Identifying interaction opportunities for CAD systems within a gastroenterology exam room

Diogo Abrantes

2015
Acknowledgments

I would like to thank all the staff at IPO-Porto for being so kind and helpful at all times.

This work is dedicated to all the people in the VISION group, especially Miguel, Pedro, Ana and Sérgio, for all the patience, encouragement, advice, and laughs.
Abstract

Cancer is one of the leading causes of morbidity and mortality worldwide. Gastric cancer is the third most lethal type, causing more than 700,000 deaths a year. Metastasis, a defining feature and major cause of death, occurs in most of people suffering from gastric cancer. Gastroenterology imaging is an essential tool for this battle, since an early diagnosis typically leads to a good prognosis. With the design and development of computer assisted decision systems able to aid gastroenterologists on the diagnosis process, comes an increase of the performance of endoscopic examinations, with better precision and less error. In this thesis we aim to find interaction opportunities for CAD systems within a gastroenterology exam room. This required a workplan grounded on tried and tested human-computer interaction (HCI) research methodologies, combined with contextual studies in real gastroenterology operating room scenarios, in order to research how we can develop clinically useful interactive systems inside those environments. We identified and modeled all relevant personas (doctor, nurse, patient), with the objective of obtaining accurate descriptions of the user (needs, ability to use technology), as well as modeling the scenario and its restrictions in terms of space and equipment. Action models of relevant medical procedures were also created to better understand the staff routine. With the implementation of a workstation and an extra monitor, we could add a new system working side-by-side with what already exists, while not disrupting current practice routine. We also found out that interactive technologies, such as touch-based devices and voice recognition, are better suited for controlling a system in this environment than gesture or face recognition. An early design of a possible user interface was proposed, which when evaluated, scored positive results.
Resumo

O cancro é uma das principais causas de morbidade e mortalidade em todo o mundo. O cancro gástrico é o terceiro tipo mais letal, causando mais de 700.000 mortes por ano, sendo que a metastização ocorre na maioria das pessoas que sofrem deste cancro. A imagiologia gastrenterológica é uma ferramenta essencial nesta batalha, já que um diagnóstico precoce normalmente leva a um bom prognóstico. Com o design e desenvolvimento de sistemas de apoio à decisão capazes de ajudar gastroenterologistas no processo de diagnóstico, vem um aumento no desempenho de exames endoscópicos com melhor precisão e menos erro. Nesta tese encontrámos oportunidades de interação para sistemas CAD dentro de uma sala de exames de gastrenterologia. Exigiu um plano de trabalhos fundamentado em metodologias de pesquisa provadas e testadas em interacção pessoa-máquina (IPM), combinadas com estudos contextuais em cenários de sala de exames de gastrenterologia, para descobrir como podemos desenvolver sistemas interactivos clinicamente úteis dentro desses ambientes. Foram identificados e modelados todos os utilizadores relevantes (médico, enfermeiro, paciente), com o objetivo de obter descrições precisas dos mesmos (necessidades, à vontade com tecnologias de informação), assim como a modelação do ambiente e das suas restrições em termos de espaço e equipamento. Modelos de acção de procedimentos médicos relevantes também foram criados para entender melhor a rotina do serviço. Com a implementação de um computador e monitor extras, podemos acrescentar um novo sistema a trabalhar lado a lado com o que já existe no serviço, não interrompendo a rotina da prática actual. Também descobrimos que as tecnologias interactivas, como dispositivos baseados no toque e reconhecimento de voz, são mais adequados para controlar um sistema neste ambiente do que controlo por gestos ou reconhecimento facial. Uma sugestão inicial de um possível interface de utilizador também foi proposta que, quando avaliada, teve resultados positivos.
# Table of Contents

Acknowledgments .............................................. v
Abstract ...................................................... vii
Resumo ......................................................... ix
Table of Contents ............................................. xi
List of Figures ............................................... xiii
List of Tables ............................................... xv
Acronyms ...................................................... xvii

1 Introduction .................................................. 3
   1.1 Motivation .............................................. 3
   1.2 Context ................................................. 6
   1.3 Objectives .............................................. 6
   1.4 Contributions .......................................... 7
   1.5 Thesis Structure ...................................... 8

2 Endoscopy ..................................................... 11
   2.1 Procedures ............................................ 11
   2.2 Technology ............................................ 12
   2.3 Teaching and Training ................................ 15

3 State of the Art .............................................. 19
   3.1 Interactive systems in medicine ..................... 19
   3.2 CAD Support Systems for Gastroenterology ....... 20

4 Gastroenterology Room Procedures Modelling ............ 25
   4.1 Methods ................................................ 25
   4.2 Observational study in regular context ............ 27
      4.2.1 Overview ........................................ 27
      4.2.2 Results - User Modelling ...................... 28
      4.2.3 Results - Scenario Modelling ................. 29
      4.2.4 Results - Action Modelling ................. 30
   4.3 Discussion ............................................ 32
5 Viability assessment of technology ............................................. 37
  5.1 Objectives ................................................................. 37
  5.2 Capture and Visualization of Video and Image ......................... 37
    5.2.1 Hardware setup .................................................... 37
    5.2.2 Available outputs and resolutions .............................. 40
    5.2.3 Capture methods and results .................................... 42
  5.3 Gesture Control .......................................................... 43
    5.3.1 Users ................................................................. 44
    5.3.2 Materials and Methods .......................................... 44
    5.3.3 Results ............................................................... 45
  5.4 Voice Control ............................................................ 46
    5.4.1 Users ................................................................. 47
    5.4.2 Materials and Methods .......................................... 47
    5.4.3 Results ............................................................... 47
  5.5 Face recognition .......................................................... 48
  5.6 Touch-based interaction .................................................. 49
  5.7 Discussion ............................................................... 50

6 A Touch-Based Interactive System for Gastroenterology .................. 53
  6.1 Methodology ............................................................. 53
    6.1.1 Analysis/Requirements stage ................................... 54
    6.1.2 Design stage ....................................................... 57
  6.2 Evaluation ............................................................... 62
    6.2.1 Cognitive Walkthrough .......................................... 63
    6.2.2 Heuristic Evaluation ............................................. 65
  6.3 Discussion ............................................................... 67

7 Conclusions and Future work .................................................. 71

8 References ................................................................. 75

9 Appendix ................................................................. 81
List of Figures

Figure 1.1 Incidence and mortality of gastric cancer, produced at (International Agency For Research On Cancer, 2012) ................................................................. 4
Figure 1.2 High definition endoscopy, Narrow-Band imaging, Capsule endoscopy .......... 4
Figure 1.3 Components of a CAD system ................................................................. 6

Figure 2.1 Anatomy of the digestive system and stomach ............................................. 11
Figure 2.2 Components and tools of an endoscope, adapted from (Valdastri et al., 2012) ... 13
Figure 2.3 OLYMPUS EVIS EXERA III VSC, adapted from http://medical.olympusamerica.com/ 14

Figure 4.1 Steps of the Contextual Design Process, adapted from (Soegaard and Dam, 2012) . 26
Figure 4.2 A gastroenterology exam room ................................................................. 28
Figure 4.3 Scenario modelling of the exam room ........................................................ 30
Figure 4.4 Action model of an endoscopic procedure .................................................. 32

Figure 5.1 Hardware specs of the workstation .......................................................... 38
Figure 5.2 Blackmagic Intensity Pro video card ........................................................ 39
Figure 5.3 Specs from the Olympus CV-190 .............................................................. 39
Figure 5.4 A 4-pin S-video to RCA cable adapted with BNC connectors ....................... 41
Figure 5.5 Back of a CV-190 video processor ............................................................ 42
Figure 5.6 Images extracted from the BlackMagic video card ...................................... 43
Figure 5.7 Using a Leap Motion Controller ............................................................... 44
Figure 5.8 Leap Motion Controller Gestures, adapted from www.assetstore.unity3d.com 45
Figure 5.9 Nodal points used in facial recognition, adapted from www.extremetech.com 48
Figure 5.10 Resistive and capacitive systems, adapted from www.extremetech.com ...... 50

Figure 6.1 A Software Development Life Cycle ........................................................ 53
Figure 6.2 User centered Design, adapted from Kellogg S. Booth, Introduction to HCI Methods, University of British Columbia, Canada .................................................. 55
Figure 6.3 Use Case for the system ............................................................................. 57
Figure 6.4 A system model ......................................................................................... 58
Figure 6.5 Login Flow ............................................................................................... 59
Figure 6.6 Sync Flow ............................................................................................... 59
Figure 6.7 Force Sync Flow ..................................................................................... 60
Figure 6.8 Select Case Flow ..................................................................................... 60
Figure 6.9 Capture Process Flow .................................................. 61
Figure 6.10 Norman’s Diagram, adapted from Kellogg S. Booth, Introduction to HCI Methods, University of British Columbia, Canada ................................................. 64

Figure 9.1 Slide 1 ............................................................................ 82
Figure 9.2 Slide 2 ............................................................................ 83
Figure 9.3 Slide 4 ............................................................................ 83
Figure 9.4 Slide 6 ............................................................................ 83
Figure 9.5 Slide 7 ............................................................................ 84
Figure 9.6 Slide 8 ............................................................................ 84
Figure 9.7 Slide 10 ......................................................................... 84
Figure 9.8 Slide 16 ......................................................................... 85
Figure 9.9 Slide 17 ......................................................................... 85
Figure 9.10 Slide 19 ....................................................................... 85
List of Tables

Table 5.1  Results from the Gesture Control Assessment  

Table 5.2  Results from the Voice Recognition Assessment  

Table 6.1  Results from the Cognitive Walkthrough  

Table 6.2  Results from the Heuristic Evaluation
# Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVI</td>
<td>Audio Video Interleave</td>
</tr>
<tr>
<td>BNC</td>
<td>Bayonet Neill-Concelman</td>
</tr>
<tr>
<td>CAD</td>
<td>Computer Assisted/Aided Decision</td>
</tr>
<tr>
<td>CCD</td>
<td>chargecoupled device</td>
</tr>
<tr>
<td>CMOS</td>
<td>complementary metal-oxide semiconductor sensor</td>
</tr>
<tr>
<td>CBIR</td>
<td>Content Based Image Retrieval</td>
</tr>
<tr>
<td>CDSS</td>
<td>Computer Decision Support Systems</td>
</tr>
<tr>
<td>CPU</td>
<td>Central Processing Unit</td>
</tr>
<tr>
<td>DOB</td>
<td>Date of Birth</td>
</tr>
<tr>
<td>DOP</td>
<td>Date of Procedure</td>
</tr>
<tr>
<td>DVI</td>
<td>Digital Visual Interface</td>
</tr>
<tr>
<td>ECC</td>
<td>Error-Correcting Code</td>
</tr>
<tr>
<td>EGD</td>
<td>EsophagoGastroDuodenoscopy</td>
</tr>
<tr>
<td>EHR</td>
<td>Electronic Health Record</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and Drug Administration</td>
</tr>
<tr>
<td>GI</td>
<td>Gastro Intestinal</td>
</tr>
<tr>
<td>GPU</td>
<td>Graphics Processing Unit</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td>HDMI</td>
<td>High-Definition Multimedia Interface</td>
</tr>
<tr>
<td>HCI</td>
<td>Human-Computer Interaction</td>
</tr>
<tr>
<td>IID</td>
<td>Iterative and Incremental Development</td>
</tr>
<tr>
<td>IR</td>
<td>Infra Red</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technologies</td>
</tr>
</tbody>
</table>
NBI Narrow-Band Imaging

NTSC National Television System Committee

PACS Picture Archiving and Communication System

PAL Phase Alternating Line

RAD Rapid Application Development

RAID Redundant Array of Independent Disks

RAM Random Access Memory

RCA Radio Corporation of America

RGB Red Green Blue

S-video Separate Video

SDI Serial Digital Interface

SDLC Software Development Life Cycle

SSD Solid State Drive

TGA Truevision Graphics Adapter

TOP Time of Procedure

UEGF United European Gastroenterology Federation

UGIE Upper Gastro Intestinal Endoscopy

UML Unified Modeling Language

WCE Wireless Capsule Endoscopy
1. Introduction
1. Introduction

1.1 Motivation

Cancer is one of the leading causes of morbidity and mortality worldwide. According to the World Health Organization and the International Agency for Research on Cancer, it accounted for 8.2 million deaths globally in 2012, with gastric cancer being the third most lethal type in both sexes (after lung and liver), causing more than 700,000 deaths (Figure 1.1). There were approximately 14 million new cancer cases that year (950,000 of them regarding the stomach), a number which is estimated to rise by about 70% over the next 2 decades. More than 70% of cases occur in developing countries and half the world’s total occurs in Eastern Asia (mainly in China). The highest mortality rates are also in Eastern Asia, followed by Central and Eastern Europe, and the lowest are encountered in Northern America. Age-standardized incidence rates are about twice as high in men as in women. Also, around one third of cancer deaths are due to the five leading behavioural and dietary risks like high body mass index, low fruit and vegetable intake, lack of physical activity, tobacco use and alcohol use (International Agency For Research On Cancer, 2012; World Health Organization Media Center, 2015). The complicating factor with cancer is that most times people do not know they have it until they start to get symptoms. By the time this happens, it may already have progressed and spread out to other parts of the body. Metastasis, which is a defining feature and major cause of death, occurs in approximately 80 to 90% of people suffering from stomach cancer, with around a 65% chance of survival rate in individuals that are diagnosed in earlier stages, opposing to the less than 15% that are diagnosed in later stages (American Cancer Society, 2015). To make things worse, even patients who present the favorable conditions and who undergo curative surgical resection, often die of recurrent disease (Liu et al., 2012).
Gastroenterology imaging is an essential tool for this battle, since an early diagnosis typically leads to a good prognosis. However, this is a rapidly evolving technological area with novel imaging devices such as the endoscopic capsule, narrow-band imaging or high-definition endoscopy (Figure 1.2).

Adapting to these technologies has a high time-price cost, even for experienced clinicians, motivating the appearance of interactive environments with decision support features that can accelerate training processes, improve diagnostic accuracy, and support technology adaptation periods. With the design and development of systems able to aid gastroenterologists on the decision making process, comes an increase
of the performance of endoscopic examinations, with better precision and less error. One of the reasons for these errors is that human perception and response times are limited at some levels and vary depending on attention, focus and external factors at any given time, changing from user to user, thus justifying the need for technology that can help this process. As information technologies evolve in the most diverse fields, lately there have been strong incentives for them to be used in the healthcare area, in order to provide more reliable, quicker and cheaper services. Besides streamlined storing, organization and sharing of electronic patient records, among other advantages are the improvement of doctor-patient relationships, and doctor-doctor communication (using exclusive social networks, for example), that allow better understanding between peers, as well as easier exchange of information and second opinions. Telemedicine and e-Health are also gaining their space, as patients living in remote areas, with less mobility (or availability) and less income, can have adequate treatment from specialists, without the need to go to a hospital. The explosive growth of the number of mobile apps and devices along with the increased availability of internet access means the patient can now be more aware of how to eat, sleep and live better, and even be part of online support groups that share a certain pathology. Along with all this ease of creation, administration and management of information, interactive systems that could help healthcare professionals on the process of decision making started to appear. Computer Decision Support Systems (CDSS) or Computer Assisted/Aided Decision (CAD) systems try to enhance health care by providing clinicians and other staff with knowledge and person-specific information, intelligently filtered and presented at appropriate times (Osheroﬀ et al., 2007). Examples on problem prevention include pop-up alerts to potential drug interactions when prescribing new medicine, clinical prediction rules to assess risks of pain medications, clinical guidelines for treatment, and follow-up reminders. Proper implementation and use of CAD systems is a key factor towards major improvements in quality and cost of care from the use of health information technologies (Linder et al., 2007). Several stakeholders were identiﬁed in this project, along with a critique of their impact on the adoption of this system. Doctors will have their job made easier in that they’ll have better and quicker access to information, complementing their clinical knowledge and helping on the validation of their opinions, translating into a higher degree of certainty in each diagnosis. Nurses could beneﬁt as well, since they’re acquiring experience and expanding their knowledge in the ﬁeld by assisting doctors who use the system. Interns will beneﬁt in being provided with what could be a unique mean of learning/training which will allow them to acquire skills more quickly and effectively. Productivity gains would be enjoyed by all. Ultimately, the patient also beneﬁts greatly, in that the system could bring about shorter exam duration, and a reduction of medical errors made possible by quicker and more accurate diagnoses. By demonstrating how the beneﬁts of the system outweigh its costs, we can even get the attention of hospital administration boards, as the guarantee of a more expedient and accurate diagnostic system will translate into an optimization of their resources. Nevertheless, the system has to be made in accordance with the needs and opinions of healthcare professionals, and it must be user-friendly so as to be widely and expeditiously accepted in workplaces as a valuable asset. Furthermore, when thinking about the development, one should worry not only about the system itself, but also with how the potential user will respond to, or use it. Interactive technologies can have a crucial role in the promotion of these systems, since new or better ways of interaction could improve the current way of doing tasks, or even provide the possibility of unlocking new ones, bringing an easier, more enjoyable, and more powerful experience.
1.2 Context

A CAD system usually has a type of user and five different components, as depicted in Figure 1.3.

![Diagram of CAD system components]

Figure 1.3: Components of a CAD system

Three of these modules are not contemplated in the scope of this thesis, so we are going to provide a brief description of their functions inside CAD systems:

- The database is a collection of organized data within a structure and it can be accessed by the middleware and the artificial intelligence component;

- The artificial intelligence component is a group of specific algorithms responsible for the analysis and processing of various types of data, in order to provide the user with a second opinion. In the system envisioned for this thesis, we would find in this component algorithms specialized in computer vision and machine learning;

- The controller is responsible for the interoperability and exchange of data between presented components, as well as being designed to support new components that might be added to the system;

The interaction between the user and the other two components, input and output, will be the main scope of this thesis. Assuming that a CAD system is only made up by hardware and software is a common mistake. CAD systems are composed by two major subsystems, such as the human decision makers and the computer system itself. A decision by definition cannot be programmed because its nature and structure are complex. The human decision maker comes into place as a component of CAD systems to exercise its judge or intuition throughout the entire decision making process. The user interface sub-system is the only component with which the user will have to deal. Hence, we need to make it as efficient, versatile and intuitive to the user as possible, taking various issues into consideration, such as choice of input/output devices, data and information presentation formats, among others.

1.3 Objectives

We want to understand if and how these technologies can be implemented in a gastroenterology exam room. There is no reason for us to think that a traditional desktop software provides the best interaction a gastroenterologist could experience. Let us ponder the example of a high-definition endoscopy being
performed in a typical clinical environment: a physician is holding the endoscopy tube in one hand, the control handle on the other while observing the video feed in a monitor on the wall or endoscopy cart. Implementing gesture, face, or even voice recognition with a video-camera, motion control device or microphone, could allow a physician to wave his hands and request for a second opinion regarding a section of the image he is suspicious about, where computer vision algorithms can provide an automatic classification such as the probability of the tissue observed being cancerous, or vocally request for the retrieval of exams that have lesions that may look like one being observed on the screen. There are many possibilities, making it easy to come up with ideas for interesting interaction mechanisms. Not everyone has gone through an endoscopy procedure, but a lot of us know in what one consists. However, that knowledge alone is not enough to implement an additional system that fits in the daily routines of those involved in an exam. A lot of systems (including medical ones) are produced without first defining the specific context in which they will be used. That can lead to failure in a lot of aspects which ultimately results in system rejection. The main objective of this project will require a workplan grounded on tried and tested Human-Computer Interaction (HCI) research methodologies, combined with contextual studies in real gastroenterology operating room scenarios, in order to research how we can develop clinically useful interactive systems inside those environments, which enable doctors to access and produce medical information. Given the exploratory nature of this main objective, addressing it requires addressing a set of more specific sub-objectives:

- Identify needs and interaction opportunities by modelling the users and the environments of gastroenterology operating rooms, in order to identify and model all relevant personas (doctor, nurse, patient), with the objective of obtaining accurate descriptions of the user (needs, ability to use technology), as well as modeling the scenario and its restrictions in terms of space and equipment. Action models of relevant medical procedures will be created to better understand the staff routine;

- Understand the capabilities and limitations of currently available interactive technologies that are potentially suitable, seeking ways of adding to, or modifying what already is being used;

- Design and create low and medium-fidelity prototypes of possible solutions, applying low-cost evaluation methodologies for early assessment, deploying the proposals that exhibit both high feasibility and impact potential;

1.4 Contributions

The key scientific contributions of this work are:

- User, scenario and action modelling of a gastroenterology exam room;

- Assessment of available interactive technologies that could be implemented in a gastroenterology exam room;

- Proposal of the structure and interface for a future computer assisted decision system that could be used a gastroenterology exam room;
1.5 Thesis Structure

The thesis is structured as follows:

- In the first chapter we point out the importance of bringing interactive and decision support technologies that are be able to help doctors in their diagnosis, into gastroenterology exam rooms. This is motivated by gastric cancer being one of the most severe, holding high incidence and mortality rates;

- In the second chapter we explain how a common endoscopic procedure is done, the technology involved in the process, as well as the teaching and training structure of gastroenterology trainees;

- In the third chapter we explore the state of the art of interactive systems in medicine and computer assisted decision support systems for gastroenterology;

- In the fourth chapter we explain how the contextual studies for a gastroenterology procedure were conducted, as well as the results that came from modelling the user, the scenario, and the actions;

- In the fifth chapter we assess the viability of relevant interactive technologies that could be implemented in a gastroenterology scenario. We performed video capture tests, as well as case studies for face recognition, gesture control, voice recognition and touch-based interactions;

- In the sixth chapter we present the methodology in the design and evaluation of a possible user interface for the system in question. User stories, use cases and flow diagrams were produced, as well as the design of a system image;

- In the seventh chapter we discuss the results of this project, while providing several suggestions for future work;
2. Endoscopy
2. Endoscopy

2.1 Procedures

Gastroenterology is a medical specialty focused on the structure, function and associated pathologies of the digestive system. The digestive endoscopy involves the use of flexible tubes with maneuverable ends with a camera, known as endoscopes, which allow for the visualization of a live video feed of the digestive system (Figure 2.1) in a monitor. There are several types of Gastro Intestinal (GI) endoscopic procedures with associated guidelines that help physicians decide the ones more suitable to evaluate the patient at hand. The most commonly requested tests are: the Upper GI Endoscopy or EsophagoGastroDuodenoscopy (EGD), in which an endoscope is inserted through the mouth and affords an excellent view of mucosal surfaces of the esophagus, stomach, and proximal duodenum; the Colonoscopy, in which a low endoscope (colonoscope) is introduced anally and allows examination of the entire colon and rectum, and frequently the terminal ileum.

![Anatomy of the digestive system and stomach](image)

Figure 2.1: Anatomy of the digestive system and stomach

These tests are easily performed on an outpatient basis, and with very low complication rates when performed by a Gastroenterologist with adequate training (Sociedade Portuguesa de Endoscopia Digestiva, 2015; Early et al., 2012). Complications associated with endoscopy are related mainly to sedation, with cardiorespiratory problems being the most common (0.03% to 20% incidence) (Nelson et al., 2012). Bleeding can also occur (0.2% to 2.1% incidence), as well as occasional perforation when excessive forces
are applied to the GI wall (0.1% incidence) (Varadarajulu et al., 2011). Other limitations of current practice include the fact that biopsy samples are processed off-line rather than intraoperatively. If the biopsy is positive, a second endoscopic procedure should take place to treat the pathology, in which the same spot should be located again. Standard diagnostic functions for both procedures include inspection, biopsy, photography, and video recording. Diagnostic observations are made concerning focal benign or malignant lesions, diffuse mucosal changes, luminal obstruction, motility, and extrinsic compression by contiguous structures. Common therapeutic endoscopic procedures include polypectomy, dilation of strictures, stent placement, removal of foreign bodies, gastrostomy, treatment of GI bleeding with injection, banding, coagulation, sclerotherapy, and endoscopic therapy of intestinal metaplasia. Video capsule endoscopy also provides the capability to visualize the GI tract by transmitting images wirelessly from a disposable capsule to a data recorder worn by the patient. Specialized capsules for imaging the esophagus and small intestine are currently approved by the United States Food and Drug Administration (FDA). This type of endoscopy is generally indicated for the evaluation of obscure GI bleeding in a patient in whom upper and lower endoscopy have not identified a cause, screening and surveillance of the small bowel in patients with inherited polyposis syndromes, or suspected small intestinal tumors, since the small intestine (jejunum and ileum) is relatively inaccessible to regular endoscopy (Early et al., 2012). For the purposes of this study, as the collaboration with IPO-Porto would allow us to get more data on this approach, we have focused on the upper GI tract, which includes the esophagus, stomach and duodenum (EGD or Upper Gastro Intestinal Endoscopy (UGIE)). The lower GI tract includes the anus, rectum, colon, and cecum (anoproctosigmoid colonoscopy or lower GI endoscopy).

2.2 Technology

The endoscopes that are used for the examination and treatment of the GI tract consist of three main parts (Figure 2.2) (Valdastri et al., 2012):

- The control handle (Figure 2.2, C), which is held and moved by the gastroenterologist’s left hand and has two stacked control dials that deflect the instrument tip up/down, left/right, and can also lock it into place; separate buttons for air or water insufflation, suction, image freeze/capture (the capture function is normally assigned to a foot-controlled pedal), switching to or activating advanced imaging features, among others, that depending on the model, can be assignable; an entry port for inserting accessories (Figure 2.2, F through J) through the working channel;

- The insertion tube (Figure 2.2, D), controlled by the right hand, is a flexible shaft embedding the channels that allow passage for working tools, and enable suction and air/water insufflation, along with the angulation control wires for tip deflection. The tip of the insertion tube of video endoscopes contains a digital imager [either a chargecoupled device (CCD) or a complementary metal-oxide semiconductor sensor (CMOS)] for color image generation, a light guide illumination system, an opening for the air/water channel, an objective lens and a water jet to clear that lens. The lens may be oriented for different types of views, and the length, diameter, and flexibility of the insertion tube can also be different. They vary among endoscope type, and manufacturer;

- The connector section (Figure 2.2, A) attaches the endoscope to an image processor, a light source with electrical power supply, air or CO2 source, and water (Varadarajulu et al., 2011);
Endoscopes must be cleaned and sterilized between patients, so they are designed to be waterproof and chemical resistant. Several flexible instruments are used with endoscopes to perform diagnostic or therapeutic tasks. These are introduced through the working channel (or channels, as some have dual working channels) of the endoscope and are manually operated by an assistant or a nurse. For example, during a GI examination, the gastroenterology will routinely wish to acquire a tissue sample from visible lesions or areas of mucosa that differ from their surroundings, for histological purposes. Biopsy forceps are commonly used for this procedure in GI endoscopy (Gonzalez et al., 2010). Other instruments are designed to remove or treat polyps, such as endoloops or hot/cold snares (Figure 2.2, K through O).

Most current video endoscopes have a black and white, solid state image sensor called a CCD mounted at the tip of the endoscope, which allows an image to be transmitted to a video processor for monitor display. Light is provided by a 100-300W Xenon lamp. Some light sources also have a 75W Halogen lamp for backup. This signal is converted to a color image by an Red Green Blue (RGB) sequential system, or a color CCD system. The first uses the lamp to arc white light through a rotating RGB filter located between the lamp and the light guide. When tissue is illuminated, the reflected red, green, and blue images are sent through a CCD in the instrument tip and transmitted to the image processor. The second one differs in that a micro-mosaic color filter is mounted over the CCD chip itself. White-light illumination is provided by the lamp, and the reflected images on the CCD surface are processed by circuitry in the image processor. Some image processors incorporate the light source into the same chassis, but other companies prefer to separate these devices into modules. Endoscopes are compatible only with image processors from their manufacturers, and RGB type endoscopes are not compatible with CCD-type image processors. A variety of image displays and monitors are currently available in the market (Varadarajulu et al., 2011). Figure 2.3 shows an OLYMPUS EVIS EXERA III Video System Center.
Standard GI endoscopy is based on visible white-light image acquisition. However, this approach can be insufficient as studies have shown that even experienced endoscopists can miss up to 6% of advanced adenomas and up to 30% of all adenomas when using standard white-light colonoscopy (Hasan and Wallace, 2009; Postic et al., 2002). Also, Barrett’s esophagus and some dysplastic and early neoplastic changes that can occur in specialized intestinal metaplasia may not be readily identifiable using standard endoscopy. There are newer endoscopic imaging techniques that provide the possibility to easily differentiate normal from abnormal mucosa, guide biopsy collection, or in some cases, even eliminate the need to obtain biopsy specimens (Valdastri et al., 2012). High-definition GI endoscopy was a natural step from the increased availability of HD vision sensors for consumer products. High-resolution endoscopes with high-density CCD sensors produce high-magnification images for the detection of microscopic abnormalities in tissue, up to 100 times compared with 30 times of standard endoscopes (Hasan and Wallace, 2009). It does not seem to be a game-changing technology, but however, as costs are reduced, endoscopists will likely adopt this technology for the more visually appealing images produced. Narrow-Band Imaging (NBI) uses filters to narrow projected light to blue (415-nm) and green (540-nm) wavelengths in order to generate images. This type of light enhances superficial mucosal capillaries and surface patterns, as the greater absorption of illuminating bands by hemoglobin causes the blood vessels to look darker. It has been theorized that NBI might improve the assessment of neoplastic lesions, due to the change of shape and density of the microvessels in these lesions (Valdastri et al., 2012). Chromoendoscopy is based on the use of various dyes to colour the mucosa of the GI tract, which enhance subtle changes that are difficult to see. This procedure is indicated for both lesion detection and characterization, and its demonstrated efficacy has resulted in its incorporation into guidelines for surveillance in patients with certain pathologies (Hasan and Wallace, 2009). However, staining remains a time-consuming process and may lead to operator-dependent results.
2.3 Teaching and Training

Although our key focus of this thesis is the interaction with CAD systems, there is another key area that can improve our ability to screen diseases as early as possible: training. Medicine students have contact with Gastroenterology through a course lectured somewhere between the 3rd and the 4th year, depending on their University. They learn about the associated pathologies of the digestive system and also hepatology, which is considered a sub-specialty of gastroenterology. Apart from all the theory, they also watch several live procedures such as upper GI endoscopies, colonoscopies or eec endoscopies, a number of times. This is all done in pre-graduate status. They also learn the so called non-technical skills, which are crucial for better leadership, teamwork, decision making and communication with the patients, as well as with the other clinical staff. Doctors who want to pursue a specialty in Gastroenterology are hired by Hospitals or other health institutions in a 5 year program. Each training program should have an expert endoscopist and teacher who is designated as the endoscopy training director. He will monitor the intern’s acquisition of technical and cognitive skills, as well as periodically review his methodologies and quality of work. Interns in GI endoscopy acquire their skills through a program of hands-on experiential training under the mentorship of the training director. Training follows a natural progression as they learn key principles of anatomy and scope manipulation. They will also practice sedation techniques, learn to tell normal from abnormal endoscopic findings, as well as integrating those findings into a plan of treatment. They will also develop skills for appropriately document all the procedures (Adler et al., 2012). They have to be supervised for the whole residency, but as they get more confident, comfortable and experienced, the supervision will decrease to the point when they are allowed to perform endoscopies alone, calling out the supervisor when a doubt arises or a second opinion is needed. The teaching methods can vary from hospital to hospital, and from doctor to doctor. A good example is the methodology used by Professor Mário Dinis-Ribeiro, head of the Gastroenterology Department at IPO-Porto, who prefers a step-by-step approach. In 10 different stages of a whole procedure, the intern should perform the first, and he will perform the remaining 9. In the next procedure, the intern will perform the first two stages, he will perform the remaining 8, and so on. This incremental approach is designed to solidify and mechanize the procedure. There are virtual computerized models to train endoscopies. Most of them typically consist of a proxy flexible gastrointestinal endoscope introduced into an interface device that transmits movements to a computer that displays images of the upper and lower intestinal tracts. Some reproduce events such as patient discomfort and some even used ex vivo animal organ models to enhance tissue realism (Singh et al., 2014). Teaching endoscopy to novice trainees is a challenge. Despite the increasing use of simulators, endoscopy skills are best developed during clinical experiences with real patients where clinically relevant time constraints apply. A study showed that trainees feel that a more streamlined approach to teaching endoscopy would help them during the early part of their learning curves, and additional endoscopy experience to increase their confidence and help them meet their perceived goals of competence (Cooper et al., 2014). Understanding how this clinical experience is, although with studies more strongly motivated towards the design of CAD systems, may shed light on how to create better interactive systems for teaching and training of young gastroenterologists, overcoming some of the limitations of current simulators and the restriction of requiring real patients for this process.
3. State of the Art
3. State of the Art

3.1 Interactive systems in medicine

Information healthcare systems are now based on a variety of devices that must be both reliable and easy to use, as they exploit technologies such as mobile devices, location and tracking tools, as well as an array of sensors. Designing interactive computing systems that take advantage of the potential of such a variety of devices and contexts to provide reliable solutions to real problems is no easy job. Some of these systems are or will be used by people without extensive training, if any. If nurses, doctors or patients misread the systems or make slips when setting up drug doses, potentially dangerous situations may arise. Therefore, these systems need to be dependable, predictable and consistent. Healthcare is certainly complex. Systems are bought by hospitals or other health facilities in large quantities, and consistency, compatibility and interoperability between systems continues to be a serious issue. Features appear to make devices more useful, yet increasing numbers of features increases risks of feature selection errors during use. Manufacturers are businesses, and commercial pressures do not yet significantly drive dependability, particularly in areas of user error identified after a system design is certified, since certification implies that the design is right and that any resultant harm is the responsibility of the medical practitioners (Blandford et al., 2011). Interactive technologies are usually split into input and output. The first one concerns how the user could give orders or insert data into the system, and the second one on how the system will respond and how information will be presented. Input technologies include not only the typical keyboard/mouse, but especially voice and motion control, and touch-based devices. The latter (which can work as a dual channel), along with TV screens and monitors, provide feedback and appropriate content to the user in a perceptible way. Multimodal interaction can provide the user with increased usability and multiple ways of interfacing with a system. The advantage of using these types of systems relies not only in the fact that they are more efficient, but also in the possible reduction of critical errors. There is an increasing demand for a human-centered system architecture with which humans can naturally interact so that they do no longer have to adapt to computers, but vice versa. Therefore, it is important that the user can interact with the system in the same way as with other humans, via different modalities such as speech, gestures, etc. This kind of multimodal human-machine interaction facilitates the communication of the user of course, but it is still quite challenging from the system’s point of view (Gieselmann and Denecke, 2003). We are challenged to understand what is the best way to understand these requests and deliver this information to clinicians in a variety of contexts. Unobtrusive interaction within clinical scenarios (including gastroenterology) already has some early research. Motion capture using low-cost solutions such as the Leap Motion controller (Mauser and Burgert, 2014), or Microsoft
Kinect (Svendsen et al., 2014) have been tested and explored for other medical fields such as urology (Ruppert et al., 2012) and surgery (Kim et al., 2014). Another interaction paradigm with some research is voice recognition, for both control (Hotker et al., 2013) and reporting (Fox et al., 2013). Besides the Leap Motion controller, the Microsoft Kinect also provides a quick, cheap and easy way of analyzing position and tracking movement. Voice recognition systems are also better nowadays, allowing users to speak more naturally as they dictate commands to a system. Command recognition rates are usually high and remain stable, but the rate of accidental or unintended commands is also high and should be reduced, for both technologies. They could prove useful in:

- Interactive procedures that take place within aseptic environments as opposed to diagnostic situations, which introduce the need to separate sterile from non-sterile contexts and require no-touch navigation. This potentially writes off familiar techniques for information manipulation like keyboards, mice, and even touch screens, as it would violate the premise as well as being rather impractical;

- Procedures that depend on manual dexterity and a correct type of stance and hand-eye coordination, where motion analysis could also be an inexpensive and simple method to teach or improve correct movements, and monitor competence in a non-biased way;

- Dictating reports, a convenience that can save time to physicians on a daily basis;

However, the available gesture control systems have some drawbacks, such as gesture recognition being relatively loose and often resulting in physical fatigue. Also, not any kind of gesture will do, as they have to be easy to learn and intuitive to use. Voice recognition also has its limitations, as it simply cannot handle rushed, run-on, or slurried dictation styles, although the adaptive technology may be able to learn from the user in time. It is necessary to define and discuss the set of commands that are required in each situation, as the collection may differ due to personal user preferences, available hardware, or type of intervention. Feasibility is proven in various types of scenarios, but further testing should be done to assess the real value of the implementation of these approaches within healthcare providing. Moreover, these pilot studies need to be expanded and adapted to gastroenterology operating rooms, from the perspective that these types of interaction will be the best way to deliver the pieces of information that the clinicians need at that moment.

### 3.2 CAD Support Systems for Gastroenterology

Development of CAD systems for gastroenterology have already been reported on various sets of studies that began spurring interest about a decade ago, being bleeding, polyp and tumor detection and classification considered to be the most mature fields (Liedlgruber and Uhl, 2011). A good example for the motivation behind the design of these systems was the introduction of the Wireless Capsule Endoscopy (WCE) and the huge time consuming task brought by the large number of images generated in a single session that needed analysis by a medical expert (Liedlgruber and Uhl, 2011). Decision support systems can sometimes be directed to classification (polyps, lesions, cancer), using previously trained classifiers or feature detectors to support their decisions (Hwang et al., 2007), or focused on Content Based Image Retrieval (CBIR). They usually are differentiated into passive, active or cooperative systems.
(Berner, 2007). The first helps in the decision making process only by collecting data and organizing it effectively, providing no suggestion or recommendation, while the second one actually processes that data and adds explicit suggestions and/or actions. The third one allows the user to modify or redefine the recommendations for actions before sending them back to the system for validation. In case of an automated decision support system, its output is a suggestion of a final diagnosis or additional information for a diagnosis. This output is usually generated without any intervention, potentially allowing for real-time lesion detection while the endoscope is advancing. CBIR systems, on the other hand, have an interactive nature and can present the physician with a number of similar images on demand, or even past exams from that particular patient, allowing for comparisons to be made that facilitate diagnoses. With the advance of new imaging modalities that greatly increase the efficiency of endoscopy, medical experts should get familiarized with these new systems as they can serve as an expert training tool to predict or detect pathologies or as an educational resource, proving to be a great asset on the exam room. Currently, the biggest challenges to these systems’ implementation include the detection and handling of image degradation (reflections or sensor noise), a lack of robust features to detect and classify different pathologies properly, and finding regions of interest in an automated fashion. Also, the available hardware still imposes some limitations, which leaves room for improvement as well (Liedlgruber and Uhl, 2011).

All in all, even with the advance of new imaging modalities that greatly increase the efficiency of endoscopy, medical experts should get familiarized with these new systems as they can serve as an expert training tool to predict or detect pathologies or as an educational resource, proving to be a great asset on the exam room. From a different angle, the way doctors integrate and interact with these systems is very important, as it will often determine its embrace or rejection, and any gains in productivity will hinge on how comfortable they are with it.
4. Gastroenterology Room
Procedures Modelling
4. Gastroenterology Room Procedures Modelling

4.1 Methods

We hold it fundamental to understand the abilities and limitations, the routines, and the desires of users, and use that knowledge to design our system. Inside the User Studies, the ethnographic approach consists in the prospective study of users in their workplace so as to discern their socio-environmental requirements by means of contacting with them, while interfering as little as possible. This study provides us with detailed descriptions of work practices which might become broken with our system, as well as underused/obsolete practices which might become useful once again. A good understanding of current practices translates into a greater ability for predicting the impact of new methods, or a redesigning of existing ones. We can consider the study as ‘full ethnographic’ when there is the time and opportunity available to be at all times with the users in their environment, and actually becoming one of them. When this isn’t possible, a lighter version of this approach, called contextual studies (and sometimes referred to as contextual inquiry), can be pondered. In this pre-design phase, we used a contextual design method (Beyer and Holtzblatt, 1997; Holtzblatt, 2001) (yet another ethnography-based approach), which is a user centered design process that consists of a structured method of gathering and presenting data from field studies, study workflows and develop human computer interfaces. The data will be used to create and prototype concept, while iteratively testing and refining those concepts with the users. By focusing on the user, we gave a top-down approach to our systems engineering, listing their necessities and difficulties in order to obtain a wide-ranging understanding of how the work is performed currently (fix what’s broken, keep what’s good), and the problems that come with it (inefficiency, frustration, confusion, lack of functionality). In any case, care is required in involving users with the development of the project (participatory design), in that there are certain factors to consider:

- Many projects are not very long and users have little time to spare;

- There is a high percentage of users with a conservative bias in regards to current practices;

- A lack of IT knowledge/skills might prevent them from effectively articulating what they really want in a system, or drive them to ask for unrealistic or hopelessly difficult function implementations. Thus, the solution was to consider them as passive participants which can complement the limitations and intuition of the designer, by bringing in real-world user experience;
Within the contextual design method, we went through a few phases relating to understanding the user's work. Contextual inquiry comprises a mixture of observation and an informal interview. By observing the user going about his work, one gets a more accurate notion of his reality than one would by interview alone, which might lead to an excessively abstract, and consequently 'idealized' model of the user. We have also the possibility to observe relevant situations in which the user might not have considered, if these situations are part of his routine. By being there we manage to not just complement our data with information which the user would be unlikely to impart by means of description, but also to gather performance indicators, speed issues, errors, learning curves, among others. As written above, the observation should be complemented by a personal interaction with the user, or an interview (structured, unstructured or semi-structured), depending on the information needed. These are normally conducted so as to discern the users' opinions, particularities, and responsibilities. A rapport ought to be established with them, keeping the interaction focused on what matters, sharing control and ideas as they come. The user may even expand or correct the researcher's understanding of the observation. Additionally, other techniques such as focus groups can also be used. Questionnaires may also be posed to the users in order to better understand the general opinion on various matters, and to save time on issues that only need a quick, direct answer.

These are just the first steps of the Contextual Design process. There is a lot more to it, as we can see in Figure 4.1:

Figure 4.1: Steps of the Contextual Design Process, adapted from (Soegaard and Dam, 2012)

It is mostly divided in two major phases: where the requirements and solutions are found, and where the concepts are defined and later validated. Simply put, after the contextual inquiry, a work modeling is done in order to capture the routine of individuals in diagrams, providing different perspectives on how things are done (flow, capture, sequence, physical, and artifact models); data consolidation allows us to see common patterns and structure without losing individual variation. Affinity diagrams can bring together issues and insights across all users to reveal the scope of possible problems. The work models previously mentioned can also be consolidated, in order to reveal common strategies and intents while
retaining and organizing individual differences; the redesign phase, sometimes called visioning, is all about using the consolidated data in order to improve work by using technology to support new practices. It includes the system, its delivery, and support structures to make them successful, being intentionally rough and high-level (setting a possible design direction) without going into much detail. This enables the team to see the overall structure of the solution and ensure its coherence; storyboards describe how users accomplish tasks in a new system. The User Environment Design captures the floor plan of that new system, showing each part of it, how it supports the user’s work, exactly what function is available in that part, and how the user gets to and from other parts of the system; finally, testing is an important part of any system development. It is generally accepted that the sooner problems are found, the less it costs to fix them. That way, it’s important to test and iterate a design early on (horizontal prototyping), ahead of investing time writing code (vertical prototyping). Paper prototyping develops rough mockups of the system using notes and hand drawn paper to represent windows, dialog boxes, buttons, menus, and the other user interface elements the user will need. Once the structure and interaction design are largely stable, the team can develop and test interaction and visual design options with users (Soegaard and Dam, 2012).

4.2 Observational study in regular context

The results of this section are heavily based on an article already published at the Medical Informatics Europe conference of 2015 (Abrantes et al., 2015):

4.2.1 Overview

This study was conducted at the Gastroenterology department of the Instituto Português de Oncologia (IPO) Porto, comprising 6 physicians (2 of which were interns), 6 nurses and 8 operating technicians. Beforehand, we reviewed some basic principles of gastroenterology and familiarized ourselves with gastric cancer etiology as well as with the basic procedure of an endoscopy. Once all permissions were given, we conducted our observational study of routine non-sedated endoscopy procedures, which make up for the vast majority of the hospital’s total. They are done by utilizing two distinct analysis systems, according to availability and case nature. The OLYMPUS EVIS EXERA II Video System Center, which consists of a CV-180 image/video processor (acquires and processes images and video) and a CLV-180 light source (regulates lighting according to the approximation to the organs’ walls); the OLYMPUS EVIS EXERA III Video System center, consisting of a CLV-190 image/video processor and a CLV-190 light source. This one features its own monitor in the cart itself, which makes it easier to be moved from room to room. We witnessed 10 endoscopies total, in two separate days, 7 having been conducted by interns and 3 by attending physicians. We wrote down the exams’ protocol, as well as the tasks of every worker involved in them. These observations familiarized us with the methodology and praxis of gastroenterology exams, and we further complemented and consolidated our data by conducting semi-structured, informal interviews with 4 physicians (one of them an intern) and 3 nurses, all from the department. Additionally, a small questionnaire was posed to every potential user (except one doctor) in the department with a view to determine their aptitude for Information Technologies (IT), along with their age group and years of practice (Appendix). We also took pictures of the empty room and its equipment so we could have
a concrete idea of the positioning of everything in the environment and study it later without having to disturb the daily routines of the service (Figure 4.2).

![Image of a gastroenterology exam room](image)

Figure 4.2: A gastroenterology exam room

### 4.2.2 Results - User Modelling

The first identified user of our system is the Doctors, responsible for performing the endoscopy and associated procedures, providing a diagnosis and producing a report of the exam. The second identified user is the Nurses, responsible for patient care and support, as well as assisting the doctor on all procedures. The operating technician and the anesthesiologist (when required by the exam), although part of the daily practice and proper functioning of the routines, pose no active part in the use of our system because the first leaves the room as soon as the patient consent is signed, and the latter isn’t normally there because most of the exams at IPO are non-sedated. Although, when they are present, their degree of attention to patient sedation is very high, leaving no room for handling other equipment. Besides observing the users in their natural habitat, the questionnaire we posed to 11 health professionals working in the department (5 gastroenterologists, 6 nurses), along with the interviews, helped us characterize the type of users we were dealing with. From the questionnaire, we learned that

- 73% of the users have had IT-related training;
- 100% had computers at home and used the Internet frequently;
- 91% had smartphones and 91% had a tablet;
- 100% admitted that most of their work is done electronically;

In an age range between 27 and 48, and years of experience between 3 and 18, we believe it is safe to say that our potential users are highly aware of modern interactive technologies and are comfortable in using them. From the interviews, we learned that users are comfortable with the various software packages
they currently use for different tasks (Glintt HS\(^1\), SiiMA Gastro\(^2\)), but at the same time are able to point out flaws and improvements that they would like to see, suggesting feasible ideas. They are receptive to new implementations and we believe them capable of adapting to a new system with a soft learning curve. Nurses admitted they could have more power within an endoscopy exam, and that the addition of a simple task should not diminish the ability of providing the same degree of attention to the patient, as well as the ability to assist the physicians. The latter admitted they did not have any major concerns delegating an extra task on the nurses, since it would bring productivity gains for all parts involved and ultimately would also benefit the patient. All this information guided us into aligning some remarks:

- Main user: nurses can be potentially marked as the main user of our system, as neither them nor doctors have a problem with it. Nurses were and are keen to learn, and contact with new technologies along with more participation in the procedure would value even more these professionals. Moreover, the communication between doctor and nurse would be forced to improve, as the one who has final say in all decisions would have to use the nurse as a proxy for system control;

- Technology: all users deal with IT and electronic devices on a daily basis. A mobile device can be a viable solution as the potential users are competent with this type of technology. Being the case, we can suppose that the learning curve would not be that steep and its introduction would be well accepted;

- Conception: both doctors and nurses (but mostly nurses) face new implementations with moderate enthusiasm, all of them emphasized that even if the procedure is more conventional, the degree of attention to the patient does not decrease. A big red button kind of design should be the way to go, focusing on simplification and not leaving too much to the user’s imagination, as they don’t have a lot of time to dispose in extra tasks;

### 4.2.3 Results - Scenario Modelling

Although we did all scenario modeling in one of the two available exam rooms, we found that the other room layout is similar enough for results to be valid for it too. In addition to the pictures we took, the environment analysis during endoscopy observation resulted in the following diagram depicted in Figure 4.3:

---

\(^1\)www.glintt.com  
\(^2\)www.first-global.com
Patients leave their belongings in the space located at the bottom left corner and take position 1, lying on their side on the gurney;

The attending physician or the intern conducting the procedure occupy position 2, facing the patient;

A nurse takes up positions 3, 4 or 5, determined by a number of factors:

- personal taste, which dictates the possibility of remaining in any of the three, knowing that the CV-180 system (behind the physician) is in use;
- position 3 or 5 if the CV-190 mobile system is in use and taking up position 4;
- position 5, mandatory, regardless of the system in use when the exam requires specific assistance to the physician, such as biopsy collection and handling;
- positions 4 and 5 are only easily distinguishable if the CV-190 is in use;

As the exam progresses, the doctor observes and analyses the image feed on the monitor positioned at the top left of the diagram (part of the CV-180 system) or on the one included in the CV-190 cart. Both use proprietary RGB cables to output the feed, also with the option to choose from Digital Visual Interface (DVI), Separate Video (S-video) or firewire formats. We can directly intercept the feed that comes out of each one of the processors to an additional computer (with a compatible video card), where our system will run. Its output would be on an additional monitor placed just beside the one that is already there. A small tower placed near a high speed internet connection would provide us with the necessary resources for high resolution image and video processing. Regarding system control, having for example the option of using a tablet via Bluetooth or wireless, it would be possible to have it on the counter ready to be used whenever necessary.

### 4.2.4 Results - Action Modelling

Before the exam, the physician and the intern consult the request for the next procedure, as well as a variety of patient information: demographics, clinical history, risk factors, comorbidities, current medication, etc. Doctors at IPO-Porto have expertise in diagnostics relating neo and metaplasia, but
the procedures themselves are quite standard and can be performed by any gastroenterologist. After interviewing the doctors and the nurses we learned that the average number of endoscopies performed by each doctor at IPO, per week, is 13. This number also corresponds to the average of endoscopies that each nurse assists in the same period of time. Once the patient enters the room and signs the consent form, the nurse’s jobs are to put the patient at ease and explain the whole procedure (especially to first timers), administer Simethicone orally (an anti-foaming agent that reduces bloating and pain), spray Lidocaine on the patient’s throat to numb it, and put him a bite guard that allows the safe passage of the endoscope tube. All healthcare professionals perform the exam standing, and use gloves (and sometimes masks) during the whole process, that are discarded in the end. The lights are dimmed, the patient lies on his left side on the gurney and the examination begins with the insertion of the endoscope tube, and subsequent visualization of all stomach regions and duodenum. As it progresses, the professional can take multiple pictures, regular and NBI, plus saving the video on demand to help with the diagnosis. The use of both hands is extensive, with the right one guiding the tube and the left one on the controller. This controller has several buttons, depending on the model, to which are pre-assigned functions at the doctor’s will, such as freeze frame, image save, switch to NBI (when available), among others. Extra functions may be available at a press of a pedal, right beside the physician’s foot. The nurse will divide her actions between patient care (stability, comfort, tolerance to the exam), talking to them and trying to ease the discomfort, and physician assistance in case biopsies or other procedures are needed. During biopsy collection, both the nurse and the doctor have their hands full. Should it be the intern to perform the examination, the physician will validate his decisions and give suggestions. The endoscope is withdrawn and the endoscopy comes to an end, having an average of 5 to 7 minutes, depending on patient tolerance, need for biopsies (or difficult access to biopsy target), or case severity. The lights are switched on and a casual conversation takes place, as an appointment for a follow-up consultation may be scheduled. The patient may not be fit to be discharged right away, even without sedation, so he may sit on the gurney for a few minutes. He then will be accompanied by the nurse or an operating technician to the exit most of the times, but can also go by himself sometimes. The physician or the intern will then check the exam on the computer, while the operating technicians clean the equipment. They usually prepare the report shortly after, depending on the ease of diagnosis (need of a second opinion) or available time. Some doctors exchange impressions with nurses regarding the elapsed exam, consulting their ruling on a possible diagnosis. This practice appeals to nurses, who feel more valued. Regarding abnormalities that may occur during an endoscopy, there is equipment malfunctioning, lack of communication between the doctor and the nurse, or a major non-tolerance of the patient to the exam. In any case, although possible, these are rare events. Figure 4.4 depicts an action model of an endoscopic procedure.
After exchanging impressions with the doctors, we concluded that there were no significant differences in the general procedure for endoscopies with sedation, so that wouldn’t become a problem when thinking about system implementation. We didn’t study noise or lighting levels, but as the rooms are quiet and the exams are always performed at low natural light, we could speculate that these could be two positive factors if the use of voice recognition and gesture control were to be contemplated.

4.3 Discussion

An endoscopy, being a complementary exam that is part of a patient’s complete diagnosis, is a fairly standardized procedure, for which there are guidelines and processes that work well, without any major problems. However, an improvement on clinical decision support to medical experts in the examination room, as well as in supporting the learning of interns, ultimately translating into patient and institution benefits, would be viewed favorably. Physicians’ suggestions were taken into account, such as the possibility of complementing the endoscope’s freeze function with an extra pedal, to transfer the images to our system or another database in real time, or the attempt to access various information on the patient, currently spread across different systems. Everything we observed, in addition to the interviews, helped us to better understand which steps we have to take from here on out, taking into account functionality and usability with equal concern. Nevertheless, it is imperative that the commands for system control are as simple as they can get, so the professionals won’t lose more time than the bare minimum.

Regarding system interaction, being on the field helped us to put our initial interaction ideas into perspective:

- Facial recognition seemed inadequate at a first glance, since the environment does not allow for a camera to be put at a reasonable distance for it to work. Plus, the fact that the exam is performed at very low light could make the process of recognition difficult or impossible for this type of technology;
• Voice control use has to take into account one thing: the IPO-Porto percentage of exams in which
the patient is not sedated is high. Could the discussed issues and commands dictated to the system
be uncomfortable or unsuitable (even in ethical terms) to have in front of a conscious patient, or
would the commands not differ much from regular in-exam professional communication? If the first
becomes a problem, this option may still be interesting to endoscopic modalities that use sedation
most of the time, or to training environments in which these requirements might not be so strict;

• Gesture control will probably have to be discarded for in-exam use since the doctor spends prac-
tically all of the time with both hands full, as the degree of attention that the nurse must have with
the patient does not facilitate the use of this technology;

• A mobile device such as a tablet may be the most viable (and reliable) option, compared to the
others. Regarding the conclusions on gesture recognition, it would be controlled by the nurse at the
physician’s request. The addition of this task could not diminish the ability of providing the same
degree of attention to the patient, as well as the ability to assist the physicians. The latter admitted
they did not have any major concerns delegating that extra task on the nurses, and these proved to
be receptive in experimenting it, since it would bring productivity gains for all parts involved and
ultimately would also benefit the patient;

Despite being considered, one cannot say that all technologies are sufficiently mature to the point
of using them without any kind of flaws, such as reception problems, noise interference, complex initial
training, or even plain unsuitability for this particular environment. However, we should consider all
the options valid until they are rendered too difficult or useless to implement. Unfortunately, we didn’t
have the chance to observe endoscopies in a teaching environment. From what we were told, the gas-
troenterology department of IPO-Porto sometimes hosts procedure observations for students from the
Faculty of Medicine of the University of Porto. The crowded room (which is even more crowded if there
are anesthesiology students) makes an ethnographic study more difficult, but it would be interesting to
know the thoughts and opinions of those attending these observations. This way, we could understand if
and how pre-grad learning could be improved via interactive systems. As for interns, practice will make
perfect in due time. Nevertheless, apart from technical skills, the implementation of computer assisted
decision or interactive systems (both in-exam and post-exam) could result on a learning curve decrease,
or even a more streamlined way for them to learn, theory wise. This would be a valid challenge, since
the United European Gastroenterology Federation (UEGF) organized a consensus-based strategy meet-
ing and decided they should focus in future activities that included designing new educational training
methods and providing appropriate assessment tools for measuring learning impact, and the development
of training tools to increase competence in finding and evaluating new research for use in clinical practice
and to support clinicians who wish to implement integrated GI care (Berberat et al., 2010).

In the next chapter we case studied all interaction possibilities so we could claim with some degree of
certainty the viability or not of these options, by ascertaining the functionality of these technologies in
the study context.
5. Viability assessment of technology
5. Viability assessment of technology

5.1 Objectives

After analyzing the previous results, we learned that we can only influence to a minimum the system/routine that already exists, and anything we want to implement cannot decrease the current efficiency of patient care. There are a few windows of opportunity, but our research hints that the interaction with any kind of CAD systems cannot easily be done with the contemplated technologies. Two things become essential:

- Identify the available technology already present in gastroenterology exam rooms. Understanding how does it work, and what can we get from it, could guide us in deciding how would it fit with outside hardware we want to introduce in the room;

- Conduct case studies of the interactive technologies at hand, so as to better ascertain their functionality in the study context. We need to take into account all work and action flows that are part of the professionals’ daily routine in order to assess implementation viability, and the value they could bring to the exam room;

5.2 Capture and Visualization of Video and Image

Regarding video capture, there are two parts involved in the process: a PC/workstation with a compatible video card, and the endoscope. Our objective in this section was to understand if and how we could extract the video feed from the endoscopes’ video processors that are used in the department, stream it, and store it for posterior processing. We should be able to do it parallel to the default video stream that goes to the exam room’s monitor(s) and desktop PC. The best quality image and videos have, the better, so we should perform some tests regarding connection type, data transmission and frame rate stability, using the endoscopes’ outputs with the inputs of the machine which will likely run the system.

5.2.1 Hardware setup

There’s not another common medium as data-heavy as video, and we should be certain that our hardware is up to the load. A workstation class machine will ensure we get the stability and performance demanded from the system, as the internal workings are held to a higher standard than those of a PC. They commonly have features like Error-Correcting Code (ECC) Random Access Memory (RAM), multiple processor cores (more cores are preferable to higher clock rates), Redundant Array of Independent
Disks (RAID), Solid State Drive (SSD), and optimized Graphics Processing Unit (GPU) (to help take some of the load off the Central Processing Unit (CPU)). The most important thing when building a computer for video processing is balance though. The performance will only be as good as the weakest link in the chain, and this chain has 4 key links: motherboard, CPU, GPU and RAM. If either of these four points is below the other 3, you will bring down the overall performance of the workstation, and if one of the components is significantly higher performing than the other three you can be wasting money on it and creating bottlenecks in other areas. Video is streamed from disk, so a SSD should also be used, as it will greatly increase boot times, application load times, and overall system performance. Even with the scarce space to exam rooms usually have to spare, the option should always be a mid-size or full tower. The added space comes in handy for when adding in future components, but the bigger case makes running cables and the installation easier. This also provides for better air circulation, which keeps the system cooler. Below in Figure 5.1 are the specs of the machine we used to run the tests:

![Operating System](image)

Figure 5.1: Hardware specs of the workstation

Our setup also features a BlackMagic\(^1\) Intensity Pro video card (Figure 5.2). The card has two High-Definition Multimedia Interface (HDMI) input/output ports, and another one for an analog breakout cable with support for National Television System Committee (NTSC), Phase Alternating Line (PAL), component, composite or S-video. We need this to complement the GTX560 graphics card because it is specialized in high-performance video capture and compression, and enables a quick pass-through of the video from one place to another.

\(^{1}\)www.blackmagicdesign.com
The gastroenterology service of IPO-Porto features several video endoscopy system centers. From older models, such as the OLYMPUS EVIS EXERA II CV-165 series, to the top of the line OLYMPUS EVIS EXERA III CV-190 series, passing through the OLYMPUS EVIS EXERA II CV-180 series. These systems can be used interchangeably, according to availability and case nature of the scheduled exam.

We dealt mostly with the CV-180 and 190 systems. The latter was vastly preferred, because in terms of overall quality, it brings better focus, better NBI quality, better image processing with reduced noise and halation, and a new type of freeze that works a bit like a car’s anti-lock braking system. The freeze action is divided into several micro-freezes, and the sharpest frame of them all is then selected and returned to the user, saving time. Concerning signal output, as far as our purposes go, the 190 adds a DVI extraction option, which can be turned into HDMI using a simple converter. Below in Figure 5.3 are some specs taken from the CV-190 tech sheet.
5.2.2 Available outputs and resolutions

The available outputs from the endoscopes are RGB (or Component/YPbPr), S-video, Composite, Serial Digital Interface (SDI), DVI and Firewire (IEEE 1394).

Component video is a video signal that has been split into two or more components. This term is commonly used when referring to analog video that is transmitted as three separate signals. RGB is a simple color component of Red, Green, and Blue. It usually requires additional conductors for video display synchronizing. When discussing component video, the YPbPr term is often used interchangeably, although their color channels are coded differently. It is designed to be used in analog systems and was developed to save on cable bandwidth requirements and provide enough separation of the signal to provide a quality image. Another term, YCbCr, is sometimes credited as being the same as YPbPr, but there is a small difference in usage, as YPbPr is generally used for analog signals, and YCbCr for digital signals.

S-Video stands for separate video. However, professionals prefer the name Y/C video rather since it is more clearly description of the signal format. This technology is used for sending analog video signals after separating them into two different signals: the S-Video connection keeps the all-important black and white (Y) information separate, and combines the colour difference signals into a single colour signal (C), which results in two separate signals going to the display device, instead of the three that component video provides. Combining the two colour signals results in a minor degradation of the colour information, but we still can get a very good picture from this signal. The connection has 4-pins, being two for ground, one for brightness (Luma or Luminance), and one for the color information (Chroma or Chrominance). When these signals are sent to a receiver such as a television set, it results in sharper images than composite video, but they also don’t support HD video signals.

Composite video is the format of an analog signal, where video information is transmitted as one signal over one wire or a single signal. It is also often called the CVBS, which is an acronym for Colour, Video, Blank and Sync. Composite video cables do not carry audio and do not have high definition video support. As the name suggests, it is a composite signal from three different sources called the Y, U and V. In this case, the Y represents luminance, and the U and V carry the hue and saturation, which together constitutes the chrominance. This way, U and V together carry the information on the color signals.

S-Video and composite video mix the signals together by means of electronic multiplexing, causing degraded signal when the display is not able to separate the signals properly. The advantage of YPbPr over these, like other forms of component video, is offering enough separation of the signals so the quality of the final video is quite identical to the original signal.

SDI is a family of video transmission interfaces that uses YCbCr encoding that is serialized at very high speed. The signal is distributed on a BNC (Bayonet Neill–Concelman) connector, and the same coaxial cables used for Composite and RGB signal can be used. Usually, depending on the video card input, an RCA (Radio Corporation of America) plug (sometimes named phono connector, or A/V jack) is adapted to the BNC connector. Figure 5.4 depicts the cable we used when connecting the endoscope’s video processor to our machine.
This makes it a popular signal since the same cabling can be re-used and the quality is digital on long distances. Cable length can actually be a problem in other formats. In S-video, color shifting can appear if the cable is too long.

DVI was created to convert analog signals like VGA into digital signals to accommodate both analog and digital monitors. It has been superseded by HDMI in televisions, but still has wider acceptance in the PC industry. Both DVI and HDMI deliver the signal in a digital format and are also based on similar specifications, since the latter was derived from the first.

The following video resolutions and their associated compression standards can be obtained from the endoscope models used:

- 1920×1080p@60Hz from RGB/Component/YPbPr;
- 720×576i@50Hz or 720×480i@59.94Hz from S-Video;
- 720×576i@50Hz or 720×480i@59.94Hz from Composite;
- 1920×1080p@60Hz and several others below, from SDI;
- 2560×1600p@60Hz or 3840×2400p@33Hz from DVI;

Although the endoscopes have no direct output for HDMI, the cables support resolutions of 1080p and beyond, including advanced display technologies such as 4K, 3D, and Deep Color.

Both RGB and SDI outputs are usually booked for external main monitor use (and generally use proprietary cables), since they produce the highest quality. With the CV-190 we had the option of using DVI out to HDMI in on the BlackMagic video card. That would allow us to extract video at least at 1920x1080p@60Hz, which is the endoscope’s native output rate.
5.2.3 Capture methods and results

We connected the CV-190 video processor to the BlackMagic Intensity Pro with a DVI to HDMI cable. Using the software from the same company (BlackMagic Media Express\textsuperscript{2} 10.3.5) for recording (Figure 5.5), we were hoping to get a stream at 1080p, or at least 720p. After several experiments, we couldn’t get the video to get through, for no apparent reason. Even at different refresh rates, lower resolution, and with different cables. Updated versions of MediaExpress (10.4.0/10.4.2) that now include automatic input resolution detection, were also of no avail. We got the same results when we tried to use other video capture utility software such as VirtualDub\textsuperscript{3} or VLC Media Player\textsuperscript{4}.

![Figure 5.5: Back of a CV-190 video processor](image)

Given these limitations, we decided to study alternative video resolutions and standards. We used the video card’s breakout cable input, and the endoscope’s S-Video and Composite outputs successfully, capturing video in PAL format at 720x576p@50Hz with the first one. Image capture to the exam room’s desktop, where the doctor produces the report, is done by a composite output and a Pinnacle Dazzle video card. So, interestingly enough, our captured image happens to have slightly better quality than the ones that are used on actual reports. We can speculate that other video cards (maybe a Matrox\textsuperscript{5} MXO2 LE, which has SDI input) might be able to overcome the HD capture limitations found. Given time limitations, this is left for future work.

Using the BlackMagic Media Express, images are stored in the Truevision Graphics Adapter (TGA) lossless format, weighing around 1.6Mb per image (Figure 5.6). Videos are kept in Audio Video Interleave (AVI) format, weighing around 1Gb per minute. It is a lot by today’s storage standards, but a morning’s work consists of about 8 endoscopies. With a mean time of 6 to 7 minutes per procedure, if all video was recorded, that would still leave us with enough disk space for maybe a whole week of procedures,

\textsuperscript{2}www.blackmagicedesign.com/products/intensity/mediaexpress
\textsuperscript{3}www.virtualdub.org
\textsuperscript{4}www.videolan.org/vlc
\textsuperscript{5}www.matrox.com
depending on how busy the department could be. Given the feasibility of this approach, a decision was made to not compress the images and keep the whole videos for the purposes of this study.

![Image of test pattern and hand](image)

**Figure 5.6: Images extracted from the BlackMagic video card**

With all the options tested, there is a barely noticeable delay in the video feed when using types of outputs other from the proprietary RGB/SDI. In conversation with an OLYMPUS technician, he warned us that the video interception to our system could experience a slight decrease in image quality or differences in shades of color depending on the adopted connection. The captured data has better, or at least the same quality as the one that is captured for reports, so we can speculate that the quality decrease from RGB should not have an influence on a possible diagnosis. However, physicians would have to test this claim by comparing and evaluating the images.

### 5.3 Gesture Control

To build a 3D gesture system you need software and hardware. Together, they can capture your movement and translate it into a command. The software does the job of recognizing gestures by picking the data gathered by camera sensors. Hardware-wise, a camera is generally part of the required parts, and normally has other elements built in to better perceive depth, such as infra red projectors or sensors. Other technologies (from Elliptic Labs\(^6\) for example) make use of ultrasound instead of cameras. The projector emits IR light in pulses, which is outside the spectrum of visible light for humans. The sensor can detect the IR light reflected off what’s in front of the projector, and then a timer will measure how long it takes for the light to leave, bounce and return. The amount of time it takes the light to travel will vary as objects move, and the system can use that data to interpret it as movements and commands. Some systems use a combination of multiple technologies in order to detect movement, not relying on a single technological approach. One of another methods is called structured light, in which the projector emits IR light in a grid pattern, and as the grid encounters physical objects and distorts, a sensor can measure these distortions and translate it into commands (Jonathan Strickland, 2012). There are quite a few options in the market regarding gesture and motion control, like GestureTek, Microsoft Kinect, Myo Armband, Nintendo WiiMote, Playstation Move, Leap Motion Controller (Figure 5.7), among others.

\(^6\)www.elipticlabs.com
5.3.1 Users

In order to perform our experiments on gesture control, eight users were recruited. All were comfortable with information technologies, having no trouble whatsoever in understanding what was asked for them to do. Two of them fall in the 20-25 age gap, three of them in the 31-35 gap, one of them in the 45-50 gap, two of them in the 60-65 gap.

5.3.2 Materials and Methods

We selected a Leap Motion Controller as our gesture recognition technology, given its easier positioning inside a gastroenterology room. Microsoft Kinect is a very interesting alternative but its accuracy is heavily dependent on room configuration and distances, which requires more extensive studies than the time available for this thesis. We started by testing the device with the bundled applications, and its precision and effectiveness were good. Infra Red (IR) waves from sunlight have a direct impact on the gesture recognition as they interfere with the IR sensors of the device, so it does work better in places where sunlight is close to a minimum. The controller comes with some predefined gestures such as circle (circular movement by a finger), swipe (straight line movement by the hands with fingers extended), screen tap (forward tapping movement by a finger), and key tap (downward tapping movement by a finger). While writing this thesis, we had the opportunity of working with Vitor Afonso, a student from the DCC-FCUP who needed a topic for a project and took interest in the Leap Motion. Following our advice, he successfully developed a simple, user friendly Graphical User Interface (GUI), with several gesture combinations which made it possible to open an image from a set, draw annotations in two different ways, save or discard these annotations, close the image, and go back and forward in the image directory. Although the commands look simple and effective at a first glance, they aren’t as swift and flawless as required by a live procedure. We didn’t have access to the application in order to run studies regarding responsiveness or muscle fatigue from gesture combination. One of the suggested features was
the possibility of mapping your own gesture to a predetermined function, but due to limited time, it was not implemented. We decided to assess the Leap Motion controller tracking capabilities using its sandbox applications, using their proposed gestures. Assuming that the proportion of success in the dichotomous experience is of 85%, and to ensure an estimate of the proportion within 10% of the real proportion, a sample of size 49 is needed (with a 95% confidence interval). We experimented with the 8 different users, performing 5 different gestures 50 times (10 times each gesture at a time, for washout purposes). This way we can try to prevent the user from sticking to an unnatural or biased way of making the gestures. Contemplated gestures were **swipe right**, **swipe left**, **zoom in**, **zoom out**, and **grab**. The first two are done by a simple waving of your hand over the controller from left to right, and from right to left. Zoom in and out is done by making a circle with your finger, clockwise being zoom in, anti-clockwise being zoom out. Grab is done by a pinching motion using your thumb, index and middle fingers. These are represented in Figure 5.8 below:

![Leap Motion Gestures](image)

Figure 5.8: Leap Motion Controller Gestures, adapted from www.assetstore.unity3d.com

### 5.3.3 Results

The results are below in Table 5.1. Insuccess includes both the gesture not being recognized or doing an unexpected command:

<table>
<thead>
<tr>
<th>User</th>
<th>Swipe Right</th>
<th>Swipe Left</th>
<th>Zoom In</th>
<th>Zoom Out</th>
<th>Grab</th>
<th>Success Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>47/50</td>
<td>42/50</td>
<td>44/50</td>
<td>42/50</td>
<td>23/50</td>
<td>79.2%</td>
</tr>
<tr>
<td>2</td>
<td>46/50</td>
<td>45/50</td>
<td>40/50</td>
<td>43/50</td>
<td>19/50</td>
<td>77.2%</td>
</tr>
<tr>
<td>3</td>
<td>46/50</td>
<td>40/50</td>
<td>45/50</td>
<td>47/50</td>
<td>28/50</td>
<td>82.4%</td>
</tr>
<tr>
<td>4</td>
<td>41/50</td>
<td>48/50</td>
<td>42/50</td>
<td>39/50</td>
<td>30/50</td>
<td>80.0%</td>
</tr>
<tr>
<td>5</td>
<td>44/50</td>
<td>41/50</td>
<td>46/50</td>
<td>41/50</td>
<td>12/50</td>
<td>71.6%</td>
</tr>
<tr>
<td>6</td>
<td>45/50</td>
<td>43/50</td>
<td>44/50</td>
<td>44/50</td>
<td>25/50</td>
<td>80.4%</td>
</tr>
<tr>
<td>7</td>
<td>45/50</td>
<td>47/50</td>
<td>45/50</td>
<td>42/50</td>
<td>22/50</td>
<td>80.4%</td>
</tr>
<tr>
<td>8</td>
<td>46/50</td>
<td>41/50</td>
<td>46/50</td>
<td>41/50</td>
<td>23/50</td>
<td>78.8%</td>
</tr>
<tr>
<td>Success Rate</td>
<td>90.0%</td>
<td>86.7%</td>
<td>88.0%</td>
<td>84.6%</td>
<td>45.5%</td>
<td></td>
</tr>
</tbody>
</table>

The success of each gesture apart from ‘Grab’, which is a bit hard to interpret by the device, was more than 80%, and the success rate of each user completing all gestures was nearing 80%. However, they
have to be done slowly and the device needs to be in an optimal position in front of the user and under the hand. Taking into account these results, and our time and conclusions assisting the procedures, we speculate that in-exam use of this technology would most certainly be difficult. This isn’t necessarily a drawback as we can surely contemplate post-exam interactions, but not in this particular study for now. Our personal opinion after this study is that touchless, gesture and motion control technologies applied to the medical field would find their purpose in areas where:

- Sterilization and inability to touch things other than medical equipment and the patient are of utmost importance, like surgery;
- May help the rehabilitation of stroke victims, or patients who need to undergo physiotherapy;
- Medical students could learn better from the implementation of virtual reality models;

5.4 Voice Control

Speech recognition is one of the most complex areas in computer science, as it involves a blend of linguistics, mathematics, and computing. There are four different ways a computer can take to turn spoken sounds into written words: pattern matching (where each verbalized word is identified in its entirety); feature analysis (where each word is separated into fragments and then recognized from its key features, such as vowels ratio); language modeling and statistical analysis (where grammar knowledge and the probability of certain words or sounds coming one after another is used to improve accuracy and speed up the process); artificial neural networks (where computer models can reliably recognize patterns after intensive training)(Chris Woodford, 2015; Ed Grabianowski, 2006). The systems start by listening to an utterance, which is a chunk of sound, through a microphone. That sound is digitized through an analog-to-digital converter. The digital data is converted into a spectrogram, using a mathematical technique called a Fast Fourier Transform (FFT), and then split into a series of chunks called acoustic frames. These are digitally processed in different ways and analyzed to find the components of speech they contain. After the separation of utterances into words and the key features of each one are identified, they are compared with a phonetic dictionary and we can determine what probably has been said. For any given segment of sound, there are many things the speaker could potentially be saying. The quality of a recognizer is determined by how good it is at refining its search, discarding poor matches and selecting the most likely. This depends in large part on the quality of its language and acoustic models as well as the effectiveness of its algorithms, both for processing and searching across the models. Several factors can reduce accuracy. Any extra noise coming from a number of sources introduced into the sound will interfere with the recognition. Apart from the environment, low-quality sound cards, often do not come with enough shielding from the electrical signals produced by other computer components, which results in the possible introduction of hum or hiss into the signal. Running the statistical models needed for speech recognition requires the computer’s processor to do a lot of heavy work, as they need to remember each stage of the word-recognition search in case the system needs to backtrack to come up with the right word. Current systems also have difficulty isolating simultaneous speech from various users, and homophones are also a problem, such as “There” and “their”, “two” and “too”, “be” and “bee”. However, extensive training of systems and statistical models that take into account word context have greatly improved their performance(Chris Woodford, 2015; Ed Grabianowski, 2006).
5.4.1 Users

For the purposes of this study, we selected 14 users which were comfortable with information technologies, having no trouble whatsoever in understanding what was asked for them to do. Two of them fall in the 20-25 age gap, three of them in the 26-30 gap, six of them in the 31-35 gap, three of them in the 60-65 gap.

5.4.2 Materials and Methods

Microsoft Windows\textsuperscript{7} is shipped with Windows Speech Recognition for quite some time now, and we decided to put it to the test. In order to assess the software reliability and robustness, we first need to define a group of adequate test words that could be used to control the system: **Hide, analyze, compare, store, discard, back, start, and end.** This set of eight words would probably encompass all the commands. Once again, assuming that the proportion of success in the dichotomous experience is of 85%, and to ensure an estimate of the proportion within 10% of the real proportion, a sample of size 49 was needed, for a 95% confidence interval. So, our process was to test each spoken word 50 times (10 times each word at a time, again for washout purposes), using 14 different users, and tracking how many times the correct word would be picked up.

5.4.3 Results

The results are below in Table 5.2. Lack of success was considered when the target word would be understood by the software as another one:

<table>
<thead>
<tr>
<th>User</th>
<th>hide</th>
<th>analyze</th>
<th>compare</th>
<th>store</th>
<th>discard</th>
<th>back</th>
<th>start</th>
<th>end</th>
<th>Success Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27/50</td>
<td>26/50</td>
<td>46/50</td>
<td>41/50</td>
<td>47/50</td>
<td>42/50</td>
<td>44/50</td>
<td>29/50</td>
<td>75.60%</td>
</tr>
<tr>
<td>2</td>
<td>28/50</td>
<td>27/50</td>
<td>43/50</td>
<td>38/50</td>
<td>46/50</td>
<td>40/50</td>
<td>41/50</td>
<td>27/50</td>
<td>72.50%</td>
</tr>
<tr>
<td>3</td>
<td>30/50</td>
<td>25/50</td>
<td>47/50</td>
<td>40/50</td>
<td>40/50</td>
<td>45/50</td>
<td>46/50</td>
<td>25/50</td>
<td>74.60%</td>
</tr>
<tr>
<td>4</td>
<td>22/50</td>
<td>19/50</td>
<td>44/50</td>
<td>42/50</td>
<td>38/50</td>
<td>37/50</td>
<td>41/50</td>
<td>21/50</td>
<td>66.00%</td>
</tr>
<tr>
<td>5</td>
<td>27/50</td>
<td>18/50</td>
<td>44/50</td>
<td>42/50</td>
<td>38/50</td>
<td>42/50</td>
<td>42/50</td>
<td>26/50</td>
<td>70.30%</td>
</tr>
<tr>
<td>6</td>
<td>32/50</td>
<td>20/50</td>
<td>46/50</td>
<td>41/50</td>
<td>44/50</td>
<td>43/50</td>
<td>46/50</td>
<td>28/50</td>
<td>75.00%</td>
</tr>
<tr>
<td>7</td>
<td>25/50</td>
<td>27/50</td>
<td>45/50</td>
<td>42/50</td>
<td>40/50</td>
<td>48/50</td>
<td>42/50</td>
<td>23/50</td>
<td>73.00%</td>
</tr>
<tr>
<td>8</td>
<td>30/50</td>
<td>23/50</td>
<td>42/50</td>
<td>40/50</td>
<td>37/50</td>
<td>35/50</td>
<td>40/50</td>
<td>29/50</td>
<td>69.00%</td>
</tr>
<tr>
<td>9</td>
<td>25/50</td>
<td>20/50</td>
<td>41/50</td>
<td>42/50</td>
<td>44/50</td>
<td>33/50</td>
<td>43/50</td>
<td>28/50</td>
<td>69.00%</td>
</tr>
<tr>
<td>10</td>
<td>23/50</td>
<td>17/50</td>
<td>40/50</td>
<td>39/50</td>
<td>45/50</td>
<td>39/50</td>
<td>45/50</td>
<td>25/50</td>
<td>68.20%</td>
</tr>
<tr>
<td>11</td>
<td>30/50</td>
<td>20/50</td>
<td>46/50</td>
<td>41/50</td>
<td>40/50</td>
<td>44/50</td>
<td>47/50</td>
<td>28/50</td>
<td>74.00%</td>
</tr>
<tr>
<td>12</td>
<td>27/50</td>
<td>22/50</td>
<td>44/50</td>
<td>44/50</td>
<td>42/50</td>
<td>43/50</td>
<td>40/50</td>
<td>26/50</td>
<td>72.00%</td>
</tr>
<tr>
<td>13</td>
<td>26/50</td>
<td>19/50</td>
<td>43/50</td>
<td>43/50</td>
<td>45/50</td>
<td>40/50</td>
<td>45/50</td>
<td>22/50</td>
<td>70.10%</td>
</tr>
<tr>
<td>14</td>
<td>28/50</td>
<td>21/50</td>
<td>45/50</td>
<td>45/50</td>
<td>43/50</td>
<td>41/50</td>
<td>44/50</td>
<td>30/50</td>
<td>74.30%</td>
</tr>
</tbody>
</table>

We used a microphone array from a laptop, which is far from being the best choice. “Hide” was frequently recognized as “High”, “Analyze” with “Penalize”, and “End” with “And”, among others. Still, the results were rather satisfactory. A possible solution could be the association of these words that sound alike to the same instruction, so that if the word isn’t perceived exactly as what the user said, the

\textsuperscript{7}www.microsoft.com/en/windows
command would still go through. For better outcomes, the solution could include changing the commands to words that have a more ‘unique’ sound, and of course, using a better microphone. We recorded some data aside, purposely with background noise such as conversations, television, and even a train passing by outside, to check if ambient noise would interfere with the results. The success rate frequency was the same. Nevertheless, the rooms are very quiet, so we could speculate that background noise levels wouldn’t be a problem. Voice recognition has a lot of potential though, and further studies in the future should be contemplated, knowing that there should be limitations when the exam features non-sedated patients.

5.5 Face recognition

Face recognition technology is considered the least intrusive and fastest to use biometric technology, as it doesn’t require people to place their hand on a reader or position their eye in front of a scanner. It analyzes the characteristics of a person’s face through a digital video camera, as each face has several distinguishable landmarks that make up facial features, with approximately 80 nodal points. Some of them are represented in Figure 5.9.

![Figure 5.9: Nodal points used in facial recognition, adapted from www.extremetech.com](image)

Some of the points measured by this technology are the distance between the eyes, nose width, eye socket depth, cheekbone shape, or jaw line length. A faceprint is the numerical code created by the measuring of these nodal points. It represents the face in the database. 3D facial recognition can capture a real-time 3D image of a person’s facial surface, using distinctive features of the face to identify the subject at different view angles with the potential to recognize up to 90 degrees. The system goes through a series of steps to verify the identity of an individual: detection, alignment, measurement, representation, matching and verification. The image may not always be verified or identified solely by facial recognition. There is another process that uses skin biometrics, called Surface Texture Analysis and it works quite the same way as facial recognition. A skinprint is taken, which is a picture of a patch of skin, and is then broken up into smaller blocks. Algorithms are used to turn the patch into a mathematical, measurable space, and the system will then distinguish lines, pores and the actual skin texture. This whole technology is not perfect though, as there are some factors that could get in the way of recognition, including glare on glasses, long hair in front of parts of the face, poor lighting and
lack of good image resolution (Kevin Bonsor, Ryan Johnson). With no access to a proper digital video camera, our option was to go with a laptop’s webcam. After testing face recognition software packages like CameraMouse\textsuperscript{8} or eViaCam\textsuperscript{9} (KeyLemon\textsuperscript{10} and FaceRig\textsuperscript{11} are also good options, but are paid), we considered that they work surprisingly well and the commands are received without any problems, even at low lighting. The reality in this case is that blinking your eyes or tilting your head would simply not be enough to cope with the task speed or the amount of commands we envision our system having. Not too many of course, but still more than 3 or 4. This ended up meeting the initial assumption we had, that these types of algorithms are very good are detecting faces and are more driven towards access control or for helping people with disabilities when no other options are available, than for quick application control.

5.6 Touch-based interaction

Touchscreens are input devices that are highly desirable user interface component and are now introduced in a vast array of electronics, allowing the user to interact directly with what is displayed without having to use intermediate devices. Three systems are used to recognize touch: resistive, capacitive (Figure 5.10) and surface acoustic wave.

- The resistive consists of a glass panel, covered with a conductive and a resistive metallic layer, being held apart by spacers. An electrical current runs through the two layers and when a user touches the screen, they make contact, and the changes in the electrical field are noted in those exact spot coordinates. Once the coordinates are known, the touch is translated into a command. Price-wise, it is the cheapest of all;

- In the capacitive, a layer that stores electrical charge is placed on the glass panel of the monitor. When someone touches the monitor with a finger, some of the charge is transferred to it, decreasing the charge on the capacitive layer. This decrease is measured by circuits located at each corner of the monitor, and from the relative differences in charge at each corner, the exact place where the touch event took place can be calculated;

- In the surface acoustic wave, a receiving transducer and a sending transducer are placed along the X and Y axes of the glass plate, along with reflectors that send an electrical signal from one transducer to another. The receiving transducer is able to tell if the wave has been disturbed by a touch event, locating it accordingly. It is the most expensive of the three;

One advantage that the capacitive system has over the resistive system (other than being stronger to being scratched by sharp objects) is that it can transmit roughly 90% of the light from the monitor, while the resistive can cope only with 75%. This advantage results in a much clearer picture. The acoustic wave setup, however, allows for 100% light throughput and perfect image clarity (Alfred Poor, 2012).

\textsuperscript{8}www.camermouse.org
\textsuperscript{9}eviacam.sourceforge.net
\textsuperscript{10}www.keylemon.com
\textsuperscript{11}www.facerig.com
Figure 5.10: Resistive and capacitive systems, adapted from www.extremetech.com

Tablets nowadays are small, light, somewhat inexpensive and easy to use. Also, quite easy to clean, which is of importance if we are to introduce them to sterile clinical environments. They are essentially bigger versions of smartphones with pretty much the same capabilities, and remote communication via bluetooth could be an elegant solution, since it brings a familiar and (almost) foolproof interaction. However, for this to work, the interface must be well thought of, and then tested and evaluated for usability.

5.7 Discussion

The system should be able to capture an endoscope’s video stream with the best possible quality, and be as easy and swift as possible to interact with. The results obtained from the video capture were not the best, as we couldn’t extract video in HDMI, but were very satisfying. A new experiment with a different video card should be in order in a near future so we could ascertain if better results can be achieved. Interaction-wise, the case studies in which we could be more detailed, presented fairly good results. However, good functionality does not always account for good usability. We will then put face recognition and gesture control in the group of options that, although viable, at a first glance, would not be adequate for the system we are currently envisioning for now, or would be more suitable for other types of environments. Voice control and touch-based interaction, on the other hand, seem more promising at the moment. Both technologies have their pros and cons against each other, but they would work essentially at the same level in terms of controlling the system. Voice control has the advantage over touch-based interaction of not having to bring a tablet to the equation (by using a lapel or shotgun microphone, for example), which is good for when both hands are busy, but brings the disadvantage that commands may not be recognized at times, which could be bad when time to spare is not much and the attention span to other things other than the patient is scarce.

In the next section we will cover system and interface design, as well as its evaluation. As we believe that touch-based interaction could be the most viable option, we favored this technology to be studied a bit more in depth, so we will be directing the design towards its use.
6. A Touch-Based Interactive System for Gastroenterology
6. A Touch-Based Interactive System for Gastroenterology

6.1 Methodology

The methodology of building a system like this is based on the Software Development Life Cycle (SDLC) used in software engineering, which depicts stages to develop and test high quality software that meets or exceeds customer expectations, while reaching cost estimates and deliverables within time. It contains several paradigms, models and methods, such as the waterfall, spiral development, prototyping, various types of agile methodologies, Iterative and Incremental Development (IID), Rapid Application Development (RAD), among others. There isn’t really a consensus with the stages’ characterization and naming, so we will be using 5 different phases to illustrate the process (Figure 6.1): Analysis/Requirements, Design, Coding, Deployment, and Support (Centers for Medicare and Medicaid Services, 2008).

![Figure 6.1: A Software Development Life Cycle](image)

After the identification of an opportunity and a concept proposal is created, the Analysis/Requirements stage is where the user, functional and technical requirements should be understood. Identification and capture of the stakeholder requirements is normally done by interviews and surveys. A gap analysis, in where actual performance is compared with potential or desired performance, should be done to evaluate if the user and its routines are operating as they should or could. A feasibility analysis will uncover the
strengths and weaknesses of what currently exists and ultimately evaluate if the project implementation can be successful. This stage may also contemplate a cost-benefit analysis and a risk management plan. Development of User Stories and system Use Cases should also be done when gathering requirements. The Design stage transforms the detailed requirements into something more tangible by focusing on delivering the required functionalities. This approach has to provide a detailed system specification and clearly define all the architectural modules (illustrating a specific set of trade-offs inherent in the structure and design) of the product, along with its communication (user interface and interaction) and data/work flow representation (external/internal procedures) with the external and third party modules, if any. If the next step (coding) is expected, the database model should also be thought of. Sometimes, this stage may be the proof of concept’s location. Coding usually includes development and integration. It’s the stage where a team of developers works on the software, and where the design is converted to a complete information system. This includes acquiring and installing system environments, creating and testing databases, prepare testing procedures, compiling, refining and validating. The integration should be made with the existing IT structure, while beta testing various iterations of the system and running efficiency tests. Deployment deals with client-side installation and training of the user. The software may be first tested in a limited segment of the market (User Acceptance Testing). After all is set up, Support handles optimization and usefulness tests, as well as the analysis of working activity and the monitoring of load and stress. This will be useful for a continuous improvement of the system.

Since software coding and what goes beyond it is not included in this thesis, we will only contemplate the first two stages and not worry with choosing an appropriate SDLC framework by recognizing its key points. The analysis stage is composed by most of the described work until now (chapters 4 and 5), apart from the construction of a user story and a use case diagram, which will be done in the next section. We will then proceed to the design stage, creating a flow diagram and developing the system image. We need to conceptualize a system that is able to respond to what is proposed: an interactive environment which can provide the gastroenterologist with information that can help with his diagnosis within a gastroenterology room, while not decreasing current patient care efficiency. There is, however, an alternative way to develop these systems, mostly centered in usability, which is depicted below in Figure 6.2.

6.1.1 Analysis/Requirements stage

After all the observations which led to the gathering of opportunities and requirements, we need to put our findings into perspective by defining possible and wanted tasks and functionalities and the actions required to perform them. A user story is a tool that captures the description of features from an end-user viewpoint. The user story describes the type of user, what they want and why, often similar to a set of use cases, but written in plain text. It helps to create a simplified description of a requirement and it is meant to be relatively short, usually in the language of the customer so that its desires are clear to the development team.
Figure 6.2: User centered Design, adapted from Kellogg S. Booth, Introduction to HCI Methods, University of British Columbia, Canada

The development team’s job is satisfying the requirements of the user story while writing the code. Sometimes, collaboration between users and developers is needed to clarify the details as the project progresses. At this time, the technology we will use regarding system control is also defined. As stated in the previous section, we think that a tablet should provide the best hassle-free and most effective interaction with the system we want to implement in the room. The user story follows below, as we divide the actions into 1 ‘false’ state, and 5 ‘real’ states in which the user could perform actions:

• **NO STATE**

Every day at a scheduled time, the system automatically requests a synchronization with the hospital’s Electronic Health Record (EHR)/Picture Archiving and Communication System (PACS), in order to obtain a list of all the scheduled procedures for the day. The request also includes the exam code and patient demographics, which will be used to create a patient record inside our system, along with the snapshots taken and other data. The daily procedures’ list is sent from the system to the tablet.

• **LOGIN STATE**

The user is presented with a login screen. Failure to login will prevent the user from accessing the system.

• **HOME STATE**

The user can access to the daily procedures list. The user can access the recorded patient database of our system. The user has the possibility of forcing the synchronization of the procedures, in case the automatic sync does not work for some reason.

• **SELECTION STATE**

The user can select an exam from the list of daily procedures. The system sends a message to the tablet “initiating exam”, burn on the screen with the same message. Tablet and system enter in recording state.
• RECORDING STATE

If the doctor freezes an image for $\geq 3$ seconds (the system has no response for less than that) the system
snapshots a frame and keeps it, with a burn on the screen "snapshot taken". System sends the snapshot
to the tablet with a selection of three options (option state): hide, analyze, compare.

• OPTION STATE

Option "hide" leads to a choice between store (option for images that don’t need analysis or comparison,
but would still be good to keep) or discard (option for accidental long freezes). Store keeps the image in
the patient record, burn on the screen "image stored". Discard keeps the image in the system, but not
on the patient record (although associated by a number), burn on the screen "image discarded". Tablet
and system go back to recording state.

Option "analyze" assumes a computer vision analysis is required. Tablet request an analysis to the
system, burn on the screen "analyzing". System analyses the image, burns the result on the screen, and
also sends it to the tablet. Image is kept in the patient record along with the associated computer vision
score.

Option "compare" assumes the user wants to compare the snapshot. Tablet request a comparison to
the system, burn on the screen "comparing". System analyses/compares the image, burns on the screen
the comparative histological statistics of other snapshots already classified in the system. System sends
to the tablet the same statistics.

Option "finalize exam" assumes the user wants to end data collection, burn on the screen "exam
finalized". Tablet and system leave recording state. Tablet goes back to selection state.

We speculate that this could be an acceptable action sequence, based on the routine course of the
exam, convenience, and usefulness inside a gastroenterology room. All pondered options are covered, and
process feasibility is also accounted for.

A use case, on the other hand, is a user-centric methodology used in system analysis to identify, clarify,
and organize system requirements, while depicting how users will perform tasks and the interactions
among the elements of a system, by a list of actions of event steps. Diagrams are employed in Unified
Modeling Language (UML), a notation which provides a standard way to visualize the design of a system.
They usually contain boundaries, actors, preconditions and relationships. Use cases help explain how the
system should behave, providing a list of goals that can be used to establish its cost and complexity. We
now have two different actors, the doctor and the nurse, performing actions inside our system. At this time
we are already defining the main user of our system. As we’ve chosen a tablet for interaction purposes,
the nurse would be responsible for the control, as doctors spend practically the entire procedure with
both hands occupied. If in the future a hands free interactive option is considered (like voice recognition),
the main actor could easily be changed to the doctor if necessary. The use case follows below in Figure
6.3:
6.1.2 Design stage

With all of this in place, we can begin to prototype the interface that will be subjected to evaluation. In this case we need to conceptualize not only the system itself, but also where will it be implemented. Based on our previous image depicting the structure of the environment of the exam room, we can add the existing and considered hardware connections, as a system model (Figure 6.4). We included the portable CV-190 with his own monitor, although the CV-180 could be updated to a 190 in the near future and there would be no need for a portable solution to be there. Nevertheless, this scenario will contemplate the use of different endoscopes that can be brought to the exam room, while also considering the users’ positions in space. Existing connections from the endoscope to the room monitor and desktop PC are in black, considered connections are in red, while adding ourselves an extra tablet, monitor, and workstation for system implementation. This solution allows for side-by-side implementation while not disturbing the current patient care and procedure *modus operandi*. 
A flow diagram, also called a process flowchart, will usually be a representation of how an object will flow through various rules and system states. They are often used for documentation purposes because people with different backgrounds may use that documentation and it’s easier to follow the steps, as opposed to pseudo-code, for example. The flow shows how the system behaves internally and how the external environment is configured. It stands apart from the use case diagrams because it describes how tasks can be accomplished rather than showing what we want the system to do. We produced five different diagrams for better viewing, which explain how processes would flow. They are composed by five different actors, where OLYMPUS is the endoscopy system, GEMINI is the workstation where our system is, SCREEN is the extra monitor we will add, TABLET is the tablet which will be used for control, and EHR/PACS regarding to the integration between our system and the Electronic Health Record system/Picture Archiving and Communication System. As to the description of the diagrams:
- **Login** gives access to the application, provided a valid username and password are inserted (Figure 6.5);

![Login Flow Diagram](image)

**Figure 6.5: Login Flow**

- **Sync** depicts the automatic synchronization between our system and the hospital’s system when asking for the daily procedures (Figure 6.6);

![Sync Flow Diagram](image)

**Figure 6.6: Sync Flow**
- **ForceSync** refers to a synchronization like the above, but forced by the user (Figure 6.7);

![Figure 6.7: Force Sync Flow](image)

- **SelectCase** represents the choosing of a procedure by the user, with the option of starting and finishing it (Figure 6.8);

![Figure 6.8: Select Case Flow](image)
- **CaptureProcess** details the process of the user freezing an image on the endoscope, and the results of the actions that can be performed on that snapshot (Figure 6.9);
To close the conceptual design chapter, the system image (or interface) can be designed when the previous steps are done. It should be done by trying to coincide our conceptual model with the user’s mental model. GUIs should be clear, coherent, provide enough information and make all possible commands visible. This way, a user could discover by itself the available tasks, and quickly learn how to perform tasks. Mockups, also called low-fidelity prototypes or paper prototypes, give us the opportunity to acquire feedback from users about design ideas, as they can function as a discussion medium with them. This way we can exchange ideas, analyze the, redo what may not be intuitive, in an effective and inexpensive way. Just by looking at a mockup, users should get a good notion of how the system would feel as a whole, even if the functions aren’t yet working. This is called horizontal prototyping. A horizontal prototype is a simulation of the interface where users are able to execute all navigation and function commands, without any real tasks being actually performed. They are fast and easy to implement, as opposed to a fully functional vertical prototype, which would require coding and a greater amount of time to produce. Balsamiq\(^1\) provides a very good tool in which we can produce mockups. The system image figures we built are in the Appendix, as they were too many to insert here.

This system image was created having a few things into account: all pre-exam commands could be a little more complex, but yet intuitive to the user. In-exam commands must be as visible and seamless as possible. Hiding and Analyzing only require two touches from the user to be executed and closed. Comparing requires the same two, adding more depending on how many comparisons we would like to check. We believe that the user’s mental model will be close to our idealized conceptual model, but the evaluation below will give us more information on that matter.

6.2 Evaluation

The development of an interactive system is iterative, and evaluation is a key factor of this process. There is an emphasis of the importance of bringing the users into the interface design process, but you’ll also notice that the time they can devote to your project is almost never a free or unlimited resource (especially in the medical field). You should catch all trivial bugs early so when they do take the time to look at your design, it should be as error-free as possible. Usability evaluation should start early in the design process, better yet in the stages of low-fidelity prototyping, for a better chance of detection and correction of critical flaws, as well as problems with system interaction. According to Nielsen and Mack, "Usability is a fairly broad concept that basically refers to how easy it is for users to learn a system, how efficiently they can use it once they have learned it, and how pleasant it is to use" (Nielsen, 1994b).

In this chapter we use two different approaches for interface evaluation that focus on the basic principles of usability:

- The cognitive walkthrough, which is a task-oriented method that also evaluates the mental model of the user, helping to understand the system’s learnability;

- The heuristic evaluation, which makes use of a list of predetermined heuristic rules (check-list approach) that catches a wide variety of problems regarding the system image;

\(^1\)https://balsamiq.com
There are other types of approaches, such as action analysis, which measures the time that an expert user would need to perform a task, and the heuristic walkthrough, which combines aspects of the cognitive walkthrough and heuristic evaluation, among others.

### 6.2.1 Cognitive Walkthrough

The cognitive walkthrough is a usability inspection method in which one or more evaluators use the interface to work through a series of tasks, and ask a set of questions that a typical user will need to accomplish. Its main focus is to understand the system’s learnability for new or infrequent users, hence it was originally designed as a tool to evaluate walk-up-and-use systems like ATMs, ticket machines or public information systems, where the design needs to be so self-explanatory, that users with little to no training are able to use it effectively. Being quick and inexpensive to apply, as well as providing suggestions on how to improve learnability and taking explicit account of the user’s tasks, are among its common advantages. However, it doesn’t estimate a frequency or severity of an identified problem and it can become labor intensive. Also, the analysis could be a bit superficial, as the user may focus too much in superficial aspects of design rather than deep aspects such as task appropriateness or error recovery.

When designing small blocks of the interface on your own, informal walkthroughs can be done on your own, to monitor the design as you work. As other blocks of the interface begin to fuse, a group of people (that might be composed by other designers and users) should be gathered to do a walkthrough for a complete task. A thing to retain is that this is a tool for interface development, not for validation. As so, keep in mind that you should expect to find things that can be improved.

Before proceeding to a walkthrough, you need different information:

- A description, representation, or prototype of the user interface. It’s not mandatory that it is complete, but the level of detail should be enough for proper testing;

- An idea of the user profile (or persona). Who will they be and what experience do they have. The creation of specific personas is not mandatory, but at least the type of target user should be clear, after the previous user analysis;

- A task or a piece of task, as well as a complete written list of actions (action sequence) needed to complete the task within the interface, from beginning to end;

After defining the users and conducting a context of use analysis, we should determine realistic, simple tasks (and task variants), moving to more complex ones later on. With the action sequence and the interface also defined, it’s time to conduct the actual walkthrough by providing a representation of the interface to the evaluators. There are a few key points one should be concerned about in the process:

- Users generally aren’t thinking the way designers expect them to think. Will the users be trying to perform actions that may go beyond their mental model?

- Users should be able to locate interface controls. Not necessarily identify it as the right control at a first instance, but actually just notice it exists. Once they find it and want to perform the action, an will they be able to recognize that is produces the effect that was asked?
• Users need feedback by performing even the simplest action, because they really need evidence that whatever they were trying to accomplish has been done, or got closer to being done. Will they understand that feedback and be motivated to proceed to the next task?

While performing the walkthrough, success and failure rates should be noted, as well as design suggestions, probable errors, areas of confusion, comments, and whichever things that may prove useful in a re-design. Some fixes will be easy, like making the controls more obvious/accessible, or providing better feedback. Others, not so much. An understanding of strengths and weaknesses should be achieved, and a brainstorm on potential solutions for identified problems with other designers is also advised (Polson et al., 1992; Lewis and Rieman, 1993).

Methods and Results
This type of evaluation allows to quantify if a new user is able to work with a new system, through its mental processes, with special attention to how well the interface supports exploratory learning. Using a Norman’s Diagram (Figure 6.10), we established the fulfillment of the following tasks:

![Norman’s Diagram](image)

**Figure 6.10**: Norman’s Diagram, adapted from Kellogg S. Booth, Introduction to HCI Methods, University of British Columbia, Canada

1. Start and finish a procedure;
2. Check a patient record;
3. Perform a compare action on the image;
4. Log off the application;

We then invited five different users to perform them. Two of them had a lot of experience with IT (25-30 age gap) and used them every day, three of them were fairly comfortable with IT and used them every so often (one in 45-50 age gap, two in 60-65). All five users knew what a tablet was and had used one at least one time. To evaluate task completion, we used three different levels. The user completed its task (green); the user had difficulties with task completion (yellow); the user didn’t complete the task (red). We also went ahead and recorded the times of completion of each task for each user (in seconds).
Difficulty in completion would result in more time spent on the task. The results follow below in Table 6.1

Table 6.1: Results from the Cognitive Walkthrough

<table>
<thead>
<tr>
<th>Task</th>
<th>User 1</th>
<th>User 2</th>
<th>User 3</th>
<th>User 4</th>
<th>User 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>6.22</td>
<td>9.32</td>
<td>1.31</td>
<td>13.43</td>
<td>15.17</td>
</tr>
<tr>
<td>Task 2</td>
<td>5.45</td>
<td>5.17</td>
<td>5.27</td>
<td>10.09</td>
<td>12.23</td>
</tr>
<tr>
<td>Task 3</td>
<td>8.07</td>
<td>8.33</td>
<td>10.33</td>
<td>11.38</td>
<td>22.56</td>
</tr>
<tr>
<td>Task 4</td>
<td>3.14</td>
<td>3.05</td>
<td>4.45</td>
<td>6.32</td>
<td>7.29</td>
</tr>
</tbody>
</table>

Every user was able to complete the tasks, although users 4 and 5 had more trouble getting there. This was mostly due to the interface being in a different language from the user’s native one. If this would become a problem, translating the interface to another language could also be done rather effortlessly. Regarding the results, we weren’t expecting less. The interface is very simple and direct to use, as we knew it had to be when we were gathering the requirements. Moreover, we think the recorded times for a first use are also adequate for the limited opportunities to perform commands during the procedure. These should be shorter for routine use by users, but we would need higher-fidelity prototypes in order to properly quantify.

6.2.2 Heuristic Evaluation

An heuristic evaluation is a method in which one or more expert reviewers analyze an interface through a list of design principles (heuristics) and identify where the product fails to follow them. Heuristic evaluation can be used throughout the design life cycle at any point, but the earlier, the better. Usability issues should be dealt with beforehand, and this method can then be used to evaluate versions of the interface as the design evolves. They are called heuristics because they are general principles, or broad rules of thumb, as opposed to specific usability guidelines. It’s inexpensive and intuitive to implement, and can be used early in the development process, but they usually require evaluators with experience in usability, as they can find more issues than non-experts. It also may not scale well for complex interfaces and the results may be biased by the preconceptions of the evaluators. Furthermore, they may report problems at different levels of granularity. Before proceeding to the evaluation, we will need:

- The list of heuristics accompanied by a small description;
- A list of task examples regarding the aspects of the interface you want evaluated. Some types of coverage will add more value than others;
- Access to the interface, prototype, screenshots, or even the working product;
- The form for registering heuristics violations;

Nielsen and Molich described the heuristic evaluation as “an informal method of usability analysis where a number of evaluators are presented with an interface design and asked to comment on it”. They initially presented 9 heuristics which were refined and reformulated to 10, a few years later (Nielsen and Molich, 1990; Nielsen, 1994a).

1. Visibility of system status – Appropriate feedback should always be presented to the user, to keep it informed at all times.
2. Match between system and the real world – The information should appear in a natural and logical order, providing familiar words and concepts to the user. System-specific engineering terms should be avoided.

3. User control and freedom – Undo and redo should be supported, as users will make mistakes and will need a quick exit from an unwanted state, without damaging anything.

4. Consistency and standards – Platform conventions should be followed, so users can learn that action sequences used in certain places can usually be used in other places with similar results.

5. Error prevention – When writing an error message, you should ask yourself if an how that error can be prevented. A well-thought design which can prevent problems from occurring is even better than a good error message. Users should be presented with confirmation options before selecting actions.

6. Recognition rather than recall – Objects, actions and options should be visible. Try to minimize the user’s memory load by keeping the information on the screen until it’s no longer needed.

7. Flexibility and efficiency of use – Experienced users should have an option (accelerators/shortcuts) to skip lengthy dialog and help they don’t need. They should be allowed to tailor frequent actions.

8. Aesthetic and minimalist design – Think of designs to be simple and avoid information which is irrelevant or rarely needed, as that may divert the focus of the user to other unwanted places.

9. Help users recognize, diagnose, and recover from errors – Error messages should be simple, precise, and helpful.

10. Help and documentation – A perfect (or extremely simple) system may be used without documentation, but, when there is the need to provide help and documentation, it should be easily searchable, focused, and not too large

After deciding which heuristics will be used, a team of 3 to 5 evaluators should be enough (depending on how knowledgeable they are) to provide satisfactory results. The premise is that every problem with an interface won’t be found by a single evaluator, and different evaluators will often find different problems. Each one will go through the interface alone and detect where the product violates the heuristics, in two passes. The first to understand the flow and scope of the system, the second to focus in specific elements. They will also be asked to rate how severe the violations would be (in a 0 to 4 scale) from the users’ perspective, and to provide a possible solution for the violations. The problems identified are then combined into a single list. This results is an increased assurance that if a problem can be identified with the heuristics, it can be solved. However, there might be problems that the heuristics themselves miss. Those problems might show up with some other evaluation method, such as user testing or a more task-oriented analysis (Nielsen, 1994b; Lewis and Rieman, 1993; Nielsen and Molich, 1990).

Methods and Results

The heuristic evaluation was carried out after the cognitive walkthrough, aiming to identify existing problems with the system image, thus contributing for better usability in the future. The methodology is based on the list of heuristics presented above, and we decided to carry out the assessment using 3 evaluators (in order to have a majority opinion, if needed). All of them were very at ease with IT and
were located in the 31-35 age gap. Each evaluator will locate the errors and classify them according to a severity scale from 0 to 4, being:

0. Not a usability problem;
1. Aesthetic problem only: needs no fixing unless extra time is available;
2. Minor usability problem: to fix, but with low priority;
3. Major usability problem: to fix, but with high priority;
4. Usability catastrophe: imperative to fix before product release;

A session between the evaluators and the designers was in order to discuss each heuristic and severity rate of the situations encountered on the individual evaluations. This will result in a Table 6.2 depicting the violated heuristics, the problem in each violation and the severity rate, along with a suggestion towards solving the problem.

<table>
<thead>
<tr>
<th>Violated Heuristic</th>
<th>Problem</th>
<th>Slide</th>
<th>Severity</th>
<th>Potential Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Technical language</td>
<td>16</td>
<td>3</td>
<td>Translate it to a more user friendly term</td>
</tr>
<tr>
<td>8</td>
<td>Small icon</td>
<td>4,10</td>
<td>2</td>
<td>Bigger icon</td>
</tr>
<tr>
<td>10</td>
<td>No help item</td>
<td>All</td>
<td>1</td>
<td>Add help icons</td>
</tr>
<tr>
<td>4</td>
<td>Hide function not clear</td>
<td>7</td>
<td>2</td>
<td>Promote sub-functions to main, and forget hide</td>
</tr>
<tr>
<td>3</td>
<td>Mandatory to choose one</td>
<td>8</td>
<td>1</td>
<td>Add an exit icon</td>
</tr>
<tr>
<td>1</td>
<td>No feedback</td>
<td>17</td>
<td>3</td>
<td>Rework the design in order to be more intuitive</td>
</tr>
<tr>
<td>2</td>
<td>DOB, DOP, TOP are acronyms that might evoke confusion</td>
<td>4,10</td>
<td>1</td>
<td>Exchange acronyms for the extended term</td>
</tr>
</tbody>
</table>

Seven violations of six heuristics were detected, and actions were identified to address them. Bear in mind that these results were expected for a paper prototype, as it is quite limited in terms of usability. A vertical prototype would probably lead to a more extended evaluation and analysis, of functions that simple cannot be represented in a mockup like this.

### 6.3 Discussion

Our analysis and gathering of requirements led us to believe that a tablet was the most promising solution for system interaction. Therefore, the subsequent processes of thinking and designing the system were done keeping this in mind. Use cases and process flowcharts could actually be re-used if we eventually wanted to change the interaction paradigm to, for example, voice recognition, with just a few alterations. The user interface mockup was kept simple and direct, with three main functions that could be implemented in in-exam use. Evaluating it accordingly, we learned that the interface is sufficiently intuitive, clean, and able to keep all functionalities at hand. A few heuristics violations were detected by the evaluators and taken into consideration, as it was nothing that couldn’t be handled in a quick rework of some sections of the interface. The correction would be the first step into validating this design, towards more refined models.
7. Conclusions and Future work
7. Conclusions and Future work

We were presented with the challenge of identifying interaction opportunities for computer assisted decision systems within a gastroenterology exam room. We were able to model a typical gastroenterology environment as well as its users and routine procedures, which led us to believe that not only the health professionals were receptive to having some kind of system that could help the process of diagnosis, but that there were also several approaches in how could it be implemented. From these contextual studies, two things became essential: we could only influence to a minimum the system/routine that already existed, and anything we wanted to implement could not decrease the current efficiency of patient care. Our conclusions first motivated us to case study the available interactive paradigms we initially contemplated in order to assess them in this context. Voice recognition and touch based interaction were the most promising for implementation, considering the rate of success of functionality allied to the reality of a gastroenterology procedure scenario. Face recognition and gesture control were considered impractical or more suited towards other types of scenarios. Keeping this in mind, the interface was thought of and designed as straightforward as possible, taking into account the use of a tablet for system interaction. The design was deemed functional and intuitive, as its evaluation presented good results. The annotations that were made were insightful and with low severity ratings, and reworking the design, minding the evaluation, would not be too complicated, making way to better and more refined prototypes. Hardware-wise, the tests we performed with our equipment combined with the department’s endoscopes were also solid, as we were able to extract video from the processor at a suitable quality for storage and later processing. High definition capture was not attainable in these attempts, but further testing with another video card with support for other types of outputs could provide even better results. An extra monitor and a workstation could be connected and integrated in the room without disrupting any routines. All the field work and early design is done. From here on out, there are several possibilities for future work. Gathering images from the procedures could drive the creation of a database of pathological clinical cases for research purposes, or even an e-learning platform. This could result in a more streamlined approach to teaching endoscopy to young gastroenterology trainees. Based on the previous work, coding a vertical prototype would allow us to assess a more accurate response from the end users. This way we could determine if the solution meets their expectations and if they see any value in using it. Regarding the CAD system component that is concerned to the automated decision that has to be performed on lesion detection, low-computational cost computer vision algorithms which can provide reliable answers in a few seconds should be developed and integrated to fully empower the system. A vertical prototype could be deployed at IPO-Porto’s gastroenterology department, to be used in live procedures with real patients, and be tested once more for functionality, usability and robustness. All of this could contribute into the making of a complete interactive system that could be adequate for professional use in gastroenterology.
8. References
8. References


9. Appendix
9. Appendix

Questionnaire

1. Age?
2. Role?
3. Years of service?
4. Have you ever had IT related training?
5. Do you have a computer or laptop at home? If so, is it a Mac?
6. How often do you use the internet out of work?
7. How do you deal with information at your job? Electronically or Paper-based?
8. Which software packages do you often use at work?
9. Do you have a smartphone? If so, is it an iPhone?
10. Do you have a tablet? If so, is it an iPad?

Structured Interview (Doctor)

1. Name, Age, Years of service?
2. Which patient data do you need to look at, or have, before starting an exam?
3. Would it be good or necessary to access that data during the exam?
4. How many endoscopies do you do in a week?
5. Do pedals get in the way or are they necessary?
6. Do you have a ‘unique’ way of doing the procedure or does everyone do it roughly the same way?
7. Is the image visualization enough to formulate a diagnosis or do you often need to go look at the video?
8. Do you think nurses could have more power doing the procedure? This in terms of task attribution of handling a hypothetical new piece of equipment.
9. Do you exchange impressions with the nurse after the exam is finished?

10. Do doctors often discuss the exams with other doctors for second opinions?

11. What is the biggest problem with the existing system, and what would you like to see implemented?

**Structured Interview (Nurse)**

1. Name, Age, Years of service?

2. How many endoscopies do you assist in a week?

3. What are the main cares and concerns you need to have with a patient?

4. Do you think you could have more power during an endoscopy?

5. Would you like the doctor to exchange impressions with you after the exam is finished?

6. What could go wrong during a procedure?

7. Which procedure do you think it could be more efficient, during an endoscopy?

8. What is the biggest challenge you feel during an endoscopy, being a nurse?

**System Image**

![Image of system interface]

Figure 9.1: Slide 1
Figure 9.2: Slide 2

Figure 9.3: Slide 4

Figure 9.4: Slide 6
Figure 9.8: Slide 16

Figure 9.9: Slide 17

Figure 9.10: Slide 19