{ji} = w^l_{ji} - \eta \frac{\partial C}{\partial w^l_{ji}} \tag{2.3}$$

$$b_j \rightarrow b'_j = b_j - \eta \frac{\partial C}{\partial b_j} \tag{2.4}$$

where η is the learning rate, a parameter that determines how fast the network learns.

By applying this repeatedly, we will arrive at a local minimum of the cost function. The partial derivatives of the cost function are computed using **backpropagation**. During the training process each training example is fed to the network. First the forward propagation occurs giving a certain output, afterwards the cost is computed based on the output and the ground truth, finally the error is propagated backwards in order to apply the changes to the weights and biases using gradient descent.

Let z^l_j be the weighted input to the activation function for neuron j in layer l, such that:

$$z^l_j = \sum_i w^l_{ji} a^{l-1}_i + b^l_j \tag{2.5}$$

$$a^l_j = f(\sum_i w^l_{ji} a^{l-1}_i + b^l_j) = f(z^l_j) \tag{2.6}$$

The gradients of the cost function relative to a single weight and bias (for a single training example) are given in equations 2.7 and 2.8, respectively, by applying the chain rule.

$$\frac{\partial C}{\partial w^l_{ji}} = \frac{\partial C}{\partial a^l_j} \frac{\partial a^l_j}{\partial z^l_j} \frac{\partial z^l_j}{\partial w^l_{ji}} \tag{2.7}$$

$$\frac{\partial C}{\partial b^l_j} = \frac{\partial C}{\partial a^l_j} \frac{\partial a^l_j}{\partial z^l_j} \frac{\partial z^l_j}{\partial b^l_j} \tag{2.8}$$

The equal part in both equations is commonly called the local gradient.

$$\delta_j^l = \frac{\partial C}{\partial a_j^l}\frac{\partial a_j^l}{\partial z_j^l} = \frac{\partial C}{\partial z_j^l} \tag{2.9}$$

These derivatives are first computed for the final layer and are then computed across the network adjusting the weights and biases according to equations 2.3 and 2.4.

The training process usually occurs in batches, a sub-set of training examples is fed to the network, for each training example the forward propagation, error and backpropagation computations are made, the gradient descent algorithm is applied to the batch, this is called stochastic gradient descent.

Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a specific type of artificial neural network that is commonly used in image classification. The input to this networks is an image represented by a three dimensional tensor, where the width and height represent the resolution of the image, and the depth the color channels. The values inside the tensor are the pixel values for each color channel.CNNs have three basic concepts: local receptive fields, shared weights, and pooling.

Local receptive fields: Figure 2.12 shows the connections between the input and the neurons in the first hidden layer. Unlike the first example (figure 2.11) where all neurons from one layer were connected to all neurons in the next layer, in CNNs the connections are between a "patch" of image, called the local receptive field (LRF), and a neuron in the next layer. The LRF is then slid across the entire image creating a feature map.

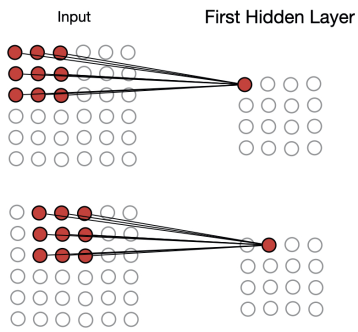

Figure 2.12: Feature map extraction

Shared weights and biases: The weights in the connections for creating each feature map are the same and represent a filter, or kernel. The filters aim at extracting specific features from the image (e.g.: lines, edges). Filters are represented by matrices, equation 2.10 shows an example of a filter with size 3×3.

$$\begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \tag{2.10}$$

The convolution operation is given in equation 2.11 and occurs between each LRF and the filter, as represented in figure 2.12.

$$P * W = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} * \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} = \sum_{m=1}^{3} \sum_{n=1}^{3} p_{m,n} \times w_{m,n} \tag{2.11}$$

where, P is the LRF, which is represented by a matrix with pixel values, and W is the filter.

The output for one feature map is given by equation 2.12

$$a_{i,j}^{l} = f\left(b + \sum_{m=0}^{2} \sum_{n=0}^{2} w_{m,n} a_{i+m,j+n}^{l-1}\right) \tag{2.12}$$

where,

- a_{ij}^{l} is the activation for neuron (i, j) in the feature map
- f is the activation function
- b is the bias
- $w_{m,n}$ is the weight at position (m, n) in the filter, in this case the filter has size 3×3, therefore m and n range between 0 and 2
- $a_{i+m,j+n}^{l-1}$ is the activation from the previous layer, which is the value at position $(i + m, j + n)$ on the feature map that is the input to the current layer

In a CNN layer, there may be many filters applied to the input, creating an output of many different feature maps.

Pooling layers: After a convolutional layer (where a filter is applied to the input), CNNs usually have a pooling layer. The pooling layer down-samples each feature map in order to reduce its size, outputting a condensed feature map. This is beneficial because it reduces computational overhead and helps to reduce overfitting. A common approach is to take a patch of 2×2 and retrieve the maximum value (figure 2.13).

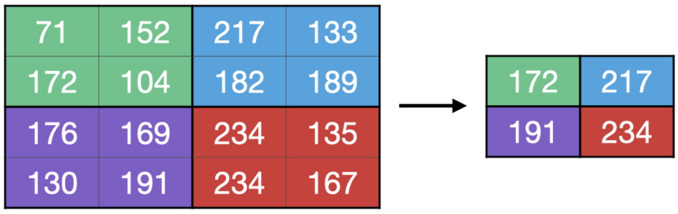

Figure 2.13: Max-pooling example

A CNN is composed of many convolutional and pooling layers. The first layers extract simple features, such as lines or edges. As the network gets deeper, it is able to detect more complex patterns, such as objects. At the end of the network there is usually one or more fully connected layers that make the high-level reasoning in the neural network. After this layer the output vector is fed to a number of neurons, each representing a class, which output the probability of the input image representing that class.

Transfer Learning: Transfer learning is a technique that allows to apply CNNs models trained in large datasets to smaller datasets. This is done by freezing the initial layers of the network, whose weights have already been trained to extract low-level features, and train only the final layers to adjust to the more specific features of the new dataset.

Object Recognition

Object recognition, or object detection, is the task of localizing objects in an image and classifying them. CNNs are used to solve this type of task. In this case, the input to the network are images and the labels to this images are a set of boxes (location of the object) and the classification of each box, as shown in figure 2.14.

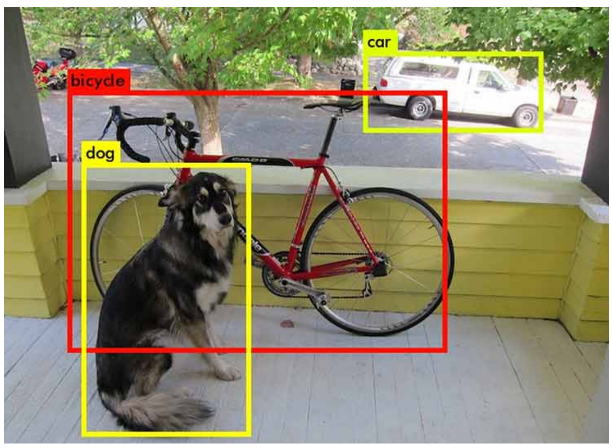

Figure 2.14: Example of data in an object recognition task

There are two object recognition model types: one-stage and two-stage. One-stage detectors adopt a fully convolutional architecture outputting classification probabilities and box offsets. Two-stage detectors first filter the regions that are more likely to contain objects (region proposals) and then feed them to the convolutional network. Two-stage detectors are generally more accurate, but one-stage detectors are much faster, making them viable for mobile applications.

A commonly used type of one-stage detector is the Single Shot Detector (SSD). The SSD is an object detection architecture that has 3 stages: 1) feature extraction, 2) detection heads and 3) non-maximum suppression.

1. The first stage extracts features from the input image using convolution layers and outputs several feature maps with a variety of sizes and number of channels. This stage uses the basis of a classification network.

2. The second stage consists of convolution layers that produce the boxes and classes predictions. The detector tries to identify which default boxes correspond to an object. For each default box, the scores for each class are computed, if a class scores higher than the background class, the box is classified as a positive detection, otherwise it is classified as a negative detection. Furthermore, the overlap of the default boxes with the ground truth boxes is measured, using intersection over union (IoU), which is to divide the area of overlap of the boxes by the union of both areas, with a threshold value usually close to 0.5. Default boxes that have a greater IoU than the threshold with the ground truth are considered positive. Other boxes may be considered negative or may be ignored, depending on the IoU value. The cost (or loss) in object detection architectures is divided into localization cost and classification cost.

3. The last stage is the post-processing stage which uses Non-Maximum Suppression to discard multiple detections of a single object. For boxes with a IoU greater than the threshold, the box with the lower confidence is discarded and the other is kept, this is done until there is only one box per detection (no boxes with the same class have a IoU greater than the threshold with each other).

2.6 Evaluation of Augmented Reality Applications

The Technology Acceptance Model (TAM) has been the most used model for the evaluation of technology acceptance since it was first introduced by Fred Davis in the 1980s. Davis based this model in two user perceptions, usefulness and ease of use, arguing that these two perceptions were the ones that mostly correlated with the acceptance of a new technology. He stated that if people believe the system would help them perform their job better, there would be a higher chance of using the system (perceived usefulness). Also, even if the system is useful, it cannot be so hard to use that the effort of using the system outweighs its benefits (perceived ease of use) [31].

The original TAM aimed at evaluating how these two variables affect attitude toward using and, consequentially, behavioral intention to use the technology (fig. 2.15). The model is evaluated by using the correlation between the variables scores, which are collected by using a questionnaire filled by people that experimented with the technology. Since F. Davis publication, several studies have used this method to predict user acceptance and others have extended the model to include other variables.

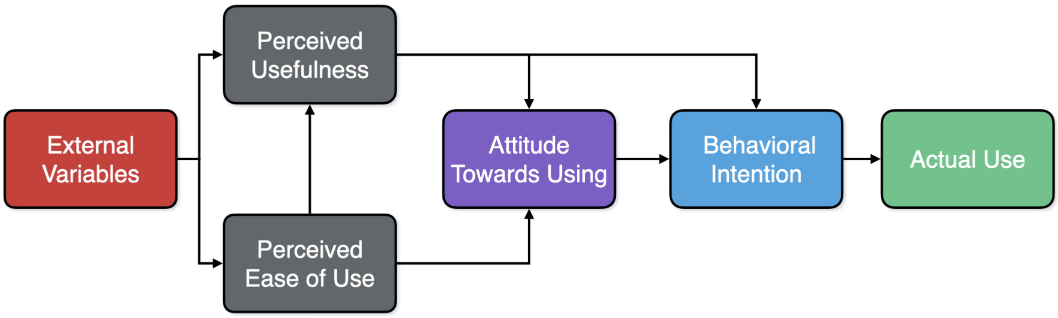

Figure 2.15: Original technology acceptance model

TAM has been used to evaluate the acceptance of augmented and virtual reality technologies. A variable that is often included in these models is perceived enjoyment, which is defined as ”the extent to which the activity of using a specific system is perceived to be enjoyable in its own right, aside from any performance consequences resulting from system use” [32]. Perceived enjoyment can be correlated with perceived ease of use and with intention to use [33].

A second variable used with augmented and virtual reality technologies is personal innovativeness (also sometimes called perceived innovativeness). This variable was proposed in 1998 by Agarwal Ritu and Jayesh Prasad [34]. They defined it as ”the willingness of an individual to try out any new information technologies”. The authors noted that a high perceived innovativeness led to a more likely adoption behavior. Other studies [35, 36] have also concluded that perceived innovativeness has a positive effect in perceived usefulness and perceived ease of use.

Yusoff et al. [36] proposed an evaluation of user acceptance of mixed reality using the TAM. Besides the original variables, the authors included two other variables: perceived innovativeness and perceived enjoyment. Holdack et al. [33] researched how perceived innovativeness and perceived enjoyment impacted the acceptance of augmented reality wearables. They concluded that these two variables positively influence intention to use. Vrellis et al. Sagnier et al. [37] extended the technology acceptance model to evaluate virtual reality acceptance, in their model they included personal innovativeness, cybersickness and presence, among others. Presence is ”a subjective sensation which allows the user to interact with and feel connected to a world outside of themselves”, being an important variable to evaluate virtual reality and some augmented reality applications that intend for users to feel like their inside a different reality.

Cybersickness is the negative side effect that might afflict a person when exposed to a virtual environment. It has similar effects as motion sickness, such as nausea, headaches and dizziness. Cybersickness symptoms can be more or less severe depending on the rendering modes, visual display and application design, they also affect people differently. They can impact the users comfort and, more importantly, their health, having the possibility of inflicting injury or decreased capacity [38]. Being a negative outcome of the exposure to virtual environments, this side effect might influence the acceptance of users to augmented reality systems, therefore this variable is sometimes included in the TAM to assess how it influences user acceptance. The Simulator Sickness Questionnaire (SSQ) [39] is often used to evaluate the symptoms due to simulation exposure.

Chapter 3

Augmented Reality Application for the Analytical Laboratory

In this chapter the development of the application and its navigation, as well as the smart glasses used are described. The objective of the Augmented Reality application is to guide the analyst through the laboratory procedures, in an natural and easy manner, concentrating different information in one place, and allowing all this to happen hands-free, in order to increase the workers productivity and reduce cognitive load.

Hovione has an internal knowledge management tool - Excellent Development and Manufacturing (EDaM) - designed to be a shared knowledge center where R&D scientists can create workflows for members to consult. In the analytical laboratory it is common for the analysts to consult these workflows to follow procedures. This tool is web-based, which makes it real-time updated. It is directly from this knowledge center that the information used in the application is retrieved.

The application allows the user to follow the procedures in a step-by-step guide, consult different types of information and supports hands-free navigation. This chapter is organized as follows: section 3.1 presents the smart glasses used; section 3.2.1 shows the user interface of the application; section 3.2.2 explains how the information is retrieved from the database using an Application Programming Interface (API); an explanation of how the step-by-step guide is constructed can be found in section 3.2.3; section 3.2.4 describes additional information and resources the user can access in order to further help with the task at hand; hands-free navigation, namely, speech recognition and gesture recognition are presented in sections 3.3.1 and 3.3.2, respectively.

The information flow between the different software components is presented in figure 3.1 and a diagram representing a simple flowchart of the application in figure 3.2.

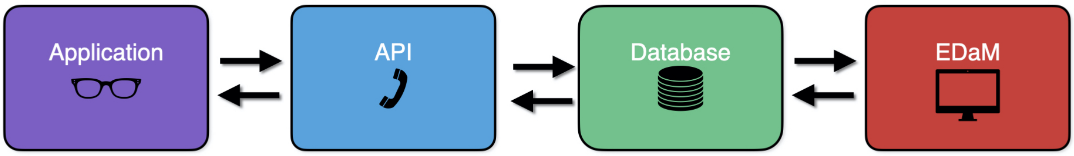

Figure 3.1: Information flow between the different software components

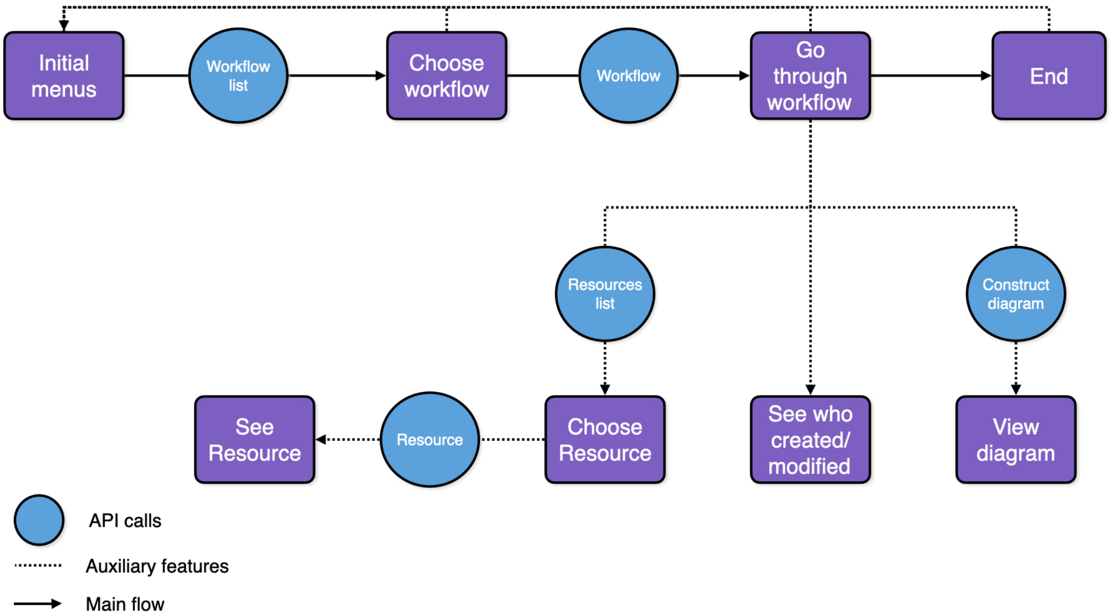

Figure 3.2: Flowchart of the application

3.1 Augmented Reality Device

The Augmented Reality device used was the Moverio BT-300 smart glasses, which were the ones chosen by Hovione for the development of this project. This section provides a short insight of these smart glasses, relevant technical specifications for the current work, and compares them to other devices available on the market. Section 4.2 shows the response of the users about some characteristics of these smart glasses, namely if the glasses adjust well to the users head, if they are comfortable to wear and if the display is seen clearly.

Display

The Moverio BT-300 smart glasses include an optical see-through display with a field of view (FOV) of approximately 23° (figure 3.3) [40].

The FOV measures how big the augmented reality image is when viewed by the user. A larger FOV means that the user will see a larger image. This is important not only because a larger image allows for the display of more information, but also because the augmentation layer will increasingly match the real layer, making the Augmented Reality experience more immersive. On the contrary, a small field of view gives the user the sense of having a screen in his camp of vision, rather than having virtual objects superimposed in the real world. Other smart glasses in the market have FOVs up to 90°. Microsoft Hololens 2 and Magic Leap 1, two commonly used HMDs in industry, provide about 50° FOV. Fraga-Lamas et al. [9] state that the field of view should not be lower than 30° in order to have a good user experience, so one can note that the FOV of the Moverio BT-300 is somewhat low, however, for the purpose of this thesis, it is sufficient.

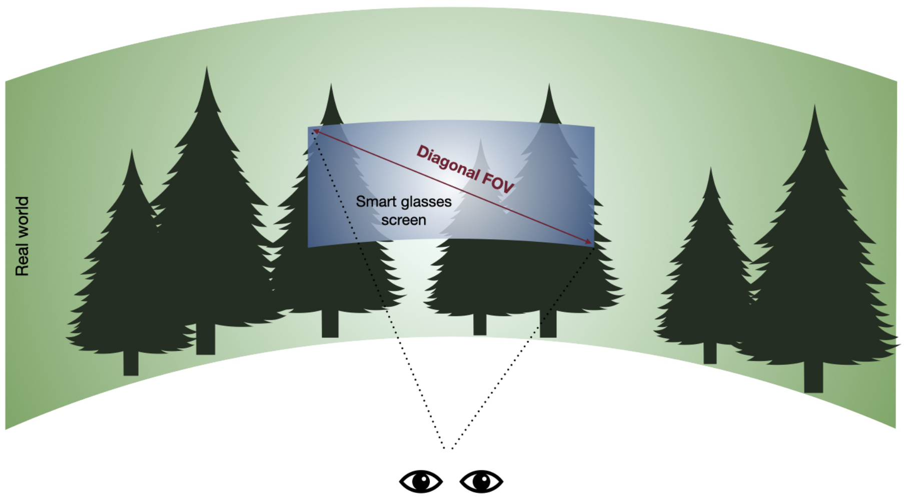

Figure 3.3: Optical see-trough smart glasses display and diagonal field of view

Operating System

The smart glasses use Android 5.1 (Lollipop) operating system, API level 22. A clear advantage of using Android is that by developing an augmented reality application for this headset one could also deploy it to other Android devices, such as tablets or even other smart glasses. Since the operative system is Android, the platform used for the development of the application was Android Studio and the programming language was Java.

Augmented Reality Device Controller

The user navigates the interface by using a controller (fig 3.4). The controller navigation is similar to that of an Android smartphone, making it easy and natural to use. The headset also incorporates a sensor that detects touch, this is a common approach of other smart glasses in the market, for example, the Google Glasses, which use this type of input as well as speech recognition.

There are many advantages of having a controller: it is easy to learn and use and it allows for more accurate input, unlike speech or gesture recognition, for example, which have an associated error and the input could be compromised in certain environments. Additionally, the controller does not discard these options, since the developer can implement this type of control.

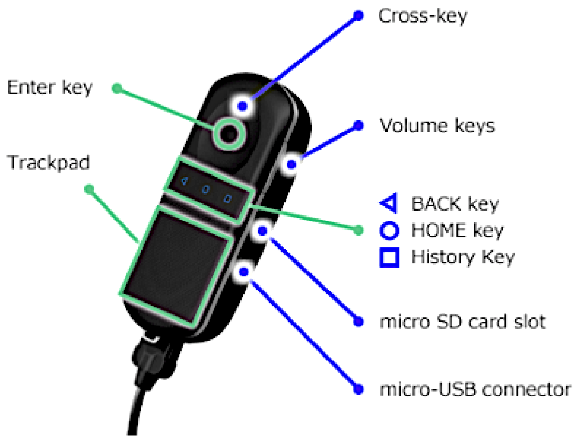

Figure 3.4: Epson Moverio BT-300 controller [40]

Other features

Moverio BT-300 are currently the lightest binocular, see-through smart glasses on the market. Lightweight is a valuable characteristic in smart glasses, since these are to be worn for long periods of time [9]. Heavy glasses are uncomfortable and can cause injuries to the worker.

The battery life of the Moverio BT-300 lasts up to 6 hours, which is above average compared to other smart glasses ($\simeq$ 4 hours) [9]. Magic Leap 1 SG have a battery life of 3,5 hours, and the Microsoft Hololens 2 have 2 to 3 hours of active battery life.

The Moverio glasses support microphone and headphones connection through an audio jack. They also have some integrated sensors that can be consulted in [40], but none were used in the context of this thesis.

3.2 Application's software features and components

3.2.1 User Interface

The user interface (UI) allows the user to view and interact with the application. The UI was made as simple as possible, not only for it to be natural, but also with the objective of not cluttering the view of the user, since the display will be overlaid with the real environment. Figure 3.5 shows the type of views present in the application.

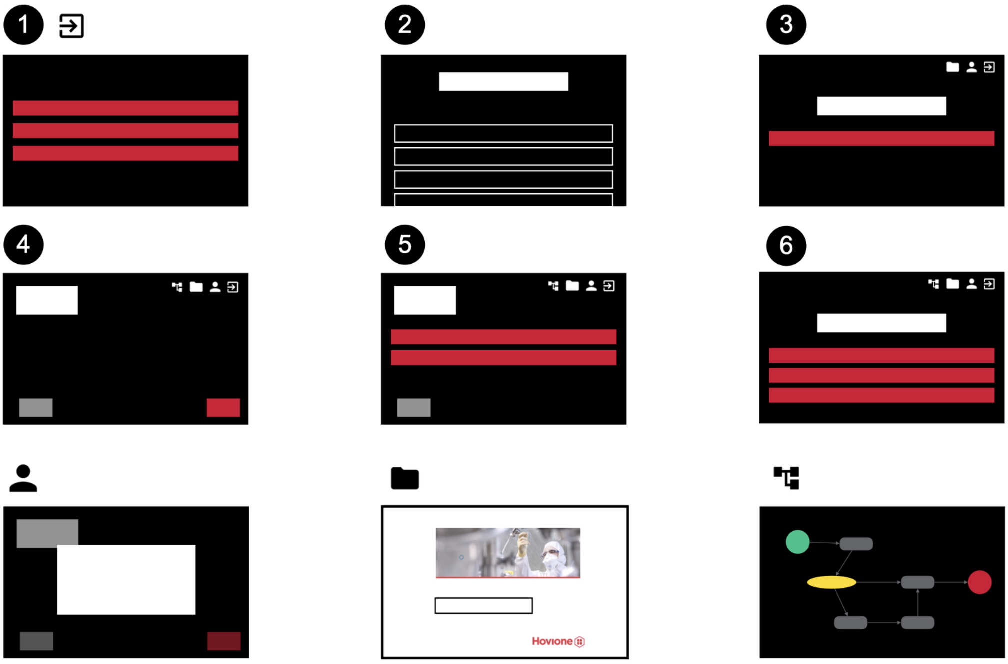

Figure 3.5: User Interface

Most views have a black background that in the glasses is not seen - the black is equivalent to transparent. The red rectangles represent buttons: these are made to stand out, in order for the user to be aware that an action is needed. View 1 represents the initial menus of the application, the symbol to the right of the number means that it is to this view that the user comes when this option is clicked on the options menu. View number 2 is a list view, the white block represents a title and the other blocks are the items of the list that the user can choose from. In view 4, a step of the workflow is represented, the white block shows the task to be made, the gray an red buttons have the options to go back or to the next step, respectively. Figure 3.6 shows an image of what view 4 would look like from the user's perspective. View 5 represents a question in the workflow, the options are placed in the red buttons. Views 3 and 6 are respectively the start and the end of a certain workflow. The last row of views shows the additional features explained in section 3.2.4 that can be accessed using the options menu present in the top right corner of certain views.

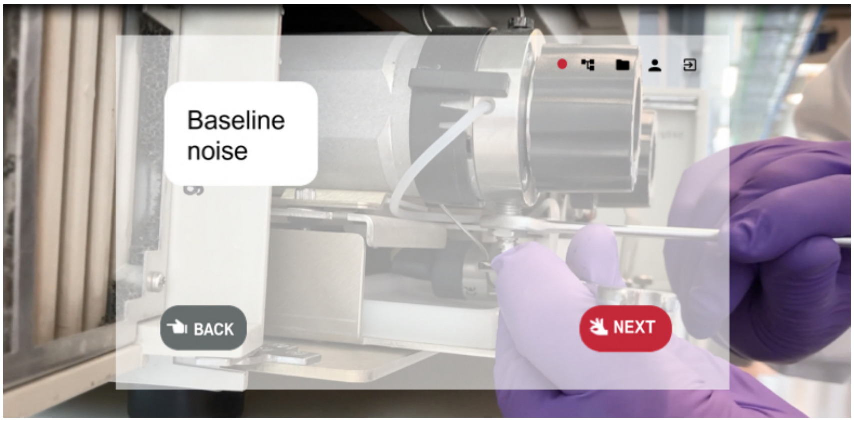

Figure 3.6: User's perspective when using the application in the Epson Moverio BT-300 smart glasses

The navigation in the application is done using either the glasses controller (fig. 3.4), voice commands (section 3.3.1) or gestures (section 3.3.2).

3.2.2 Application Programming Interface

An Application Programming Interface (API) was built in order to establish the communications between the client (application), and the company's database. The language used was Python. This API was also used to make some other operations that will be detailed bellow. Figure 3.7 represents the information flow between the client, API, and database, as well as the operations performed by the API.

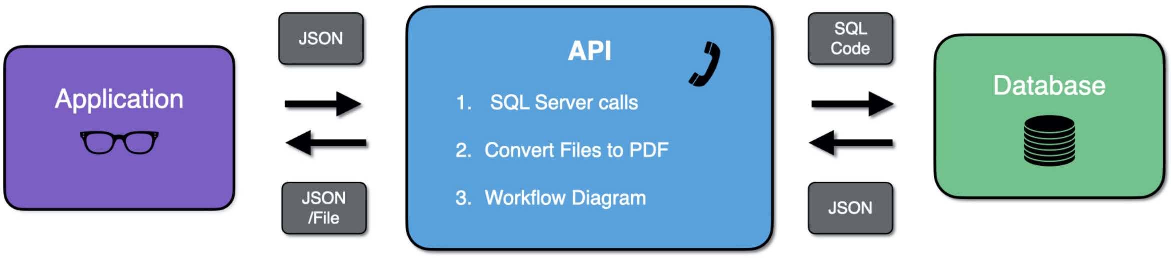

Figure 3.7: API

1. SQL Server Calls

The main purpose of the API is to retrieve information related to the workflows using SQL server calls. First a list of workflow names and the corresponding knowledge areas (KA) are retrieved. After the user selects a certain workflow from this list, the workflow itself is fetched from the database, as well as information regarding who made it and/or last modified it. The list of resources related to a certain knowledge area are retrieved when the KA is settled. A resource can then be chosen by the user using the resources option (section 3.2.4), which first shows the list of resources available and then the resource selected. Depending on the type, the resource may need to be converted inside the API, as explained bellow.

2. Convert Files to PDF

Since it is not possible to display a Word or Power-Point file inside an Android application, this type of files had to be converted to PDF format, which can be embedded in the application using a PDFView. When the user selects a document in the application, a request for the document is sent to the API that makes the SQL call to the database. A JSON is returned containing the document (in binary), its name and type. If needed this file is then converted to a PDF by using the *Python* libraries *docx2pdf* and *ppt2pdf*. The document is generated inside the API and then returned to the application that displays it.

3. Workflow Diagram

The diagram of the workflow is obtained by using the *Python* library *pyvis*. When the user selects the option to see the diagram, the JSON containing the current workflow is sent from the application to the API and is then transformed into a *pyvis* graph. This diagram is written to an HTML file that is sent to the application and displayed in a WebView (figure 3.13).

3.2.3 Application Workflow

The application transforms the workflows made with EDaM (see chapter 3) into step-by step guides in order for the analyst to follow them more easily. In this section the process of constructing the step-by-step guide is explained.

An example of an EDaM workflow is presented in figure 3.8. This example is simple and would be easy to follow by just looking at the whole diagram, but these workflows can be very complex, making it harder for the user to track its position and the task to perform next.

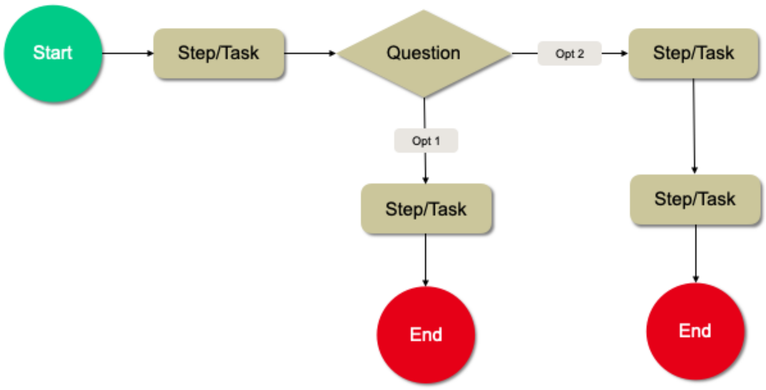

Figure 3.8: Example of a workflow diagram

In the company's database, the workflows information is stored in JSONs, that are built, edited and read by the EDaM tool. These JSONs are passed to the application by a call to the API (section 3.2.2) specifying which workflow to retrieve based on the user's choice. The information is retrieved from the JSON and is stored in objects (see figure 3.9).

Some JSONs containing the workflows information were gathered to understand which types of nodes and links were present and which information was relevant. Also from these JSONs the linking logic was retrieved and mimicked inside the application.

An example of this process is presented in figure 3.9.

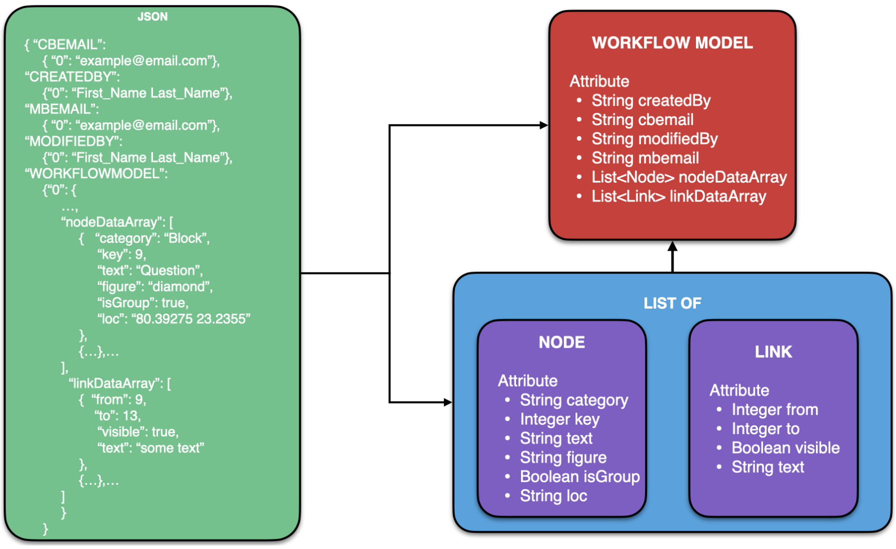

Figure 3.9: JSON string to objects

Some of the important attributes of a workflow are specified bellow.

Node: The '*category*', '*figure*' and '*isGroup*' attributes have to due with the type of node (for example if it is a '*Start*', '*End*' or '*Diamond*' node, figure 3.8). The '*text*' is the text displayed in that node, the '*key*' is the number of the node, which is used by the links to make the connections.

Link: The integers '*from*' and '*to*' are the keys corresponding to the node from which the link is coming and to which node it is going, respectively. '*visible* and '*text*' are related to the text presented in the link, if applicable. For example in figure 3.8 we have text in the links leaving the '*Diamond*' node, but not on the others.

Workflow Model: The Workflow Model object is the object that gathers all the information relative to a workflow. Not only all the links and nodes from that workflow, but also who created it and last modified it (section 3.2.4).

Once the user starts the step-by-step, the views are updated according to the corresponding node. Two examples of these views are presented in figures 3.10 and 3.11.

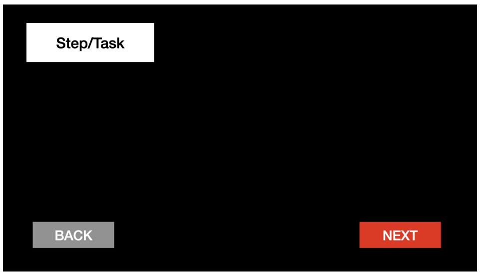

Figure 3.10: Example of a *Simple* node view

Figure 3.11: Example of a *Diamond* node view

The parsing of the JSONs add some associated challenges. Since the platform is used by different people, the workflows are not all constructed in the same way. For example, some workflows do not have a 'Start' node and the application cannot know where the workflow starts and therefore cannot construct the step-by-step guide. In this case, a message is sent to the user saying that the workflow cannot be displayed. There are also workflows that do have a start node, but it is not connected to the rest of the workflow, in this case the same message is sent. There may be other problems similar to these, but they can only be solved by a consistent way of constructing the workflows by the company members.

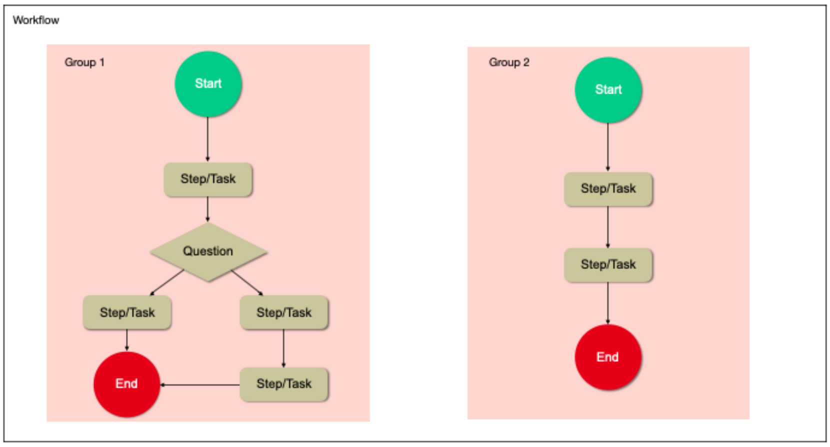

Figure 3.12: Grouped workflows example

Other challenge is related with the fact that some workflows have two or more sub-workflows, as shown in figure 3.12. For this problem, the solution found was to search the workflow for more than one 'Start' node. If more than one is found, the user has to choose which sub-workflow to follow and the other sub-workflows are ignored. This is connected to the attribute 'isGroup', when a workflow has a node of this type it means that the node is not really a node but a sub-workflow. This was also challenging when construction the diagram view. Since the user is now using a sub-workflow and not the whole workflow, a new JSON had to be constructed to be sent to the API in order to generate the diagram for this sub-workflow only. In normal workflows the JSON sent is the same as the one retrieved, so no changes need to be applied.

3.2.4 Auxiliary features

Throughout the application, in certain activities, the user can assess other features in the options menu.

Resources

The resources feature allows the user to consult documents or multimedia related to the task that is being worked on. The resources are separated by knowledge area. While inside a workflow, the user can choose this option to see a list of all the resources available and choose one from the list. This list is retrieved from the database as soon as the user decides on a KA. The resource itself is only retrieved when the user selects it. The reason for this is that the resource is a heavy file and retrieving all of them would be very time and memory consuming. The document is displayed in a *PDFView*, *ImageView* or *VideoView*, depending on the type of resource. These views are embedded in the application, no other application is opened to consult the resources. The PDFView allows the user to scroll through the document and to zoom in and out[1].

Regarding the videos, at the time of the making of this dissertation, the company had only two videos, which were not in the company's database. The videos were made specifically to integrate the application in order to make a proof of concept that this would help the analysts with tasks, more specifically with the troubleshooting of analytical equipment. For this reason the videos are not retrieved from the database, but are saved locally in the application. Nevertheless, they are associated and integrated in the list of resources available for the specific KA. If the videos proof to be valuable, more videos may be made and added to the company's database.

Created by

This feature allows the user to consult who created and/or last modified the workflow that is currently being consulted. By using this feature the user sees the name of the colleague and his email, being able to contact him in case of needing assistance.

Diagram

The diagram feature allows the user to see the full diagram of the workflow. As mentioned in section 3.7, the diagram is made inside the API, returned to the application in the form of HTML and displayed in a *WebView* (figure 3.13).

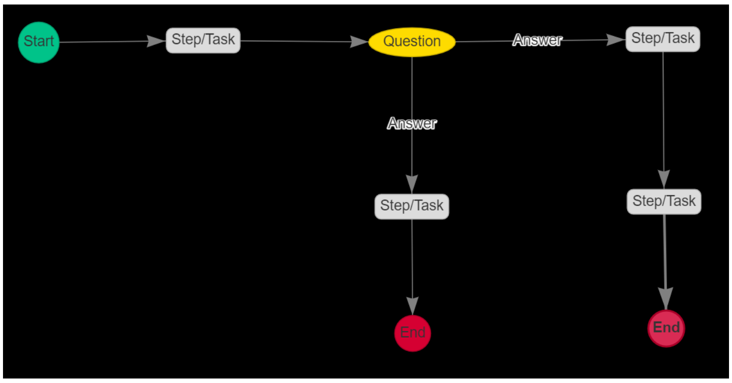

Figure 3.13: Diagram WebView example

When the workflow is complex, this diagram can be very cluttered. This form of presenting the workflow allows the user to drag the nodes to different positions and to zoom in and out on the diagram, being able to see the diagram as a whole or to get a more precise view of certain parts.

[1] Inside the *PDFView*, *ImageView* and *WebView* it is not possible to use voice commands to navigate, only to go back to the previous view.

To the Beginning

This last feature allows the user to go directly to the initial menu.

3.3 Alternative ways of navigating the application

As mentioned in section 3.1, the Moverio BT-300 SG do not provide a in-built hands-free navigation, the navigation is made through a controller. Although the controller has many advantages, mentioned in section 3.1, it is not hands-free, the user must work with only one hand while the other is handling the controller. While performing different tasks in the laboratory it was important that the user could use both hands, not only for convenience, but also for safety measures. In order to make the system hands-free, two different possibilities of navigation were implemented: speech recognition and gesture recognition. The implementation and evaluation of each of these solution is detailed in the sections bellow.

3.3.1 Navigation by Speech Recognition

For the implementation of the speech recognition feature the Vosk toolkit was used. Vosk is a free, offline toolkit compatible and easily integrated with Android, which made it the number one choice for implementation. Google Speech API is a common choice for this type of implementation, but although it is compatible with Android, it needs access to applications that are not supported by the Moverio Epson smart glasses.

Vosk library was integrated in the Android Studio Project and a service inside the application was made to implement the speech recognition feature. Each time the users speaks, the service sends the output to the current activity, which compares it to its list of commands using Levenshtein distance (LD). The Levenshtein distance, given in equation 3.1, measures the difference between two strings. It is the minimum number of single-character edits (insertion, deletion or substitution) required to change one string into the other [41]. This distance was chosen since it is the most commonly used in similar situations.[2]

$$\operatorname{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j)=0, \\ \\ \min \begin{cases} \operatorname{lev}_{a,b}(i-1,j)+1 \\ \operatorname{lev}_{a,b}(i,j-1)+1 \\ \operatorname{lev}_{a,b}(i-1,j-1)+1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases} \tag{3.1}$$

[2]The Levenshtein algorithm used was taken from [42].

	0	1	2	3	4	5	6	7	8	9
0		W	O	R	K	F	O	R	C	E
1	W	0	1	2	3	4	5	6	7	8
2	O	1	0	1	2	3	4	5	6	7
3	R	2	1	0	1	2	3	4	5	6
4	K	3	2	1	0	1	2	3	4	5
5	F	4	3	2	1	0	1	2	3	4
6	L	5	4	3	2	1	1	2	3	4
7	O	6	5	4	3	2	1	1	4	5
8	W	7	6	5	4	3	3	3	3	4

Table 3.1: Matrix for computing the Levenshtein distance between words ”Workflow” and ”Workforce”

Table 3.1 shows the matrix used to compute the Levenshtein distance between words “Workflow” and “Workforce”. Each element (i, j) represents the distance between the words $a_{1:i}$ and $b_{1:j}$, for example, element $(3, 4)$ represents the Levenshtein distance between ”Wor” and ”Work”.

For example, to compute the LD for element $(2, 2)$, applying equation 3.1:

$$\begin{aligned}\text{lev}_{a,b}(2,2) &= \min(\text{lev}_{a,b}(1,2)+1, \text{lev}_{a,b}(2,1)+1, \text{lev}_{a,b}(1,1)) \\ &= \min(1+1, 1+1, 0) \\ &= \min(2,2,0) \\ &= 0\end{aligned}$$

The result was expected since the words for element $(2, 2)$ are the same: ”Wo”.

For element $(6, 6)$, comparing the words ”Workfo” and ”Workfl”:

$$\begin{aligned}\text{lev}(6,6) &= \min(\text{lev}(5,6)+1, \text{lev}(6,5)+1, \text{lev}(5,5)+1) \\ &= \min(1+1, 1+1, 0+1) \\ &= \min(2,2,1) \\ &= 1\end{aligned}$$

Since the words are now different, the Levenshtein distance is equal to 1, which represents the exchange of an ”o” with and ”l” at the end of the words. Computing the full matrix, the Levenshtein distance is represented by the element in the bottom right corner, for this case, the Levenshtein distance is 4, shown in element $(8, 9)$.

In the application, the Levenshtein distance is calculated between the Vosk model output and each command, if this distance is zero, the function automatically returns the respective command to be executed, otherwise, an array of distances is created. From this array the smaller value is found. If there are two equally smaller values, the result is inconclusive. If there is only one, but the value of the distance is greater than the maximum of the two lengths of the words compared, the result is also inconclusive. In these two cases the user is asked to speak again. Otherwise the function returns the command that scored the minimum distance.

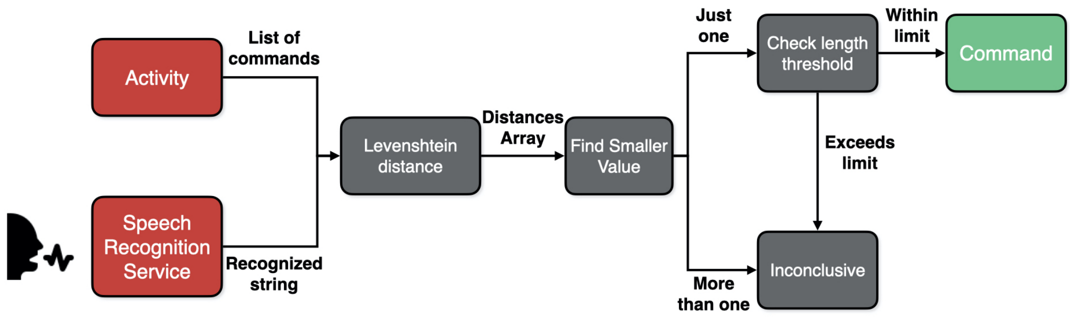

Figure 3.14: Flowchart of the speech recognition feature

A test was created to evaluate the speech service in the context of the application. The test consisted in speaking three different sequences of thirteen commands and checking how the model recognized them. The test was completed by two people. The commands presented are the ones used in the application's navigation, although some similar commands were left out (for example ”scroll down”, for being similar to ”scroll up”). Table A.1 shows all the results obtained and part of that table is presented in table 3.2. The tables show the spoken word, which corresponds to a command in the application, the Vosk model output, the number of times that prediction occurred, the Levenshtein distance between the spoken word and the outputted command (if applicable), and if the command was recognized.

Word spoken	Model Output	No of times	LD	Recognized?
Next	Next	5	0	Yes
	one axed	1	-	No
All workflows	oh workflows	1	3	Yes
	oh we're close	2	10	Yes
	oh workforce	1	-	No
	how workforce	1	7	Yes
	how workflows	1	3	Yes
Show documents	Show documents	6	0	Yes

Table 3.2: Part of the Speech Recognition Service results

The results show that of the 78 predictions, the speech model failed to correctly predict 23 (highlighted in green and red), but only 8 would not have been recognized by the application (highlighted in red) which demonstrates the importance of using the Levenshtein distance strategy. From this results we can say that the Speech Recognition service is detecting words with a 90% accuracy, although it is important to notice that this results were acquired in a different environment than the laboratory, with no background noise, and that the words were spoken by only two people. Noise, different accents and other aspects may affect the accuracy. There was only one case were the command was misinterpreted by a different command, the word spoken was ”voice commands” and the recognized command was ”show documents”. This case would be the most problematic had it happened several times, since the application responds to a command the user did not give, resulting in a poor user experience. Since this case only happened once (less than 2%) during testing, it can be considered rare.

It was noted that the command ”dismiss” was recurrently misinterpreted for ”this nice” or similar phrases that were not recognized. In order to further increase the reliability of the recognition, the phrase ”this nice” was added to the list of commands, resulting in a lower chance of this command not being recognized, also, by adding this phrase, ”this means” (and possibly other variations) will also be correctly recognized, since the Levenshtein distance between this words is smaller than between other commands.

Besides the commands shown in table A.1 the application also has to recognize the names of the knowledge areas and of the workflows, which can be variable over time. In order to prevent misinterpretations, especially for workflows whose name is an acronym, this lists were numbered, therefore the user can simple say ”number x” - x being the number corresponding to the wanted item on the list - instead of saying the text of said item.

The speech service starts running when the application is started and immediately a window appears which informs the user that the service was started, shows the basic commands, and has an option to cancel if the user does not want to use this feature. At any time the user can mute/unmute the service by clicking the volume buttons on the controller. This allows for the user to speak to a colleague while using the application and eliminates the risk of unwanted commands being recognized. To further prevent the app from taking unwanted commands, a keyword was implemented as well. When starting the step-by-step guide, the user must say the word “glasses” before any command in order for the command to be recognized.

An evaluation of the speech recognition feature by the analysts will be presented in the chapter 4.

3.3.2 Navigation by Gesture Recognition

The gesture recognition feature allows users to use hand gestures to navigate the application. This feature was added in order to further improve the hands-free component. Although the speech recognition works well, some people may be reluctant to talk to a machine in an open space environment, also, there might be some situations were the background noise is too great or where the microphone fails, so this feature serves also as a backup.

Unlike the speech recognition, the gesture recognition feature was implemented to only be used with the step-by-step guide of the workflow. This was decided because until the user starts the workflow, the hands are not required to do others tasks. Furthermore, the user needs to use the glasses controller to open the app and it can also be used to go through the steps of choosing the workflow (these last steps can also still be done using voice commands). A third reason was to keep the gestures to be used as simple as possible so that the user may remember them. Once the workflow starts, the user then has the option to navigate using gestures.

The workflow step-by-step guide is based on 3 types of interactions: go back to the previous task, go to the next task or choose from a set of options that can fall under two categories: the options from the workflow itself (from a Diamond node - fig. 3.11) or the four auxiliary features mentioned in section 3.2.4. Six gestures were thought of to navigate, which are represented in figure 3.15[3].

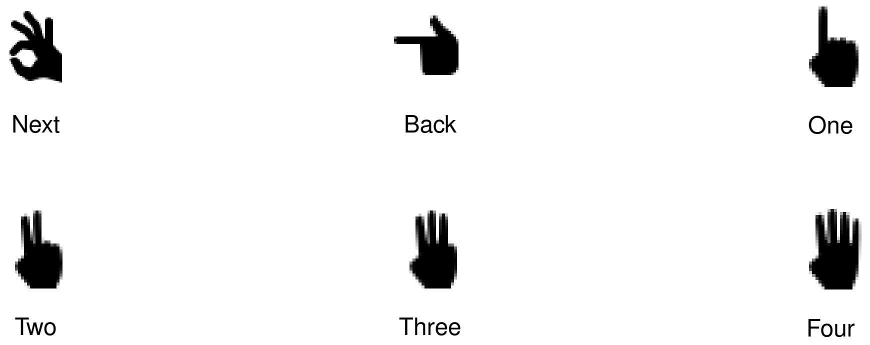

Figure 3.15: Icons representing the gestures used to navigate the application

From these gestures, numbers one through four may have different applications, depending on the type of node the user is at. If the user is at a simple node, the numbers are used to enter one of the auxiliary features (fig 3.16, left). If the user is at a diamond node, the numbers are used to choose from one of the given choices (fig 3.16, right).

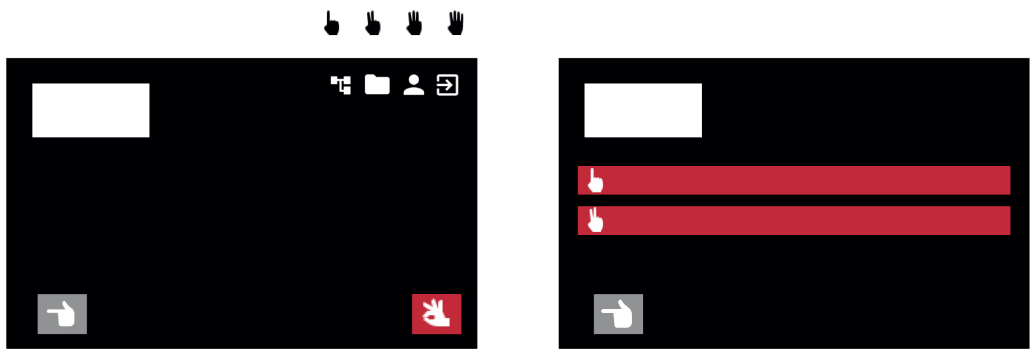

Figure 3.16: Command associated with each gesture

[3] Icons retrieved from https://icons8.com/

Machine Learning for Gesture Recognition

To implement the gesture recognition feature some machine learning models were trained to classify the gestures, using the TensorFlow Python library. The deployment of models trained with TensorFlow to mobile application is straightforward, using TensorFlow Lite, and a sample project that uses TensorFlow Lite with Android is available, which was later used to test the gesture recognition feature and to integrate it in the application.

At first an image classifier model was trained using a dataset constituted with images of the 6 hand gestures and some background images (images with no hands). The model should distinguish between 7 classes: the 6 gestures and an additional class ”none”, meaning no gesture was present. It was noticed that the classifier was not being able to classify the gestures, but was mostly identifying hands. This behavior might have been caused due to the lack of a big image dataset.

From the image classifier the strategy shifted to an object detector. The intuition was that by localizing the hand in the image the classification of the gesture would become ”easier”. The intuition turned out to be verified. To train the object detectors that will be describe below, the TensorFlow 2 Object Detection API was used, running on a Google Colaboratory notebook with access to a GPU. This API gives access to a list of models (model zoo) pre-trained on the COCO 2017 Dataset, which one can use for transfer learning [43]. Out of these models, only the SSD ones are eligible to use with TensorFlow Lite, since this type of detector has significantly lower computational power needs, compared to other detectors, and therefore can be used in devices such as smartphones and, in this case, the Moverio smart glasses.

Dataset Description

The dataset used for training is composed of images with hands making the different gestures as well as some images with hands that do not represent any gesture in particular (”None” of figure 3.17). The images needed to be taken from the smart glasses perspective and no mask should be applied to the images, these conditions were not found in any already existing dataset, therefore the images were gathered by the author. Examples of the images used are presented in fig 3.17. The number of images per label is presented in table 3.3.

Gesture	Next	Back	One	Two	Three	Four	**Total w/o None**	None	**Total w/ None**
No of Images	27	27	25	30	29	27	165	41	206

Table 3.3: Dataset constitution

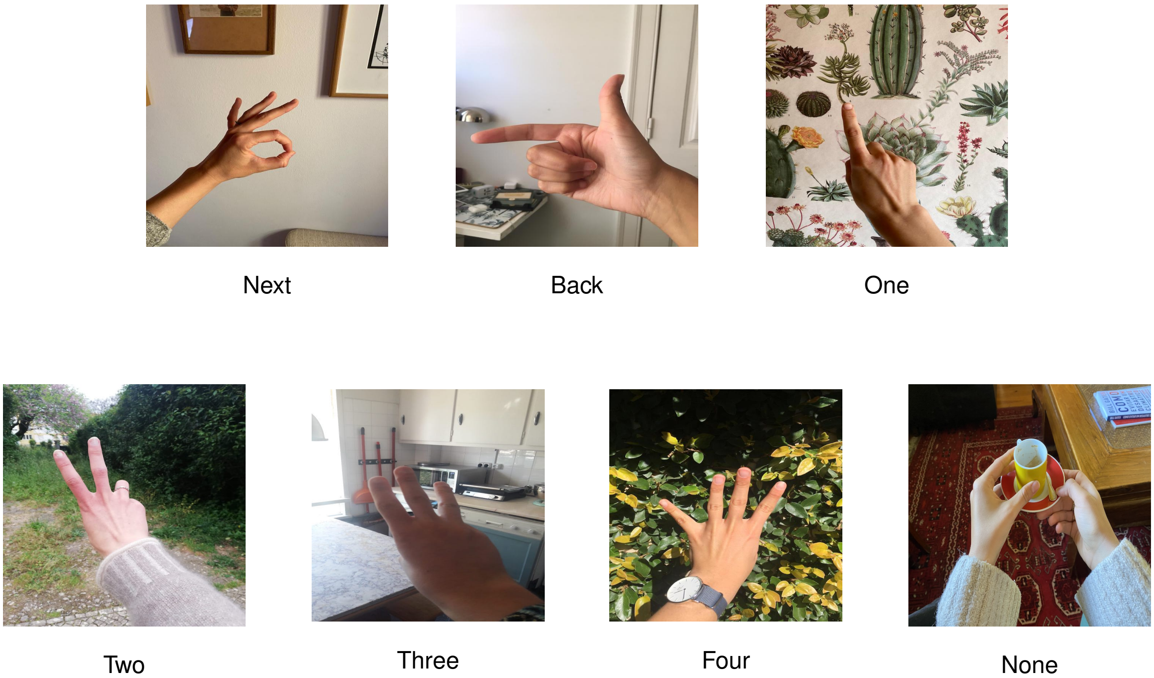

Figure 3.17: Examples of images in the dataset

Two different datasets were used in different models. One model was trained with only the six gestures images, other two models used an additional set of ”None” images. The datasets were divided into training and testing sets. The testing set is constituted by 2 images of each of the gestures (12 for model without ”None”, 14 for models with ”None”). The training set is constituted by the remaining images (153 for model without ”None”, 192 for models with ”None”).

Training

Three different models were trained and tested in order to find the one that best suited the problem. All of the models use the SSD Mobilenet v2 FPNLite architecture [44], no changes where made to the default architecture provided by TensorFlow, other than the number of classes, the number of training steps and the batch size (that was set to 4). This architecture uses the Mobilenet v2 as the classification network. The main advantage of using the Mobilenet v2 network is that it uses Depthwise Separable Convolution, which decreases the number of computations needed, with only a small accuracy loss, compared to the standard convolution method [45]. The localization loss is the weighted smooth l1 and the classification loss the weighted sigmoid focal. The threshold value used for the IoU is 0.6.

Table 3.4 shows the different models trained with the respective input sizes of the image, number of labels (6 for the one without the ”None” label, 7 for the ones with the ”None” label), number of training steps and Mean Average Precision (mAP).

Id	Input size	No of labels	No of Training Steps	mAP (IoU=0.5:0.95)
1	320x320	6	13000	0.846
2	320x320	7	15000	0.704
3	640x640	7	9000	0.708

Table 3.4: Tested models metrics

Tables 3.5, 3.6 and 3.7 show the confusion matrix obtained using the testing set for models 1, 2 and 3, respectively. Model 1 confusion matrix shows that all the predictions are correct, it is also the model with the higher mAP. The reason for this may be that with the addition of the 7th class, the model became ”confused”, since this class consists of many different gestures.

Looking at the confusion matrices of models 2 and 3, model 2 shows some false positives, but it detected all the gestures. Model 3 failed to detect correctly one image, it also shows some wrong detections.

The number of training steps of model 3 is considerably lower than the other two models, this is due to the fact that this model was soon discarded once it was tested in the android application due to its high latency while making detections in real-time. If trained for a longer period of time, the model could have achieved better results, but it would not be possible to use it in the application. This higher latency is due to the fact that model 3 has a larger image input size, which leads to a larger number of computations needed and, consequently, a decreased detection speed. Models 2 and 3 have almost the same mAP, but model 2 had more training steps, this is because the larger image input size provides the model with better accuracy, since it can gather more features from the image, this is why this architecture was tested, but by assessing its decrease in speed, this and other more computational heavy models were discarded for this task.

		Ground Truth						
		Next	Back	One	Two	Three	Four	Nothing
Prediction	Next	2	0	0	0	0	0	0
	Back	0	2	0	0	0	0	0
	One	0	0	2	0	0	0	0
	Two	0	0	0	2	0	0	0
	Three	0	0	0	0	2	0	0
	Four	0	0	0	0	0	2	0
	Nothing	0	0	0	0	0	0	0

Table 3.5: Confusion matrix model 1

Prediction \ Ground Truth	Next	Back	One	Two	Three	Four	None	Nothing
Next	2	0	0	0	0	0	0	0
Back	0	2	0	0	0	0	0	0
One	0	0	2	0	0	0	0	0
Two	0	0	0	2	0	0	0	0
Three	0	0	0	0	2	0	0	1
Four	0	0	0	0	0	2	0	1
None	0	0	0	0	0	0	2	0
Nothing	0	0	0	0	0	0	0	0

Table 3.6: Confusion matrix model 2

Prediction \ Ground Truth	Next	Back	One	Two	Three	Four	None	Nothing
Next	2	0	0	0	0	0	0	0
Back	0	2	0	0	0	0	0	0
One	0	0	2	0	0	0	0	0
Two	0	0	0	2	0	0	0	0
Three	0	0	0	0	1	0	0	1
Four	0	0	0	0	1	2	0	0
None	0	0	0	0	0	0	2	0
Nothing	0	0	0	0	0	0	0	0

Table 3.7: Confusion matrix model 3

Testing

After training, the models were deployed to the android application provided by TensorFlow in order to evaluate them in real-time, as said before, model 3 was discarded at this phase and it was not further tested. These tests were not made using the smart glasses, since they were unavailable[4], but using a different android device with similar specifications in terms of processing power. The test consisted of making three sequences of nine gestures to be detected by the android application in real-time. The nine gestures consisted of the six gestures that trigger the commands (from figure 3.15) and three additional random gestures (five, closed hand and holding a pen) to represent the "None" class, the objective of these classes was to be detected as "None" (only by model 2) or to not trigger any detection.

Five

Closed hand

Two

Figure 3.18: Examples of detections in the TensorFlow android application

[4]The smart glasses were being used in the laboratory in order to gather the users' feedback (chapter 4)

The sequences of gestures were made while running the application. In order to gather the results, the screen was recorded as the detections appeared on screen, the recordings were carefully observed and the results are presented in tables A.2 and A.3. These results were then translated into the confusion matrices of tables 3.8 and 3.9. The number of detections was not controlled, since the application was running in real-time. Both the models outputted a total of 63 detections from the three sequences of gestures. The number of detections varied for each label. Examples of the results obtained are presented in figure 3.18. It is important to note that the minimum confidence for the detections was set to 90%, detections with lower confidence were not considered.

Prediction \ Ground Truth	Next	Back	One	Two	Three	Four	Nothing
Next	9	0	0	0	0	0	0
Back	0	9	0	0	0	0	0
One	0	0	8	0	0	0	7
Two	0	0	0	6	0	0	0
Three	0	0	0	0	9	1	0
Four	0	0	0	0	0	8	6
Nothing	0	0	0	0	0	0	0

Table 3.8: Confusion matrix of real-time detections from model 1 - 90% confidence threshold

Prediction \ Ground Truth	Next	Back	One	Two	Three	Four	None	Nothing
Next	6	0	0	0	0	0	3	0
Back	0	8	0	0	0	0	0	0
One	0	0	9	0	0	0	1	0
Two	0	0	0	10	0	0	0	0
Three	0	0	0	0	8	0	0	0
Four	0	0	0	0	0	9	6	0
None	0	0	0	0	0	0	3	0
Nothing	0	0	0	0	0	0	0	0

Table 3.9: Confusion matrix of real-time detections from model 2 - 90% confidence threshold

By observing the confusion matrices, one can see that although the majority of the detections are correct, several "non-gestures" were misinterpreted. We can see that these "non-gestures" are usually confused with "One", "Four" and "Next", which is due to the nature of the gestures chosen to represent the "None" class. A "five" was chosen on purpose to see how the model reacted to similar gestures, it was confused with a "four" several times (tables A.2, A.3). Comparing the models, model 1 has less True Positives (TP) and more False Positives (FP) than model 2, although the difference is not great (table 3.14).

Class	Confidence Values				Is detected with threshold=98%			
	Correctly Classified		Wrongly Classified		Correctly Classified		Wrongly Classified	
	Higher	Average	Higher	Average	Higher	Average	Higher	Average
Next	99,98	99,75			Yes	Yes		
Back	99,90	99,63			Yes	Yes		
One	99,62	99,30			Yes	Yes		
Two	99,65	99,07			Yes	Yes		
Three	99,31	98,89			Yes	Yes		
Four	99,20	98,29	98,24	98,24	Yes	Yes	Yes	Yes
Five			97,93	95,72			No	No
Closed			99,37	98,94			Yes	Yes
Pen								

Table 3.10: Model 1

The confidences of the detections were also examined in order to better evaluate the models. The confidence is an important metric in this case since a threshold can be applied in order to only detect gestures above a certain level of confidence, this can be used to discard many wrong predictions, on the other hand, it also requires that correct predictions have a high confidence level in order for the gesture to be recognized. The higher confidence value and the average of all confidence values for each class were collected and are presented in tables 3.10 and 3.11. These tables show the outcome of applying a higher threshold to the confidence level. The 98% threshold was chosen since it was the one that, on one hand, maximized the correctly classified detected classes, and on the other hand, minimized the wrongly classified detected classes. It is possible to see here that by using model 2 (with the threshold of 98%) there would be less wrongly detected classes: one if using the higher value, none if using the average, compared to two if using the higher value and two if using the average, using model 1.

Class	Confidence Values				Is detected with threshold=98%			
	Correctly Classified		Wrongly Classified		Correctly Classified		Wrongly Classified	
	Higher	Average	Higher	Average	Higher	Average	Higher	Average
Next	99,99	99,98			Yes	Yes		
Back	99,98	99,96			Yes	Yes		
One	99,92	99,57			Yes	Yes		
Two	99,94	99,35			Yes	Yes		
Three	99,62	98,60			Yes	Yes		
Four	99,88	99,79			Yes	Yes		
Five			99,65	97,01			Yes	No
Closed	94,14		93,78	93,78			No	No
Pen	94,05							

Table 3.11: Model 2

The 98% threshold was then applied and the tests were repeated in order to see if in fact the results would confirm the theory. The results are presented in tables A.4 and A.5 and the respective confusion matrices in tables 3.12 and 3.13.

Prediction \ Ground Truth	Next	Back	One	Two	Three	Four	Nothing
Next	7	0	0	0	0	0	0
Back	0	8	0	0	0	0	0
One	0	0	7	0	0	0	7
Two	0	0	0	6	0	0	0
Three	0	0	0	0	2	0	3
Four	0	0	0	0	0	9	2
Nothing	0	0	0	0	1	0	0

Table 3.12: Confusion matrix of real-time detections from model 1 - 98% threshold

Prediction \ Ground Truth	Next	Back	One	Two	Three	Four	None	Nothing
Next	8	0	0	0	0	0	0	6
Back	0	8	0	0	0	0	0	0
One	0	0	7	0	0	0	0	0
Two	0	0	0	8	0	0	0	0
Three	0	0	0	0	5	0	0	0
Four	0	0	0	0	0	8	0	0
None	0	0	0	0	0	0	0	0
Nothing	0	0	0	0	0	0	0	0

Table 3.13: Confusion matrix of real-time detections from model 2 - 98% threshold

From the confusion matrices, table 3.14 was then constructed in order to compare the four models. It shows the percentage of True Positives and False Positives detected by each model. With these results we can see that the threshold did improve the result of model 2 by 4%, but the results for model 1 worsened.

	90		98	
	TP%	FP%	TP%	FP%
Model 1	78	22	75	23
Model 2	84	16	88	12

Table 3.14: Comparison True Positives and False Positives percentage for each model and threshold value

Evaluating all these results the natural choice would be model 2 with a 98% threshold, which performed 10% better than model 1. Nevertheless, it is important to notice that many aspects might have changed these results:

- The results were not performed with the Smart Glasses, but with a different device
- The gestures that represent the "None" class were made explicitly to the camera, which would not happen in the real environment
- Light and background variations - the tests were not performed in the environment where the system will be used

The next evaluation step for these models would be to test them in real life conditions, i.e. in the laboratory, using the smart glasses, ideally following a procedure.

Android Implementation

A second application was built to run the camera and detections in the background. This application was built on top of the TensorFlow Lite example for android that was mentioned in the paragraph "Machine Learning for Gesture Recognition" of this section. Some changes were made in order to not show the camera preview while running the app and run the camera as a service that can be started from the other application and send it the results.

This service starts when the user starts a workflow (i.e. the user clicks the button on view number 3 from fig. 3.5). A view with two buttons is displayed where the user decides to start the gesture recognition feature or to cancel it. If the choice is to cancel, the app runs normally as shown in the previous sections. If the choice is to start, the camera service is started and the views appear with the gesture icons in order for the user to know which gestures to make, as shown in fig. 3.16 (this does not happen for the auxiliary features).

Chapter 4

User acceptance of Augmented Reality System in the Analytical Laboratory

In this chapter the case study conducted to test and evaluate the augmented reality system in the laboratory is described. An extended version of the technology acceptance model was used for this purpose, which is described in section 4.1. The case study procedure and the questionnaires used are presented in section 4.2, and the respective results in section 4.3.

4.1 Technology Acceptance Model

To evaluate the user acceptance of the augmented reality system an extended technology acceptance model was used. To the original model that includes the variables **perceived usefulness** and **perceived ease of use**, were added other variables that were found to be relevant in the the acceptance of similar systems (section 2.6). These variables were: **perceived enjoyment**, **personal innovativeness** and **cybersickness**. The cybersickness component was evaluated using a version of the Simulation Sickness Questionnaire. A sixth variable was also included since it was noted that the augmented reality wearables were many times regarded as not comfortable, this variable was called "**ergonomics**" and is related to the comfort and consequent effectiveness of using a wearable device. Masood and Egger mention ergonomics as one of the most reported challenges for the implementation of AR [46]. The AR wearable devices have not yet reached maturity and need further technological development to be able to have both capable hardware, and be comfortable to wear, not injuring the users or affect their job. This variable was introduced in the model to understand the users perception of ergonomics of the

Moverio Epson smart glasses and how it affects the intention to use the system.

All the variables definitions and the questions that were included in the questionnaire to measure each variable can be found in section 4.2.3.

The expectations for the results of this experiment is that the users will find the system useful and ease to use, and that they will intend to use the system in the laboratory, providing them with a more efficient way of following the procedures. Cybersickness symptoms are expected to be minimal, since the user interface is very simple. Ergonomics is expected to be poorly evaluated since this is one of the aspects that needs improvement in smart glasses.

Hypothesis were made for the correlation between the variables which can be found in appendix B. Unfortunately it was not possible to confirm the hypothesis due to the small sample size of the case study, the reason for this small sample size is detailed in the next section.

4.2 Case Study Procedure and Questionnaires

In order to assess the acceptance of the analyst towards the developed application a case study was conducted. The case study was made in Hovione's R&D analytical laboratory and had the participation of 4 analysts. Due to the pandemic situation, the facilities had restricted access and the laboratories were operating with less people, which made it hard to get more analysts to participate. The author did not accompany the case study in the laboratory, the analysts had full autonomy during the experiment ensuring non-biased results. The author's intervention was reduced to solving technical details. This section contains the procedure of the case study and the questions included in the questionnaires.

4.2.1 Procedure

The case study was divided into the following steps:

1. It was explained to the analyst the purpose of the experiment, how to use the glasses, and the basics of the application usage

2. The analyst filled a pre-questionnaire (C.1)

3. The analyst used the glasses and the application to perform tasks in the laboratory

4. The analyst filled a post-questionnaire (C.2)

The pre-questionnaire, presented in appendix C.1, gathers some demographic information about the participants as well as the expectations and excitement towards augmented reality and the experiment.

During testing time, the analysts were free to choose the workflow performed and what features to use. Each analyst may have used different workflows and features. Each analyst was also free to test the application more than once, the number of times each analyst performed a task using the system were not recorded. The tested application integrated the speech recognition navigation, but not the gesture recognition one, since this feature was developed while the tests were ongoing.

At the end of the testing period the analyst filled the final questionnaire, which is presented in appendix C.2. At the beginning of the questionnaire, the participant pointed out which type of workflow was followed and which features of the application were used. Question regarding the speech recognition were also introduced to better understand the benefits and constraints of this feature. The second part of the questionnaire consisted of questions regarding the technology acceptance model variables mentioned in section 4.1. At the end of the questionnaire there was the possibility to add an optional comment.

4.2.2 Questions Included in the Pre-questionnaire

In the pre-questionnaire the participants were asked about the following topics:

- age group
- education level
- field of study
- time working at the current position
- previous experience with augmented reality
- excitement level regarding the experiment
- expectation level regarding augmented reality in the laboratory

These topics were included to provide a general idea of the participants background and expectation for the experiment.

4.2.3 Questions, Variables and Statements Included in the Post-questionnaire

The post-questionnaire started by asking what type of workflows and what features were used. If the participants used the speech recognition feature, they answered some related questions, if not, the reason for not using this feature was asked. The following questions were constructed in a 7-point likert-type scale, except the SSQ, whose symptoms are rated between none, mild, moderate, severe and very severe. The evaluated variables and their definitions are presented in table 4.1.

Variable	Definition
Perceived usefulness	The degree to which a person believes that using a particular system would enhance his or her job performance
Perceived ease of use	The degree to which a person believes that using a particular system would be free of effort
Perceived enjoyment	The extent to which the activity of using a specific system is perceived to be enjoyable in its own right, aside from any performance consequences resulting from system use
Personal innovativeness	The willingness of an individual to try out any new information technology
Ergonomics	The degree to which an individual thinks that the equipment is comfortable and effective.
Cybersickness	The level of physical discomfort the individual experiences due to exposure to virtual environments.
Intention to use	The degree to which a person has formulated conscious plans to perform or not to perform some specified future behaviors .

Table 4.1: Variables studied in the post-questionnaire

Bellow are the statements that were included in the questionnaire to evaluate each of the variables and the included symptoms from the SSQ.

Perceived Usefulness

- Using the system improves my job performance
- The system allows me to accomplish tasks more quickly
- The system increases my productivity
- The system is useful in my job.

Perceived Ease of use

- The system is easy to learn.
- It is easy to get the system to do what I want.
- The system is clear and understandable.
- I find the system easy to use.

Personal Innovativeness

- If I heard about an information technology, I would look for ways to experiment with it.
- Among my peers, I am usually the first to try out new technologies.
- In general, I am excited to try out new technologies.

- I like to experiment with new technologies.

Perceived enjoyment

- Using the system is fun
- I enjoyed using the system
- Completing the tasks using the system in enjoyable

Intention to Use

- I intend to use Augmented Reality in the laboratory in the future.
- If the system was available in the laboratory I would use it.
- I would recommend the system to a college.

Cybersickness symptoms: nausea, increased salivation, sweating, dizziness, vertigo, stomach awareness, burping, fatigue, headache, eyestrain, difficulty focusing, difficulty concentrating, fullness of head, blurred vision.

4.3 Case Study Results

4.3.1 Pre-questionnaire

All four participants had between 18 to 30 years of age, neither of them had experienced with augmented reality before. Information related to the background of the analysts is presented in figure 4.1.

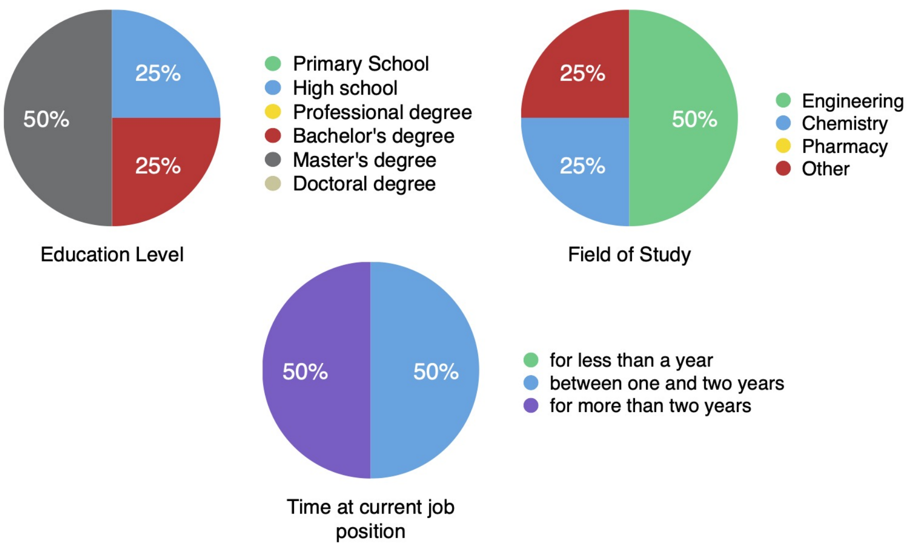

Figure 4.1: Education level, field of study and time at the current job of the participants.

The participants had somewhat different backgrounds, with different levels of education and fields of study. Regarding the time at the current job positions, two participants had between one and two years at the current job and other two were at that job position for more than two years.

Figure 4.2 shows the level of excitement of the participants towards the experiment and the expectations level regarding augmented reality in the laboratory. It is possible to see that not all participants were excited with the experiment, nevertheless the expectations for AR were overall positive.

Figure 4.2: Level of excitement regarding the augmented reality system experience and expectation regarding AR in the laboratory

4.3.2 Post-questionnaire

Figure 4.3 represents the types of workflows and application features used by the participants. All participants followed a troubleshooting workflow (related to the troubleshooting of laboratory equipment), one participant also followed a development workflow (related to the analytical development area, which contains most of the workflows used in the laboratory). As for the features, all participants used the step-by-step guide, one participant also consulted a document, and another a diagram view. There were two participants that did not use any auxiliary feature. The created/modified by auxiliary feature was not used by any participant.

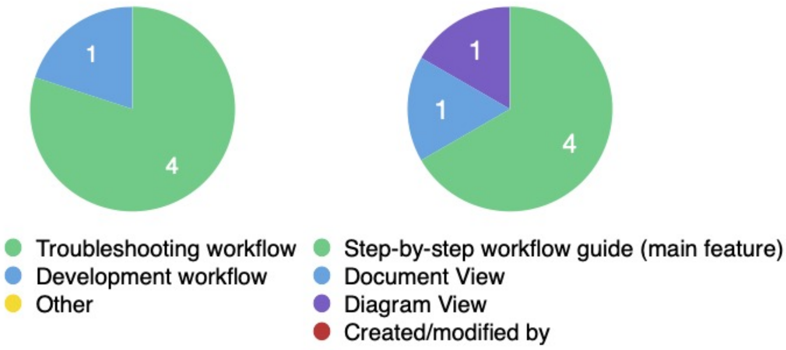

Figure 4.3: Type of workflow and application features used by the participants

Speech Recognition

The 3 participants that experimented the speech recognition feature were asked to evaluate their experience. The 1 participant that did not use the speech recognition was asked to rate the reasons for not using it. The results are shown in figures 4.4 and 4.5, respectively.

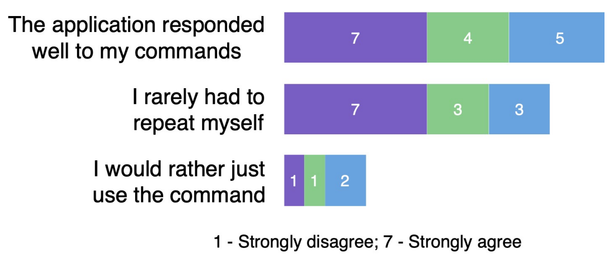

Figure 4.4: Participants evaluation of the speech recognition feature

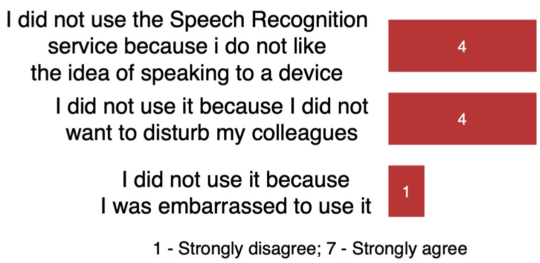

Figure 4.5: Reasons for not using the speech recognition feature

The results on the left (fig. 4.4) show that the participants would rather use the speech recognition than the glasses controller to navigate the application, even though the need to repeat commands was not infrequent for two of the participants. On the right the higher rated reasons for not using the speech recognition were that the participant did not like the idea of speaking to a device and that he/she did not want to disturb the colleagues in the laboratory.

Cybersickness

The cybersickness responses are presented in Figure A.2. One participant did not report any symptom. Symptoms reported by other participants include mild and/or moderate difficulties in focusing and concentrating, and blurred vision. Being that the participants did not report any severe symptoms, it is not likely that these symptoms had a considerate impact on the results. Although they show the flaws in the HMD hardware and advocate for its improvement.

7-Point Likert Scale Questions

This section contains the results to all 7-point likert scale questions, these include all the questions related to the variables of table 4.1, except for the cybersickness variable. The participant rated each statement with a value between 1 - strongly disagree, and 7 - strongly agree. The results to all questions can be found in figure A.1. Figure 4.6 shows the average score for each variable and participant. These averages were computed by using the results from each statement related to the selected variable (the statements per variable can be found in section 4.2.3).

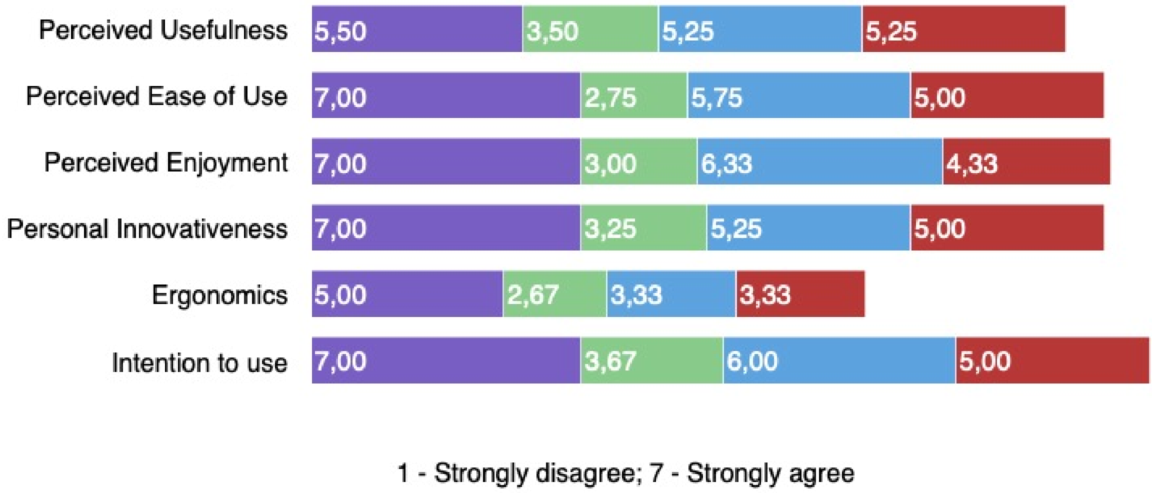

Figure 4.6: Average response for each variable and for each person

Looking at the results the variable that scored the highest sum of averages was intention to use, this is a good result since it means that the analysts intend to use the augmented reality system in the future, although there was one participant whose average was still slightly bellow the median. Inside this category are two of the highest rated statements, which include if the user intends to use the augmented reality system in the laboratory in future, and if the user would recommend the system to a colleague, both of which show the applicability of the system in the laboratory.

The variables perceived ease of use, perceived enjoyment and personal innovativeness all scored fairly similar and overall good results. The results for ease of use reveal the intuitiveness of the applications design. Perceived enjoyment and personal innovativeness are more related to personal traits, although they can influence and be related to other variables.

Perceived usefulness had an overall score a little bit lower than the previous three variables. For this variable it is important to look at the individual statement results of figure A.1. The results show that the users did not agree with the statement that says: “the system allows me to accomplish tasks more quickly”. Another statement that also got a lower score was: “the system improves my productivity”, which can be related to the previous one since productivity can be interpreted as a rate of work per unit of time. But the other two statements that evaluate perceived usefulness got high ratings. These were: ”using the system improves my job performance”, and ”the system is useful in my job”. From these results one can conclude that the analysts did find the system useful, just not in terms of time saving.

The variable that scored the lowest sum of averages was ergonomics, this variable was related to the smart glasses used. According to the results presented in figure A.1, the participants rated the statement ”the glasses are comfortable to use” with the lowest rating of all. The second lowest was also related to this variable being ”the smart glasses adjusted well to my head”. The fact that the wearable device is not comfortable can greatly impact the users intention to use the device. This is an aspect of head-mounted devices that needs to be improved in order to accomplish a greater acceptance of the technology.

It is possible to see that although the results are overall positive there is still much room for improvement. Putting these values in percentage and averaging all participants, the intention to use the system would rate 73.61%. These results can be due to the personality of the users, which there is some evidence of in the results: participants that rated lower in personal innovativeness also gave lower ratings to the other categories. Other cause could be the fact that the analysts are not used to this type of system which has a learning curve despite its intuitiveness, more time spent with the system might have had a positive effect on the results.

Additional Comments

Two participants added an additional comment at the end of the questionnaire. Both comments are included here:

1. "The glasses/application are very interactive, easy to use and the voice recognition is quite good. The key word for the glasses to respond to the command is something that should be modified, so that the glasses are not always asking to repeat themselves, because sometimes the need arises to talk to a colleague in the lab."

2. "I have found the glasses to be a useful tool in our day to day problem solving. It is easier to follow the workflows on the glasses than on the computer. Because we can stand by the equipment and follow the workflows while we do the procedure."

Regarding the first comment, The key word asked by the user was integrated in the system prior to the following participants experiment. This feature is mentioned at the end of section 3.3.1. The second comment states that the application was able to solve one of the disadvantage mentioned in the motivation section 1.3.1, expressing that augmented reality applications is a more portable tool than the convectional one, and that it helps with the workers daily tasks.

4.3.3 Results analysis

From the case study results, and considering its limitations, it is possible to conclude that the analysts found the system overall useful. It was regarded as a tool that could help in their day-to-day tasks, by providing a flexible and easier way of following the procedures in the laboratory. It was also observed that the analysts intend to use the system in the future. These results are according to the expectations mentioned in section 4.1, although it was expected that the usefulness of the solution would score higher values, this difference was due to the fact that the analysts did not find the system to save them time in the completion of tasks, as was mentioned above in the *7-Point Likert Scale Questions* segment, in this section. This aspect might have been better evaluated had the analysts more time to test the application, the perception of time would be more accurate, since the analysts would be more accustomed to the technology.

The results also show that there are still some improvements to be made. In terms of hardware, the glasses had a poor ergonomics evaluation that advocates for improvement, as expected. Also, the intention to use could be increased with force of habit, the analysts are not yet accustomed to use this type of system and a longer period of usage can improve the analysts attitude towards the technology.

Nevertheless we have to take into consideration that the case study has a very small sample, which does not allow to take further conclusions.

Chapter 5

Conclusions

This thesis proposed an augmented reality application to be used in the analytical laboratory of a pharmaceutical company with the objective of facilitating the analysts work by providing an easier and more flexible way of following procedures.

The proposed application allows the analysts to view the procedures in a step-by-step guide, with the instructions for each step being displayed in the analysts' field of vision through the use of smart glasses. These procedures are based on workflows that are retrieved directly from the company's database and are then transformed into the step-by-step guides by the application. Other features were added to the application in order to further help the user with the procedure, such as the possibility of consulting documents and who created the workflow.

To increase the flexibility of the solution, the application features hands-free navigation, which was accomplished by implementing two services: speech recognition, that allows the users to navigate the application using voice commands; and gesture recognition, that gives the user the possibility of navigating the procedure using gestures. These types of navigation allow the user to keep doing a task while still consulting the procedure, which was not possible before, when the analyst had to consult the procedures in a small computer. It also eliminates the possibility of cross-contamination considering that the analyst does not have to touch the glasses while doing the procedure.

The application running on Moverio Epson BT-300 smart glasses was evaluated by the analysts through a case study conducted in Hovione's analytical laboratory. The analysts tested the application following troubleshooting of equipment and analytical development workflows in a real case scenario. Feedback from the case study led to the conclusion that the system is a valuable asset in the laboratory and that it accomplishes its purpose of helping the analysts with their job, meeting the initial objectives of this thesis. Nevertheless, there are many aspects that can be improved to further increase the usefulness of the application and the user acceptance.

The developed work will be implemented by Hovione to give the analysts access to the tool, deploying the application to a number of smart glasses that will be permanently available in the laboratory.

5.1 Future Work

Providing that the work presented in this thesis was the first attempt of implementing an augmented reality solution in Hovione's analytical laboratory, there are several aspects that can be improved. Some possibilities of future work are described bellow divided by topics.

AR Application: The application could be turned from a view-based AR application to a triggered AR application. The start of procedures could be triggered by using markers or by detecting the equipment in the laboratory, creating an even more interactive application and eliminating the time to search and select the procedure. It could also be expanded to contain other procedures from different sources. The pharmaceutical industry is still very paper based and is trying to shift to digital formats. One solution being implemented in Hovione, with the name ”electronic lab notebooks”, is to transform analytical methods manuals, originally in paper format, to digital format, and to present them in a step-by-step guide. These manuals could be integrated inside the application, expanding its usage. Another function could be to use the application for training purposes by providing more detailed information on how to perform the procedures.

Case Study: The case study had the participation of only 4 analysts. Future research could replicate the study with more analysts to get a more representative population. With a greater sample size a statistic analysis could be made and the hypothesis of the technology acceptance model could be proved or disproved.

Gesture Recognition: The gesture recognition feature allows the user to follow the procedure using gestures, but it is not available throughout the whole application. This service could be extended to the whole application, instead of just the step-by-step guide, to serve as a backup for the speech recognition feature if there is some impediment for using voice commands, such as background noise. Furthermore the object detection models could be improved in order to increase detection accuracy. This could be achieved by gathering a larger dataset, by training the models for a longer number of training steps, or by changing the models architecture.

Bibliography

[1] M. Lorenz, M. Rüßmann, M. Waldner, P. Engel, M. Harnisch, P. Gerbert, and J. Justus. (2015) Industry 4.0: The future of productivity and growth in manufacturing industries. [Online]. Available: https://www.bcg.com/publications/2015/engineered_products_project_business_industry_4_future_productivity_growth_manufacturing_industries

[2] L. De Nul, M. Breque, and A. Petridis, *Industry 5.0 - Towards a sustainable, human-centric and resilient European industry*, 2021.

[3] FDA, "The drug development process," https://www.fda.gov/patients/learn-about-drug-and-device-approvals/drug-development-process, accessed: 2021-06-02.

[4] I. C. Reinhardt, D. J. C. Oliveira, and D. D. T. Ring, "Current Perspectives on the Development of Industry 4.0 in the Pharmaceutical Sector," *Journal of Industrial Information Integration*, vol. 18, p. 100131, 2020.

[5] E. Bottani and G. Vignali, "Augmented reality technology in the manufacturing industry: A review of the last decade," *IISE Transactions*, vol. 51, no. 3, pp. 284–310, 2019.

[6] R. T. Azuma, "The road to ubiquitous consumer augmented reality systems," *Human Behavior and Emerging Technologies*, p. e113, 2019.

[7] M. Dalle Mura and G. Dini, "An augmented reality approach for supporting panel alignment in car body assembly," *Journal of Manufacturing Systems*, vol. 59, no. January, pp. 251–260, 2021.

[8] F. De Crescenzio, M. Fantini, F. Persiani, L. Di Stefano, P. Azzari, and S. Salti, "Augmented reality for aircraft maintenance training and operations support," *IEEE Computer Graphics and Applications*, vol. 31, no. 1, pp. 96–101, 2011.

[9] P. Fraga-Lamas, T. M. Fernández-Caramés, Ó. Blanco-Novoa, and M. A. Vilar-Montesinos, "A Review on Industrial Augmented Reality Systems for the Industry 4.0 Shipyard," *IEEE Access*, vol. 6, pp. 13 358–13 375, 2018.

[10] L. F. De Souza Cardoso, F. C. M. Q. Mariano, and E. R. Zorzal, "A survey of industrial augmented reality," *Computers Industrial Engineering*, vol. 139, p. 106159, 2020.

[11] N. Murauer, N. Panz, and C. Von Hassel, "Comparison of scan mechanisms in augmented reality supported order picking processes," *CEUR Workshop Proceedings*, vol. 2082, pp. 69–76, 2018.

[12] V. Alcácer and V. Cruz-Machado, "Scanning the Industry 4.0: A Literature Review on Technologies for Manufacturing Systems," *Engineering Science and Technology, an International Journal*, vol. 22, no. 3, pp. 899–919, 2019.

[13] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, "Augmented Reality: A class of displays on the reality-virtuality continuum," *Telemanipulator and Telepresence Technologies*, vol. 2351, no. December 2013, pp. 282–292, 1995.

[14] I. E. Sutherland, "A head-mounted three dimensional display," in *Proceedings of the December 9-11, 1968, fall joint computer conference, part I on - AFIPS '68 (Fall, part I)*, vol. 5, no. 3. New York, New York, USA: ACM Press, 1968, pp. 757–764.

[15] G. Lintern and S. N. Roscoe, "Transfer Of Landing Skill After Training With Supplementary Visual Cues," *Visual Simulation and Image Realism I*, vol. 0162, no. December 1978, pp. 83–87, 1978.

[16] T. Caudell and D. Mizell, "Augmented reality: an application of heads-up display technology to manual manufacturing processes," *Hawaii international conference on system sciences*, pp. 659–669, 1992.

[17] L. Rosenberg, "The Use of Virtual Fixtures as Perceptual Overlays to Enhance Operator Performance in Remote Environments." p. 50.

[18] A. Edwards-Stewart, T. Hoyt, and G. M. Reger, "Classifying different types of augmented reality technology," *Annual Review of CyberTherapy and Telemedicine*, vol. 14, no. January, pp. 199–202, 2016.

[19] M. Gattullo, A. Evangelista, A. E. Uva, M. Fiorentino, and J. Gabbard, "What, How, and Why are Visual Assets used in Industrial Augmented Reality? A Systematic Review and Classification in Maintenance, Assembly, and Training (from 1997 to 2019)," *IEEE Transactions on Visualization and Computer Graphics*, vol. 2626, no. c, pp. 1–1, 2020.

[20] N. D. Macchiarella, D. Liu, S. N. Gangadharan, D. A. Vincenzi, and A. E. Majoros, "Augmented Reality as a Training Medium for Aviation/Aerospace Application," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 49, no. 25, pp. 2174–2178, 2005.

[21] D. Mourtzis, V. Siatras, and J. Angelopoulos, “Real-time remote maintenance support based on augmented reality (AR),” *Applied Sciences (Switzerland)*, vol. 10, no. 5, 2020.

[22] J. Egger and T. Masood, “Augmented reality in support of intelligent manufacturing – A systematic literature review,” *Computers Industrial Engineering*, vol. 140, p. 106195, 2020.

[23] W. Vorraber, J. Gasser, H. Webb, D. Neubacher, and P. Url, “Assessing augmented reality in production: Remote-assisted maintenance with HoloLens,” *Procedia CIRP*, vol. 88, pp. 139–144, 2020.

[24] L. Hou, X. Wang, and M. Truijens, “Using Augmented Reality to Facilitate Piping Assembly: An Experiment-Based Evaluation,” *Journal of Computing in Civil Engineering*, vol. 29, no. 1, p. 05014007, 2015.

[25] S. Makris, G. Pintzos, L. Rentzos, and G. Chryssolouris, “Assembly support using AR technology based on automatic sequence generation,” *CIRP Annals - Manufacturing Technology*, vol. 62, no. 1, pp. 9–12, 2013.

[26] N. Gavish, T. Gutiérrez, S. Webel, J. Rodríguez, M. Peveri, U. Bockholt, and F. Tecchia, “Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks,” *Interactive Learning Environments*, vol. 23, no. 6, pp. 778–798, 2015.

[27] J. Hahn, B. Ludwig, and C. Wolff, “Augmented reality-based training of the PCB assembly process,” *ACM International Conference Proceeding Series*, vol. 30-Novembe, no. Mum, pp. 395–399, 2015.

[28] W. P. Forrest, M. A. Mackey, V. M. Shah, K. M. Hassell, P. Shah, J. L. Wylie, J. Gopinath, H. Balderhaar, L. Li, W. P. Wuelfing, and R. Helmy, “Mixed Reality Meets Pharmaceutical Development,” *Journal of Pharmaceutical Sciences*, vol. 106, no. 12, pp. 3438–3441, 2017.

[29] L. H. Lee and P. Hui, “Interaction Methods for Smart Glasses: A Survey,” *IEEE Access*, vol. 6, pp. 28 712–28 732, 2018.

[30] C. H. Yu, W. W. Peng, S. F. Yang-Mao, Y. Wang, W. Chinthammit, and H. B. L. Duh, “A hand gesture control framework on smart glasses,” *SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications, SA 2015*, pp. 6–9, 2015.

[31] F. D. Davis, “Perceived usefulness, perceived ease of use, and user acceptance of information technology,” *MIS Quarterly: Management Information Systems*, vol. 13, no. 3, pp. 319–339, 1989.

[32] V. Venkatesh, “Determinants of perceived ease of use : integrating control , intrinsic motivation , acceptance model,” *Inorganic Chemistry Communications*, vol. 11, no. 3, pp. 319–340, 2000.

[33] E. Holdack, K. Lurie-Stoyanov, and H. F. Fromme, “The role of perceived enjoyment and perceived informativeness in assessing the acceptance of AR wearables,” *Journal of Retailing and Consumer Services*, no. November 2019, p. 102259, 2020.

[34] A. Ritu and J. Prasad, “A Conceptual and Operacional Definition of Personal Innovativeness in the Domain of Information Technology,” pp. 204–215, 1998.

[35] K. C. Yang, “Exploring factors affecting the adoption of mobile commerce in Singapore,” *Telematics and Informatics*, vol. 22, no. 3, pp. 257–277, 2005.

[36] R. C. M. Yusoff, R. Ibrahim, H. B. Zaman, and A. Ahmad, “Evaluation of user acceptance of mixed reality technology,” *Australasian Journal of Educational Technology*, vol. 27, no. 8, pp. 1369–1387, 2011.

[37] C. Sagnier, E. Loup-Escande, D. Lourdeaux, I. Thouvenin, and G. Valléry, “User Acceptance of Virtual Reality: An Extended Technology Acceptance Model,” *International Journal of Human-Computer Interaction*, vol. 36, no. 11, pp. 993–1007, 2020.

[38] L. Rebenitsch and C. Owen, “Review on cybersickness in applications and visual displays,” *Virtual Reality*, vol. 20, no. 2, pp. 101–125, 2016.

[39] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, “Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness,” *The International Journal of Aviation Psychology*, vol. 3, no. 3, pp. 203–220, 1993.

[40] “Moverio bt-300 smart glasses developer network,” https://tech.moverio.epson.com/en/bt-300/s, accessed: 2021-01-28.

[41] “Levenshtein distance,” https://en.wikipedia.org/wiki/Levenshtein_distance, accessed: 2021-05-28.

[42] “Levenshtein distance java code,” http://rosettacode.org/wiki/Levenshtein_distance#Java, accessed: 2021-05-17.

[43] “Tensorflow 2 detection model zoo,” https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf2_detection_zoo.md, accessed: 2021-05-5.

[44] Y. C. Chiu, C. Y. Tsai, M. D. Ruan, G. Y. Shen, and T. T. Lee, “Mobilenet-SSDv2: An Improved Object Detection Model for Embedded Systems,” *2020 International Conference on System Science and Engineering, ICSSE 2020*, 2020.

[45] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, 2018.

[46] T. Masood and J. Egger, “Augmented reality in support of Industry 4.0—Implementation challenges and success factors,” *Robotics and Computer-Integrated Manufacturing*, vol. 58, no. August 2018, pp. 181–195, 2019.

Appendix A

Large Tables and Charts

A.1 Speech Recognition Test Results

Word spoken	Model Output	No of times	LD	Recognized?
Next	Next	5	0	Yes
	one axed	1	-	No
All workflows	oh workflows	1	3	Yes
	oh we're close	2	10	Yes
	oh workforce	1	-	No
	how workforce	1	7	Yes
	how workflows	1	3	Yes
Show documents	Show documents	6	0	Yes
Dismiss	Dismiss	2	0	Yes
	this nice	1	-	No
	this means	2	-	No
	nice miss	1	-	No
Start	Start	6	0	Yes
Troubleshooting	Troubleshooting	6	0	Yes
Back to the beginning	Back to the beginning	5	0	Yes
	about to the beginning	1	4	Yes
Search by knowledge area	Search by knowledge area	2	0	Yes
	what search by knowledge korea	1	7	Yes
	search my knowledge area	3	1	Yes
Created by	Created by	6	0	Yes
Voice Commands	Voice commands	3	0	Yes
	why's comments	1	-	No
	where's comments	1	8	Yes
	why comments	1	6	Show documents
View diagram	View diagram	4	0	Yes
	whew diagram	1	2	Yes
	have you diagram	1	7	Yes
Scroll up	Scroll up	4	0	Yes
	screw up	2	3	Yes
Go back	Go back	6	0	Yes

Table A.1: Speech recognition test results

A.2 Gesture Recognition Test Results

Class	Test No	Confidence Values										
		Correctly Classified			Wrongly Classified							
					Four				One			Three
	1	99,9	99,75	99,86								
Next	2	99,81	99,97	99,98								
	3	98,7	99,87	99,9								
	1	99,02	99,84	99,86								
Back	2	99,84	99,83	99,79								
	3	98,7	99,87	99,9								
	1	99,62	99,58									
One	2	98,95	99,49	99,3								
	3	98,65	99,49	99,3								
	1	99,65	99,44									
Two	2	98,04	99,41									
	3	98,47	99,39									
	1	99,07	99,28	99,31								
Three	2	98,2	99,17	99,05								
	3	98,45	98,8	98,65								
	1	95,53	98,45	98,7								
Four	2	98,85	98,37	99,2								
	3	98,81	98,44									98,24
	1				97,93	94,28						
Five	2											
	3				97,19	94,04	94,54	96,33				
	1								98,67			
Closed	2								99,14	98,5	98,23	
	3								99,35	99,37	99,35	
	1											
Pen	2											
	3											

Table A.2: Results for the test made with model 1, with confidence threshold of 90%

Class	Test No	Confidence Values										
		Correctly Classified				Wrongly Classified						
						Four				One	Next	
Next	1	99,98										
	2	99,98	99,97	99,95								
	3	99,99	99,99									
Back	1	99,97	99,98									
	2	99,94	99,96	99,92								
	3	99,98	99,97	99,97								
One	1	99,67	98,33	99,1								
	2	99,7	99,92	99,7								
	3	99,86	99,92	99,91								
Two	1	95,44	99,3	99,5	99,75							
	2	99,88	99,94	99,93								
	3	99,9	99,91	99,92								
Three	1	99,36	99,62									
	2	99,22	97,9	97,81								
	3	97,34	98,8	98,73								
Four	1	99,84	99,86	99,88								
	2	99,78	99,82	99,78								
	3	99,86	99,63	99,7								
Five	1					96,41					98,63	
	2					94,91	96,64	96,42			96,26	95,02
	3					99,65	99,18					
Closed	1									93,78		
	2											
	3	94,14	92,22									
Pen	1											
	2	94,05										
	3											

Table A.3: Results for the test made with model 2, with confidence threshold of 90%

Class	Test No	Confidence Values									
		Correctly Classified			Wrongly Classified						
					Four		One			Three	
Next	1	99,01	99,45	99,26							
	2	99,69	99,61								
	3	99,96	99,97								
Back	1	99,79	99,93	99,94							
	2	99,27	99,6								
	3	99,86	99,79	99,88							
One	1	99,01	98,9	98,58							
	2	98,9	98,76								
	3	99,48	99,52								
Two	1	98,4	98,34	98,2							
	2	98,43									
	3	98,63	98,29								
Three	1	99,67									
	2										
	3	99,23									
Four	1	99,62	99,3	99,43							
	2	99,44	99,12	99,02							
	3	98,82	99,25	99,09							
Five	1										
	2				98,12	98,18					
	3										
Closed	1						99,63	99,51	99,44		
	2						98,92	99,5		98,8	
	3						98,46	98,76			
Pen	1										
	2									98,61	98,07
	3										

Table A.4: Results for the test made with model 1, with confidence threshold of 98%

Class	Test No	Confidence Values					
		Correctly Classified			Wrongly Classified		
					Next		
Next	1	99,93	99,94				
	2	99,82	99,8	99,77			
	3	99,98	99,99	99,97			
Back	1	99,78	99,88	99,95			
	2	99,73	99,86				
	3	99,96	99,96	99,96			
One	1	99,79	99,83				
	2	99,9	99,91				
	3	99,93	99,77	99,8			
Two	1	99,31	99,65				
	2	99,15	99,37	98,96			
	3	99,92	99,93	99,95			
Three	1	98,85	98,31				
	2	98,06	98,54				
	3	98,82					
Four	1	99,45	99,85	99,82			
	2	99,84	99,74	99,69			
	3	99,71	99,56				
Five	1				98,81	98,2	98,17
	2						
	3				99,65	99,53	99,63
Closed	1						
	2						
	3						
Pen	1						
	2						
	3						

Table A.5: Results for the test made with model 2, with confidence threshold of 98%

A.3 Case Study Post-Questionnaire Results

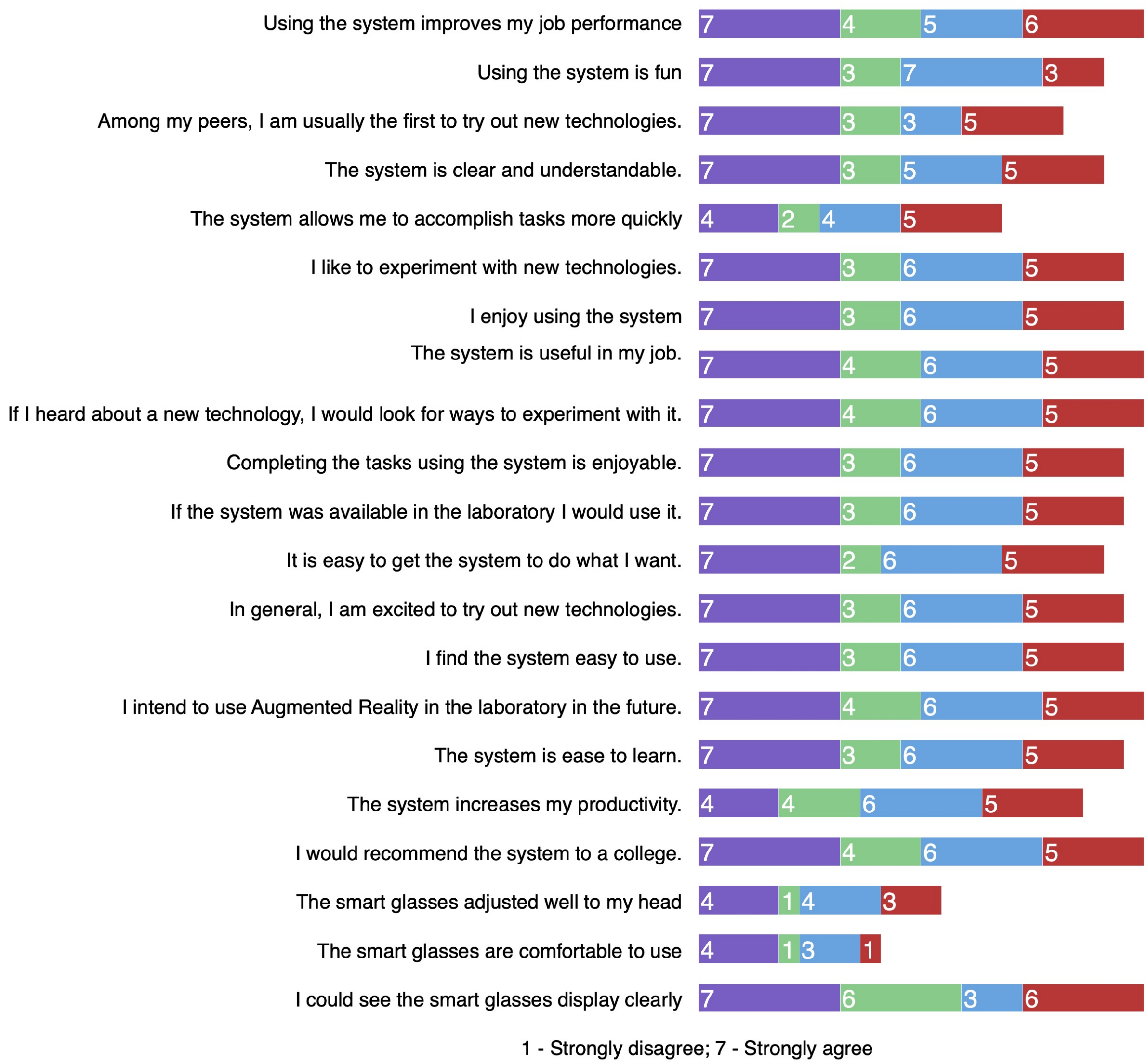

Figure A.1: Responses to the post-questionnaire 7-point likert-type scale questions

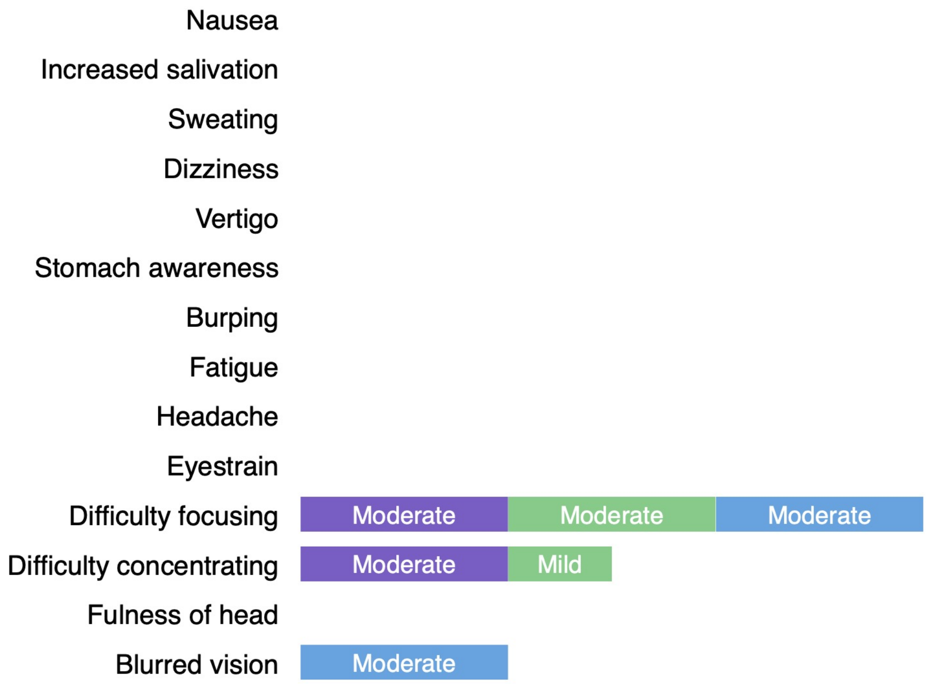

Figure A.2: Responses to the cybersickness questionnaire

Appendix B

Technology Acceptance Model Hypothesis

- **H1** Perceived usefulness has a positive effect on intention to use
- **H2** Perceived ease of use has a positive effect on intention to use
- **H3** Perceived ease of use has a positive effect on perceived usefulness
- **H4** Perceived ease of use has a positive effect on perceived enjoyment
- **H5** Perceived enjoyment has a positive effect on perceived usefulness
- **H6** Perceived enjoyment has a positive effect on intention to use
- **H7** Personal innovativeness has a positive effect on perceived enjoyment
- **H8** Personal innovativeness has a positive effect on perceived ease of use
- **H9** Personal innovativeness has a positive effect on perceived usefulness
- **H10** Personal innovativeness has a positive effect on intention to use
- **H11** Ergonomics has a positive effect on intention to use
- **H12** Cybersickness has a negative effect on intention to use

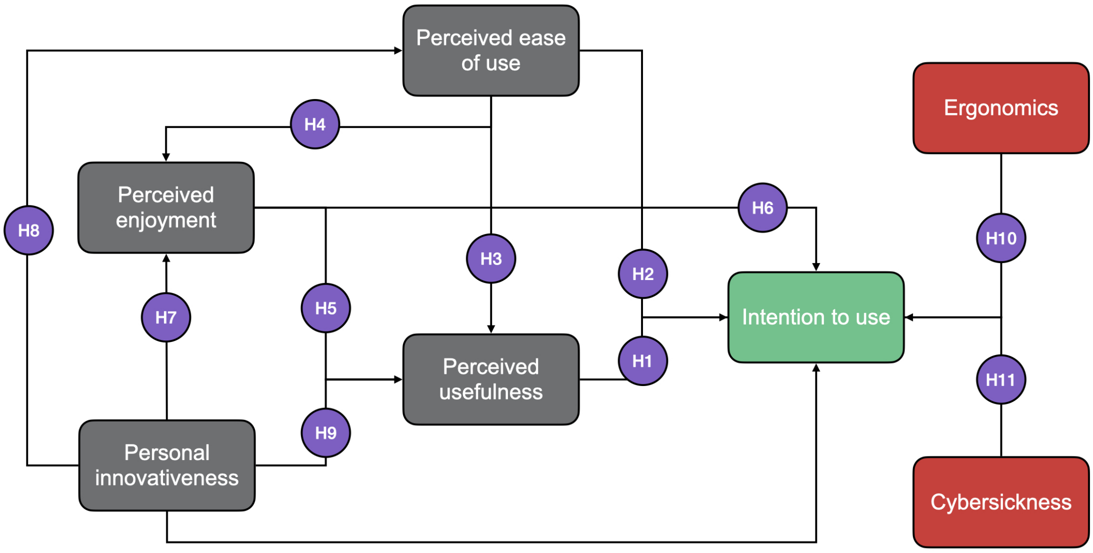

Figure B.1: Hypothesis for the augmented reality acceptance model

Appendix C

Case Study Questionnaires

C.1 Pre-Questionnaire

Augmented Reality in the Analytical Laboratory: Pre-Questionnaire

* Obrigatório

* Este formulário irá registar o seu nome, por favor preencha seu nome.

Introduction

This is a questionnaire to evaluate the user experience using augmented reality smart glasses in Hovione's analytical laboratory.

If you were not a part of this experiment, please do not proceed.

1. Do you agree to the extraction of the non-personal data given in this form for academic purposes? *

   - I agree
   - I do not agree

About you

2. My age is *

- ◯ 18-30
- ◯ 31-40
- ◯ 40-50
- ◯ >50

3. My education level is *

- ◯ Primary school
- ◯ High school
- ◯ Professional degree
- ◯ Bachelor's degree
- ◯ Master's degree
- ◯ Doctoral degree

4. My field of study is

- ◯ Engineering
- ◯ Chemistry
- ◯ Pharmacy
- ◯ [ ] Outro

5. I've been working in my current position *

- ◯ for less than a year
- ◯ between one and two years
- ◯ for more than two years

6. Regarding Augmented Reality (AR) *

- ☐ I've never used AR
- ☐ I've seldom used AR apps on my phone/tablet
- ☐ I regularly use AR apps on my phone/tablet
- ☐ I've seldom used AR with Smart Glasses
- ☐ I regularly use AR with Smart Glasses
- ☐ I've used AR in industrial settings
- ☐ I've used other types of AR
- ☐ I've never used AR, but I've used Virtual Reality (VR)
- ☐ I've never used AR, but I regularly use VR
- ☐ I use both AR and VR regularly

About the upcoming experiment

7. What is your level of excitement regarding this experiment?
(1-very low; 5- very high) *

1	2	3	4	5
◯	◯	◯	◯	◯

8. What are your expectations regarding AR in the laboratory?
(1-very negative; 5-very positive) *

1	2	3	4	5
◯	◯	◯	◯	◯

C.2 Post-Questionnaire

User Acceptance of Augmented Reality in the Analytical Laboratory

This is a questionnaire to evaluate the user experience using augmented reality smart glasses in Hovione's analytical laboratory.

If you were not a part of this experiment, please do not proceed.

The questionnaire will take about 6 min to complete.

* Obrigatório

* Este formulário irá registar o seu nome, por favor preencha seu nome.

1. Do you agree to the extraction of the non-personal data given in this form for academic purposes? *

   - I agree
   - I do not agree

2. You used the system to follow a *

- [ ] Development workflow
- [ ] Troubleshooting workflow
- [ ] Outro

3. What features of the application did you use? *

- [ ] Step-by-step workflow guide (main feature)
- [ ] Diagram view
- [ ] Document view
- [ ] Created/modified by information
- [ ] Outro

4. Did you use the Speech Recognition feature? *

- ○ Yes
- ○ No

5. If you used the Speech Recognition service, please rate the following statements:
1- Strongly disagree, to 7- Strongly agree

	1 - Strongly Disagree	2	3	4	5	6	7 - Strongly Agree
The application responded well to my commands	○	○	○	○	○	○	○
I rarely had to repeat myself	○	○	○	○	○	○	○
I would rather just use the command	○	○	○	○	○	○	○

6. If you did not use the Speech Recognition service, please rate the following statements:
1- Strongly disagree, to 7- Strongly agree

	1 - Strongly Disagree	2	3	4	5	6	7 - Strongly Agree
I did not use the Speech Recognition service because i do not like the idea of speaking to a device	○	○	○	○	○	○	○
I did not use it because I did not want to disturb my colleagues	○	○	○	○	○	○	○
I did not use it because I was embarrassed to use it	○	○	○	○	○	○	○

7. Please rate the following statements.
1- Strongly disagree, to 7- Strongly agree *

	1 - Strongly Disagree	2	3	4	5	6	7 - Strongly Agree
Using the system improves my job performance	○	○	○	○	○	○	○
Using the system is fun	○	○	○	○	○	○	○
Among my peers, I am usually the first to try out new technologies.	○	○	○	○	○	○	○
The system is clear and understandable.	○	○	○	○	○	○	○
The system allows me to accomplish tasks more quickly	○	○	○	○	○	○	○
I like to experiment with new technologies.	○	○	○	○	○	○	○
I enjoy using the system	○	○	○	○	○	○	○
The system is useful in my job.	○	○	○	○	○	○	○
If I heard about a new technology, I would look for ways to experiment with it.	○	○	○	○	○	○	○
Completing the tasks using the system is enjoyable.	○	○	○	○	○	○	○
If the system was available in the laboratory I would use it.	○	○	○	○	○	○	○
It is easy to get the system to do what I want.	○	○	○	○	○	○	○

	1 - Strongly Disagree	2	3	4	5	6	7 - Strongly Agree
In general, I am excited to try out new technologies.	○	○	○	○	○	○	○
I find the system easy to use.	○	○	○	○	○	○	○
I intend to use Augmented Reality in the laboratory in the future.	○	○	○	○	○	○	○
The system is ease to learn.	○	○	○	○	○	○	○
The system increases my productivity.	○	○	○	○	○	○	○
I would recommend the system to a college.	○	○	○	○	○	○	○

Ergonomics

8. Please rate the following statements.
   1- Strongly disagree, to 7- Strongly agree *

	1- Strongly Disagree	2	3	4	5	6	7 - Strongly Agree
The smart glasses adjusted well to my head	○	○	○	○	○	○	○
The smart glasses are comfortable to use	○	○	○	○	○	○	○
I could see the smart glasses display clearly	○	○	○	○	○	○	○