Trinity College Dublin, The University of Dublin

School of Computer Science and Statistics

Presented to the Trinity College Dublin, the University of Dublin in fulfillment of the requirements for the Degree of Doctor of Philosophy

Seán Cronin

Designing touchless interaction interfaces for medical image viewers

Supervisor: Dr. Gavin Doherty, D.Phil.

Dublin 2022

To Rhiannon

Acknowledgements

I want to thank my supervisor, Gavin Doherty, for his advice and guidance throughout this work. I would like to thank my friends, colleagues, and family, especially my parents, Marguerite and John, for their support, assistance, and motivation during my studies. I would like to give a special thanks to Dr. Bridget Kane (Karlstad University) and Dr. Euan Freeman (University of Glasgow), for their invaluable insights and collaborative efforts throughout this research.

SEÁN CRONIN

Trinity College Dublin, the University of Dublin
July 27, 2022

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Sean Cronin

July 27, 2022

> 'The only way of discovering the limits of the possible is to venture a little way past them into the impossible.'

–Arthur C. Clarke

Abstract

Clinical work in modern hospitals involves extensive use of digital medical imaging, with many specialities becoming reliant on real-time access to detailed imaging. The widespread use of technology such as digital imaging in hospitals, and the difficulty of sterilising computer controls and peripherals has increased opportunities for the spread of pathogens. Increased contact with shared surfaces, such as keyboards, can potentially increase bio-contamination between users. In turn, this can result in increased rates of healthcare-acquired infections (HAI), leading to poorer patient outcomes.

Currently, while imaging provides healthcare professionals (HP) with essential insights into a patient's condition, the requirement for sterility in some situations means that users face issues such as being unable to directly manipulate key imaging in aseptic environments. Touchless interaction provides an attractive potential means for providing aseptic access to medical imaging, by reducing contact with shared surfaces without reducing HP ability to interact with imaging systems. As touchless technologies reach maturity, it is an opportune time to investigate their application to this problem, and in this thesis, I investigate the design of touchless systems for accessing medical imaging.

The thesis investigates the needs and requirements for these systems through a systematic search of the literature and through a qualitative interview study with a broad range of clinicians. Starting with the motivations for touchless control, asepsis provides a theme and motivation for the field as a whole, but other advantages, such as solving the hands-busy problem, are also proposed. Existing practices aim to maintain sterility where possible through covering input devices, for example, but often at the cost of functionality. For example, it is common for the surgeon to instruct a clinical assistant to interact with imaging systems on their behalf in the operating room (OR). This interchange can become complex, resulting in the surgeon having to talk the assistant through what is required, increasing cognitive workload for the surgeon, as well as causing delays to procedures. The interview study shed further light on clinician perspectives and attitudes towards touchless control, as well as practices and problems surrounding the existing use of PACS.

In order to better understand the process of designing and developing touchless image viewing systems, two distinct prototype systems were developed. The process of developing an initial touchless prototype (using the Kinect V2) provided insight into some of the challenges and decisions faced when designing a touchless interaction system. The insights from HP interviews were combined with the learnings from the initial prototype to develop a second prototype (using the Azure Kinect DK). HP experiences with this second prototype were explored. As unintended input is a particular challenge for touchless interaction in medical contexts, the performance of different interface latching (clutching) techniques is investigated. Analysis of semi-structured interviews with HPs following the experiment revealed an appetite for reliable touchless interfaces, a strong desire to reduce shared surface contact, and proposed potential improvements such as combined authentication and touchless control.

Given the wide variety of issues arising from the literature review, the interview study, the construction and empirical exploration of the prototypes, I further provide an framework the development process for touchless medical imaging systems, drawing also on existing research into the design of touchless user interfaces. This provides an overview of the design space, and provides a structured means for developing the design rationale for individual touchless interfaces for medical imaging. Overall, the findings presented in this dissertation can inform the development of novel touchless medical systems and help identify opportunities for future research.

Publications

Cronin, S. and Doherty, G. (2019). Touchless computer interfaces in hospitals: A review. Health Informatics Journal, 25. doi: 10.1177/1460458217748342

Cronin, S., Kane, B., and Doherty, G. (2021). A qualitative analysis of the needs and experiences of hospital-based clinicians when accessing medical imaging. Journal of Digital Imaging. doi: 10.1007/s10278-021-00446-1

Lacey, G., Gozdzielewska, L., McAloney-Kocaman, K., Ruttle, J., Cronin, S., and Price, L. (2021). Psychomotor learning theory informing the design and evaluation of an interactive augmented reality hand hygiene training app for healthcare workers. Educ Inf Technol. doi: 10.1007/s10639-021-10752-4

Cronin, S., Freeman, E., and Doherty, G. (2022). Investigating clutching interactions for touchless medical imaging systems. CHI '22. doi: 10.1145/3491102.3517512

Contents

List of Figures

List of Tables

Glossary

AR Augmented Reality. An interactive experience that enhances the real-world with computer generated information.. 36, 38

CAQDAS Computer Assisted Qualitative Data Analysis Software.. 52, 53

COVID-19 A contagious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). First identified in December 2019, this disease developed into a worldwide pandemic. 3–5, 8, 39, 52, 53, 80, 81, 91, 115, 117, 127, 136, 138, 141, 143, 144

DICOM Digital Imaging and Communications in Medicine. This data format is the standard for the communication and management of medical imaging information and related data. Used for storing and transmitting medical images, DICOM is used in the integration of many different devices and services, including PACS. 17, 71, 92–94, 98, 104, 105

EGT Eye gaze technology. Technology that measures eye movement and position. This information can then used to control a digital interface. 21, 29, 38

EHR Electronic Health Record. A systematized collection of patient health information stored electronically in a digital format. 19, 27, 46, 67, 71

GDPR General Data Protection Regulation. A regulation in EU law on data protection and privacy in the European Union and the European Economic Area. 64, 157

GUI Graphical User Interface. A user interface for interacting with electronic devices that uses graphical icons and audio indicators rather than text-based interfaces. xv, 103, 104, 179

HAI Healthcare Acquired Infection. Any infection acquired while receiving treatment in a healthcare facility, such as a hospital, or from a healthcare professional, such as a doctor or a nurse. Examples include infections such as Norovirus, MRSA, and C. diff. v, 1, 3, 4, 117

HCI Human-Computer Interaction. A multidisciplinary field of study which focuses on the design of computer technology and the interaction between user and computer. 15, 77, 85

HIPAA Health Insurance Portability and Accountability Act. A United States federal law that required the creation of national standards to protect sensitive patient health information from being disclosed without the patient's consent or knowledge. 157

HMI Human-Machine Interface. A user interface that connects a person to a machine, system, or device. 58

HP Healthcare Professional. Any person who works in a healthcare environment, or delivers healthcare services, such as a doctor, nurse, radiographer, etc. v, vi, 1–7, 9, 34, 35, 38, 43, 45–47, 54–75, 85, 88, 103–108, 114–116, 120, 136, 143, 144, 148, 150, 151, 153, 169, 170, 180, 189

ICU Intensive Care Unit. A facility within a hospital where intensive care medicine is provided. 156, 163

IR Infrared. Electromagnetic radiation with wavelengths longer than those of visible light that is invisible to the human eye. 17, 36, 83, 88, 144

LUIS Language Understanding. A machine learning-based service to build natural language into apps, bots, and IoT devices. xvi, 109–111, 114, 168

MRI Magnetic resonance imaging. A type of diagnostic imaging that uses a magnetic field and radio waves to create detailed images of the organs and tissues within your body. 2, 17, 90, 95, 119

NFC Near-field Communication. A set of communication protocols that allow for communication between two electronic devices up to four centimeters apart. 86

NIMIS National Integrated Medical Imaging System. A computer-based system for storing and examining X-rays and scans that has been introduced to a large number of public hospitals in Ireland. 46, 67, 147

NUI Natural User Interface. A user interface that is effectively invisible. Includes vision techniques, natural language interfaces, pen-based input, and multi-touch gestural input. xv, 36, 77, 88, 89, 150

OR Operating Room. A facility within a hospital where surgical procedures are performed in an aseptic environment. v, xvi, xix, 4, 12, 15–17, 19, 20, 26, 27, 34–37, 40, 42, 62, 79, 80, 85, 88, 113, 144, 153–156, 159, 164, 173, 175, 179, 182, 187, 188, 190

PACS Picture Archiving and Communication System. A medical imaging technology that allows the efficient storage and convenient access to images from multiple modalities. In this research, PACS can also be taken to refer to equivalent systems such as NIMIS in Ireland. xix, 1, 4–7, 9, 17, 27, 43, 45–52, 54–74, 89, 92, 94, 95, 98, 103–106, 116, 117, 120–127, 133–148, 150, 153, 157, 170, 188–190

PPE Personal Protective Equipment. Any worn or held device or appliance designed to protect an individual against one or more health and safety hazards. 127, 133, 134, 139, 143, 146, 147

RFID Radio Frequency Identification. Consisting of a receiver and a transmitter, RFID allows the identification and tracking of tags attached to object. 86

RGB Red Green Blue. An additive color model in which the red, green, and blue primary colors of light are added together in various ways to reproduce a broad array of colors. 83

RIS Radiology information system. A networked software system for managing medical imagery and associated data. 46

SDK Software Development Kit. A collection of software development tools that facilitate the creation of applications. 35, 85, 98, 113, 127, 143, 166, 167

TMIS Touchless medical imaging system. xvi, 43, 101, 114, 115, 122, 144, 148–151, 182, 188–190

ToF Time of Flight. A technique for resolving distance between the subject and the camera by measuring the round trip time of an artificial light signal. 21

UX User Experience. Concerning the user's overall experience with a product, UX design is the process used to provide meaningful and relevant experiences to users. 150, 163, 164

VGB Visual Gesture Builder. Software used to generate gesture recordings using the Microsoft Kinect V2. 95, 99, 100, 102

VR Virtual Reality. A simulated experience that can be totally different from the real world. Has applications in entertainment, education, and business. 35

WHO World Health Organization. A United Nations agency responsible for international public health. 3, 59

Chapter 1

Introduction

1.1 Motivation

Modern medical work features extensive use of digital imaging technologies, such as Picture Archiving and Communication System (PACS) to inform treatment decisions throughout the hospital. These technologies are an evolution from the days of analogue film use; digital interfaces have become the norm and have brought new efficiencies to clinical work. The transition to digital imaging technologies in hospitals has caused PACS to become ubiquitous in the clinical environment. This has brought benefits, such as reducing the time taken for radiology reports to become available.

At the same time, healthcare-associated infections (HAI) remain a significant issue in hospitals and reducing their incidence while improving HP work is a major challenge facing hospitals today. In the USA, HAIs cause ninety-nine thousand attributable deaths and cost six-and-half billion dollars every year [3]. In Europe, they result in sixteen million additional days spent in hospitals, thirty-seven thousand attributable deaths, and seven billion euro in costs every year [3].

Since the 1800s, antiseptic and aseptic surgical techniques have been the norm. However, surgical procedures would present a considerable risk of infection if there is poor hand hygiene or a lack of sterility in the surgical environment. The European Centre for Disease Prevention and Control estimate that 3.8 million people acquire an HAI each year in acute settings within the EU, Norway, and Iceland [4]. An estimated ninety thousand

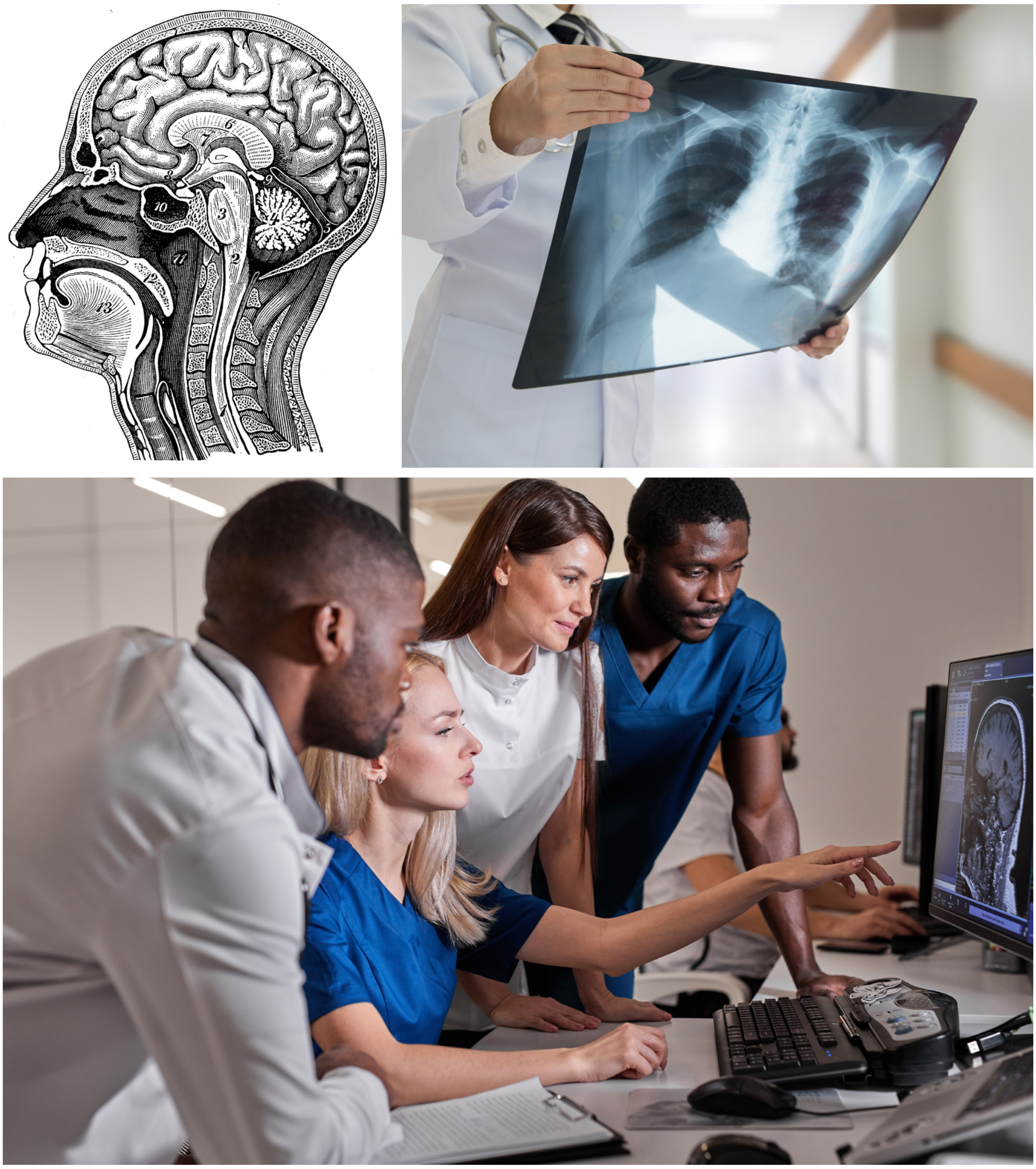


Figure 1.1: Top left: Prior to the development of medical imaging technologies, such as X-ray and MRI, healthcare professionals (HP) had to rely on anatomical diagrams in advance of surgery. Top right: For much of the twentieth-century medical imaging was presented on film stock. To review a scan, a HP would need to request a physical copy. Bottom: Modern medical imaging is stored and viewed digitally. Here a team review scans in a multi-disciplinary meeting. Images licensed from Shutterstock.

deaths in EU hospitals each year are attributable to the six most common infections in healthcare settings [5]. The economic burden of nosocomial infections (infections acquired during the process of receiving health care that was not present during the time of admission) for the EU 27 is estimated to be seven billion euro annually, not including indirect costs such as loss of income, or intangible costs such as physical and emotional pain [6]. Improved levels of sterility reduce levels of infection, resulting in reduced risk to patients, medical staff, and reduced levels of HAIs and associated costs.

Computers have further revolutionised medical care and the hospital environment in recent years. However, the introduction of computers and their benefits within the hospital environment is at an early stage of progress in clinical environments. The use of computers and related technology in hospitals comes with certain limitations. Computers and their associated peripherals are difficult to sterilise effectively; keyboards, in particular, represent natural havens for various pathogens [7]. For this reason, they are a potential source of infection. Germs are presented with the opportunity of colonising within the cracks of the mouse, the crevices of the keyboard, and on the surfaces of the touchscreen. There is growing evidence that each time a user interacts with the surfaces of equipment they are at risk of either having their hands contaminated by residual pathogens on the device, or of contaminating the device itself with their own hands [8]. In recent times, the COVID-19 pandemic has further highlighted the need to reduce the number of shared devices HPs touch. Reducing HAIs requires a multi-modal approach, including such measures as the provision of anti-bacterial gels, ongoing education and training in good hand hygiene practice, such as the WHO Five Moments and proper WHO Hand Hygiene Technique, and proper equipment sterilisation and cleaning protocols. However, when a health care worker washes their hands, it is unlikely that 100% of pathogens are eliminated [9]. A simple and effective solution for preventing contamination of surfaces by hand contact, and conversely of hands by surfaces, is to remove the need to touch those surfaces.

Touchless interaction technology has become more common in recent years. Depth camera technologies, in particular, have seen significant advancements. The advent of hardware technologies such as the Azure Kinect DK, and software solutions such as Google's MediaPipe [10], Manomotion [11], and NVIDIA's Deepstream [12] representing a new maturity in the field. Devices such as the Leap Motion Controller, Microsoft Kinect, and Apple's Face ID have brought affordable depth camera technology and gesture control into people's homes and offices.

The COVID-19 pandemic has placed new emphasis on many touchless technologies, with research into robust touchless solutions such as SigmaSense's hover technology receiving mainstream discussion [13]. Touchless interaction technology provides a number of clear benefits over traditional interaction methods, including improved sterility, the ability to interact with a system during hands-busy operations, and improvements to HP workflow in aseptic environments. It is clear that touchless technologies have reached a point of maturity whereby real-world systems can be realistically considered.

As the hardware and software required to build robust touchless systems for medical image navigation are now available, it is an appropriate time to investigate the design and implementation of touchless interaction modalities for medical imaging navigation systems, such as PACS.

To take a specific example, in the operating room (OR), a surgical user must maintain sterility where possible during a surgical operation. This presents a challenge to HP workflow, as it is not currently possible to directly manipulate medical imaging in a risk-free manner without breaking asepsis. To interact with key imaging, clinicians regularly have to break asepsis, rely on proxy users to interact with systems on their behalf, or simply forgo interacting with imaging entirely [14, 15, 16, 17]. In comparison, touchless interaction would allow the surgeon to directly interact with PACS as required while maintaining sterility. By removing the need for the lengthy process of scrubbing back in, touchless interaction enhances the clinical workflow both in terms of efficiency and sterility. Alternative solutions, such as sterile glove policies, and self-sterilizing screens and peripherals suffer limitations, with gloves being at risk of suffering micro-tearing, and self-sterilizing devices not being considered appropriate sterilization for environments such as the OR.

Touchless control outside the OR, in the wider hospital, would also contribute to improved levels of sterility. HAIs are a significant issue across many health care contexts, and hand contact is a significant vector for the spreading of pathogens throughout the hospital environment [18]. Other motivations for research and development in the field include; improvement in performance for 3D applications, resolving the busy hands interaction problem, improvements to clinical workflow, and naturalness of interaction.

The need to draw together existing research with the goals of bringing insight to acceptance by HPs and a design rationale for touchless interactions for PACS provides a motivation for this work. This dissertation explores the design and implementation of a touchless interaction system focused on potential applications for hospital environments. The needs and experiences of real-world HPs are investigated, and an experimental touchless PACS system is tested with experienced clinical users. Combining learnings from the literature, and implementing two separate prototype systems, the design path for a touchless system is set out. Each of the required design choices is investigated and discussed.

1.2 Problem Statement

Digital imaging systems, and the application software that support them, e.g., PACS, have brought significant improvements to clinical work. Such systems have improved HP access to resources, such as patient scans and radiology reports. However, these systems often require physical contact by the HP to manipulate the imaging. Increasingly, HPs are aware of the need to reduce contact with shared surfaces to help reduce the spread of pathogens; this is especially true in light of the recent COVID-19 pandemic. In aseptic environments, users are unable to manipulate key images and find themselves depending on circulating HPs who may be unfamiliar with the imaging software. Touchless technologies present a solution to these issues; however, there is little research into the experience of HPs with touchless PACS, or cohesive guidance toward approaching the design of such systems.

1.3 Research Objectives

This research addresses the problem of how to **approach the design decisions and trade-offs involved in developing touchless interactions for medical imaging systems**, integrating existing research, qualitative investigations of HP needs, learnings from developing two distinct touchless systems, and investigation of the performance and experiences of HPs with a prototype touchless PACS system.

In pursuit of these research aims, the following research questions are presented:

1. **RQ1:** What published work has been performed in the field of touchless interaction with medical image viewers, and what is the current state of knowledge regarding the design of touchless systems for hospital environments?

2. **RQ2:** How do HPs use PACS with conventional interaction modalities? What challenges do they encounter when using it? What are HP attitudes to adoption of, and experiences using, touchless interaction with PACS?

3. **RQ3:** When developing touchless interaction for PACS, how do we identify and approach significant design decisions for major elements?

4. **RQ4:** What are the opportunities and challenges presented by touchless PACS interaction when navigating medical imaging? How does clutching affect the user experience of using a touchless PACS interface?

1.4 Thesis Contributions to the Current State of Knowledge

This work makes several key contributions to touchless interaction in a clinical setting.

Firstly, a systematic analysis of existing published work is presented, providing an understanding of the history and current state of knowledge regarding touchless interaction with PACS (**RQ1**). This is in itself valuable as it not only provides context when reading the research presented herein, but will provides other researchers with an overview of the disparate contributions to this long-running research topic.

Secondly, the needs and experiences of clinicians are investigated and analyzed (**RQ2**). Existing research lacks extensive investigation of real-world HP experiences and requirements. The qualitative analysis conducted brings elements of the HP experience into focus through qualitative research using thematic analysis. These insights provide guidance and context when approaching the design of touchless medical imaging systems.

Thirdly, the decisions and trade-offs encountered when designing a touchless medical imaging system are investigated **(RQ3)**. Two separate systems were developed, an initial prototype built to investigate the major parameters in the design and development of these systems, followed by a second prototype combining the practical learnings from the initial prototype with those from interviewing HPs. The process of developing such systems ab initio required multiple design decisions. These decisions ranged from user considerations such as; ergonomics and cognitive load to technical choice of sensors and processing platforms, and the manner in which they were approached, are presented, allowing future researchers and developers to build upon this process in their own work. A set of design decisions informed through conducting this research are presented to provide a conceptual framework for future researchers and developers.

Fourthly, HP interaction experiences with a prototype touchless medical imaging system are presented. This work investigates several touchless clutch modalities (**RQ4**) and contributes additional insights into the design of hospital-based touchless systems. Findings, such as the relationship between touchless interactions and user authentication and security processes, provide promising avenues for future researchers to investigate.

1.5 Timeline

2014
Kinect V2 for Windows released to market.
July 2014
Though less successful than its predecessor, the Kinect V2 delivered several performance improvements. Discontinued in April 2015.

2017
Pilot study using first prototype.
November 2017
Using the Kinect V2 for Windows

2017
FaceID released to market.
November 2017
Enables iPhone users unlock their device using a touchless input mechanic. As of May 2022, iOS represents c.27% of market share.

2018
LUIS.ai released to market.
January 2018

2018
Touchless computer interfaces in hospitals: A review.
February 2018
Journal paper published in *Health Informatics Journal*

2019
Interviews with clinicians regarding PACS usage.
February 2019 to November 2020
Clinicians were interviewed both before and during the COVID-19 pandemic.

2019
BlazePose first commit to public GitHub.
June 2019
High performance software solutions that can run on mobile devices/generic camera hardware are becoming more common.

2019
Face Authentication HIDL released for Android 10.
September 2019
Enables Android users unlock their device using a touchless input mechanic. As of May 2022, Android 10 and above represent c.40% of market share

2019
COVID-19 first identified in Wuhan, China.
December 2019

2020
Azure Kinect DK released to market.
March 2020

2020
Experimental sessions with clinicians using second prototype.
December 2020 to March 2021
Using the Azure Kinect DK.

2021
Global 100,000,000 COVID-19 cases.
January 2021

2021
A Qualitative Analysis of the Needs and Experiences of Hospital-based Clinicians when Accessing Medical Imaging.
April 2021
Journal paper published in *Journal of Digital Imaging*

2021
Psychomotor learning theory informing the design and evaluation of an interactive augmented reality hand hygiene training app for healthcare workers.
October 2021
Journal paper published in *Education and Information Technologies*

2021
Thesis submitted.
November 2021

2022
Investigating Clutching Interactions for Touchless Medical Imaging Systems.
April 2022
Conference paper presented at *SIGCHI '22*

2022
Global 500,000,000 COVID-19 cases.
April 2022

1.6 Thesis Structure

The structure of the remaining chapters of this thesis is as follows:

Chapter 2 (p.11) Review of the research literature in the field of touchless computer interfaces in hospital settings. The focus is oriented to sterility and infection control as a major concern in the context of the hospital environment.

Chapter 3 (p.45) The methods, results, and discussion of a qualitative piece of research involving clinicians regarding PACS is presented. This research provides insights into the perceived needs and experiences of HPs with PACS.

Chapter 4 (p.75) A detailed analysis of the system design process and implementation of two touchless prototype systems is presented. The findings regarding the design process is the result of examining the existing literature, developing two representative systems, the second of which is based on the understanding of the needs of HPs gained through the qualitative study in Chapter 3.

Chapter 5 (p.115) Presents the methods, results, and discussion of experimental research conducted with clinicians investigating various clutching mechanisms for touchless navigation of medical imaging systems.

Chapter 6 (p.149) A characterisation of the development process for touchless interaction with medical imaging systems is presented, bringing together learnings from the existing literature (Chapter 2, p.11), an investigation of the qualitative needs of HPs with PACS (Chapter 3, p.45), learnings from the process of developing two separate prototype systems (Chapter 4, p.75), and HP experiences using a prototype touchless system (Chapter 5, p.115).

Chapter 7 (p.187) Concludes the dissertation by presenting the primary conclusions in conjunction with proposed directions for future work.

Chapter 2

Touchless computer interfaces in hospitals: a review

This chapter aims to address **RQ1** (Chapter 1, p.1) by presenting a review of published literature in the field of touchless interaction with medical image viewers. Researchers have been investigating the possibilities of touchless interaction in medical environments for quite some time, but the literature covers a range of contribution types in a number of domains. Given the recent improvements in touchless technology discussed in the previous chapter, it is an opportune time to consider the current state of knowledge regarding the design and implementation of touchless computer systems in hospitals. To that end, this chapter presents a review based on a systematic search of the literature. An update to this review is presented at the end of this chapter to consider more recently published work, and reflect on recent research trends.

2.1 Methods

2.1.1 Literature Search Strategy

The aim of the search was to identify papers concerning user interactions with medical devices or other information technology in a medical context, that does not involve touching them with the hands.

2.1.2 Selection Process

The literature review was performed across three databases:

Source	**Purpose**
ACM	To cover the field of computer science research
PubMed	To cover the field of biomedical research
Web of Science	To cover more general scientific research

Table 2.1: Databases included in literature review

Each database was searched using the same methodology, initially covering the period from January 2000 to January 2016. Results of an additional search covering the period from January 2016 to February 2021 is presented at the end of this chapter. The updated search is presented separately from the first search results to maintain the integrity of the initial findings. In keeping with recent practice, a further search was performed using Google Scholar to identify any papers of significance that may not have been returned when searching the other databases. Three distinct groups of key terms were combined to define a search space.

Group 1 search terms were; *“gesture recognition”*, *“voice recognition”*, *“speech recognition”*, *“gaze tracking”*, *touchless*, *contactless*, *hands-free*, and *touch-free*. These were used to restrict the corpus to papers that pertained to touchless control of any system/interface.

Group 2 search terms were; *hospital*, *medical*, *hygiene*, and *sterile*. These were used to refine the environment context and filter out those papers concerned with touchless interaction in other environments).

Group 3 search terms were; *interaction*, *interface*, *device*, and *control*. These were used to restrict the corpus to those papers in the medical devices/technology field.

After each search, paper titles were reviewed, and every paper with a title considered relevant was extracted for further analysis. Titles were deemed relevant based upon the presence of key terms in the title and reference to appropriate contexts, such as hospital or OR.

The abstracts of all papers selected were then read and analysed for relevance. Papers that were deemed relevant were then accepted for acquisition and complete review. The criteria used to determine relevance included; whether a paper contained key search terms included within the abstract, and abstracts that made reference to interaction with a medical device or computer. The final step was to read each paper to determine whether or not touchless interaction with computer interfaces in hospitals was a core theme of the paper.

2.1.3 Data Synthesis and Analysis

A summary of the search results returned by all three databases is given in Table 2.2 (p.13). For each paper, the motivations for using touchless technology were noted, along with the type of technology used, the user cohort, tasks and outcomes of any evaluation presented. User cohort is defined as the source group of individuals who used the system, and whose performance and feedback was collected and presented by the authors. Results include both quantitative and qualitative outcomes such as gesture recognition rates and subjective user experience of ease-of-use. Many of the papers discuss difficulties that were encountered, these were qualitatively analysed for common themes, along with any non-functional requirements (such as usability or reliability) reported.

Database	**Total results**	**Rejected at title**	**Rejected at year**	**Rejected at abstract**	**Rejected for no access**	**Rejected at paper**	**Final**
ACM Digital Library	1229	1208	2	5	0	5	9
PubMed	811	700	23	50	6	5	27
Web of Science	376	371	2	0	0	1	2 (both duplicates)
Overall	2416	2279	27	55	6	11	36 (38 with duplicates)

Table 2.2: High level search results (simplified)

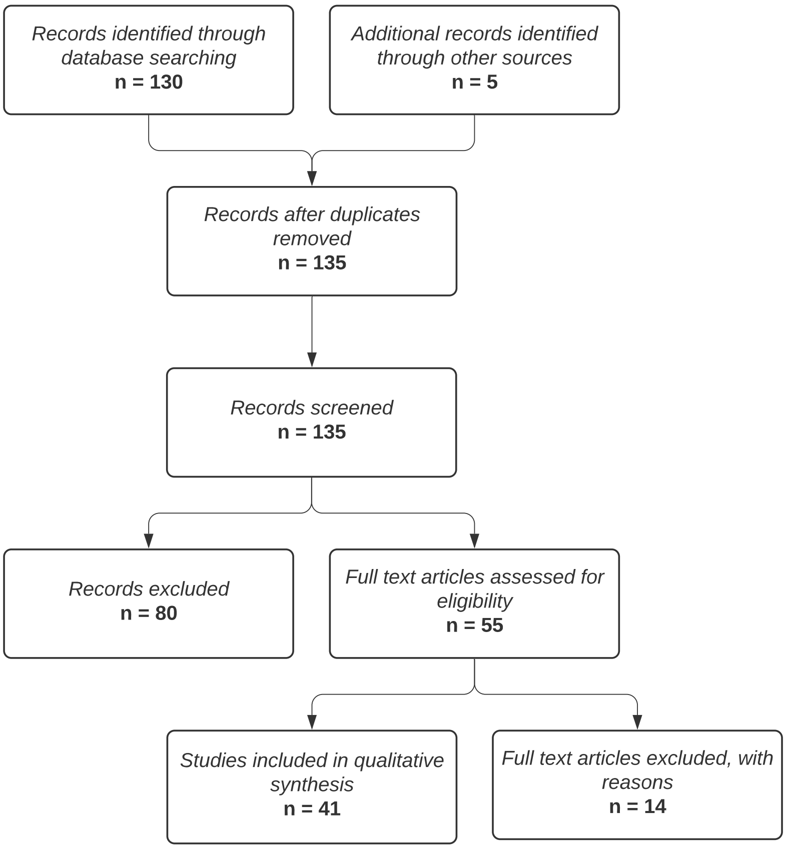


Figure 2.1: Study selection flow diagram

Country	Paper Count	Country	Paper Count
United States	14	Argentina	1
Germany	5	Australia	1
United Kingdom	4	Brazil	1
Italy	3	Colombia	1
Denmark	3	Czech Republic	1
Canada	3	Finland	1
Israel	1	Japan	1
Switzerland	1	—	—

Table 2.3: Distribution of papers by location of origin

2.2 Results

The main literature search resulted in the identification of thirty-six unique articles, all of which were obtained. An additional five papers were identified through Google Scholar (noted separately in Figure 2.1, p.14), resulting in a final corpus of forty-one unique articles. Before presenting the results, it should be noted that the corpus is highly varied, ranging from brief papers regarding technical feasibility to human-computer interaction papers with a rich discussion of user interaction. A breakdown of the geographical location of the research is presented in Table 2.3 (p.14). No studies were found that described clinical outcomes, therefore no meta-analysis is presented. The following sections describe the findings from the literature analysing themes, technologies, and contexts.

2.2.1 Motivations for Using Touchless Control

Sterility

Sterility is the most commonly cited motivation for touchless control (twenty-seven papers). Medical complications and infections arising from non-sterile interactions can be very costly [19], with both financial and human costs. It is noted by Wachs *et al.* that the current most prevalent mode of human-computer interaction (HCI) in hospitals remains the mouse and keyboard [20]. The keyboard and mouse by their very nature are a potential source of contamination (up to 95% of keyboards have been shown to be contaminated [21]). This problem is compounded by computers and their peripherals being difficult to sterilise [22].

While the increased use of technology in highly sterile settings such as the OR, and particularly the use of digital imaging, is noted, the need for access to non-sterile technology presents a problem: "*Unfortunately, the necessary divide between the sterile operative area and the non-sterile surrounding room means that, despite physical proximity to powerful information tools, those scrubbed in the OR are unable to take advantage of those resources via traditional human-computer interfaces*" [23]. Surgeons must either remember details from a prior review of each case, ask assistants to control devices for them, or use ad hoc barriers [23, 19]. One strategy used by some surgeons is to pull their surgical gown over their hands, manipulating the mouse through the gown [16]. As a result, surgeons are less likely to use computer resources if they have to step out of the sterile surgical field due to the time and effort of scrubbing back in [23]. Dela Cruz *et al.* state that breaks and interruptions to procedure workflow leads to an increased likelihood of medical error and poorer patient outcomes [24], and so should be avoided where possible. Cleaning to prevent bio-contamination during surgery after interacting with a computer can take up to twenty minutes, sometimes adding a full hour to surgery duration [25]. However, there are occasions when surgeons may need direct control of a system to manipulate images, to mentally "get to grips" with what is going on in a procedure [26].

3D applications

Twenty-one of the reviewed papers referred to 3D applications, such as manipulating the use of 3D imagery and data sets. A major advantage of 3D hands-free interaction when compared to conventional mouse use, which only operates in 2 dimensions, is the ability to navigate 3D data in 3D space [21]. Conventional analysis of 2D images is potentially cumbersome when interpreting the thousands of images that modern scanners can produce [25]. 3D interaction offers a more efficient navigation modality for 3D data [26].

Hands-busy

Five of the reviewed papers referred to the hands-busy problem, e.g., a surgeon holding surgical tools needing to manipulate a medical image [27]. Johnson *et al.* observe that at times, radiologists similarly, have their hands occupied while holding and manipulating wires and catheters [16].

Removing barriers

Touchless interaction enables not only a potential accelerator for specific tasks, e.g., image manipulation in an OR, but also provides interaction modes not previously available, as well as supporting hospital sterility. O'Hara *et al.* observe that *"the important point that we want to make here is not that these systems simply allow quicker ways of doing the same activities that would otherwise be performed. Rather, by overcoming these aspects* *of existing imaging practices, we lower the barriers to image manipulation such that they can and will be incorporated in new ways in surgical practices"* [27]. Currently, significant barriers exist to the usage of technology in sterile environments: "*In current practice, therefore, the use of modern technology in the OR is at best awkward and fails to realise its full potential for contributing to the best possible surgical outcomes*" [23].

2.2.2 Use Context

Operating Room)

The single most frequently considered use case for such touchless control is in the OR (mentioned in thirty-two papers) during surgical procedures, notably for the purpose of providing control of medical image systems such as PACS using the DICOM standard. Ten papers discuss PACS/DICOM in the OR context. Kim *et al.* found that it is feasible to manipulate surgical tools and execute simple, non-surgical tasks, such as controlling OR lights, imaging data, or positioning operating booms, using current commercially available, contactless tracking technology, such as the Microsoft Kinect [28].

Interventional radiology

Interventional radiology was discussed as a use context in eleven papers, ranking it the second most considered use case for touchless control. Image interaction in interventional radiology takes place within a surgical context rather than a purely diagnostic context, which makes sterility a key requirement [16].

Tasks

Table 2.4 (p.20) lists the procedural tasks and user cohorts studied for evaluations where this has been specified in the literature. As can be seen, the most common system application for touchless, gesture-driven interfaces has been various forms of image navigation, closely aligned with the OR and interventional radiology (IR) contexts listed above. Medical image navigation is the goal in [21, 29, 30, 16, 22, 19, 17, 27, 20, 31, 23, 32, 33, 34, 25, 35, 36], with the more specific subset of MRI image navigation being the goal in [30, 16, 22, 27, 20, 23]. The intended task context was also significant in identifying tasks for testing the systems implemented, such as the measurement of a lesion in images [29].

Table 2.4: Tasks, users and outcomes

Task	User cohort	Sample size	Main reported outcome
Medical image interaction[22]	Surgeons	10 users	Mean recognition accuracy of 97%. The system was deemed to be easy to learn and to remember, and moderately easy to perform.
Medical image navigation[19]	Radiologists	29 radiologists	69% believed the system could be useful in an interventional radiology practice.
Medical image navigation[23]	Surgeons	6 procedures	Users reported that they accessed more medical images than normal using the system.
Moving, zooming, windowing medical images [32]	Medical professionals	10 users	Mouse control was found to be faster than Kinect control, both with and without previous experience of the medical image viewing software.
Measuring a lesion on images[29]	Radiologists	29 users	The iPad was found to be the most usable, and the Kinect the least usable, with tasks taking nearly four times as long with the Kinect.
navigation[20]	Surgeons	1 operating procedure	The system was found to be easy to use, responsive, and fast to train with. Gesture recognition accuracy of 96%.
Locating an aortic stent and navigating to bifurcation [33]	Biomedical engineers	10 users	The inertial sensors were shown not to inhibit the user's movement. The system was deemed to be responsive and precise, albeit slower than a mouse and keyboard.
Medical-data browsing[21]	Hospital employees	20 users	Calibration took less than 10 seconds, and task completion rates were 95% or higher.
3D medical data navigation[25]	Unspecified	18 users	Users completed tasks faster when using interfaces with appropriate numbers of degrees of freedom.
Presentation control[37]	A Professor	99 gestures	Approximately 25% of gestures were not fully recognised because they were not performed in the working space or the gesture was not performed properly.
Classifier performance[21]	Unspecified	15 users	The system operated at 13.8 fps without classification, and at 10.69 fps with classification.
Varied[21]	Primarily computer science students	7 users	On a Likert scale of 1 to 5 (5 being the best) the system scored: response time of the system: 4.14; Adaptation to the system: 3.57 ; Comfort of gesture set: 4.0 ; Intuitiveness of the gesture set: 4.0.
Arbitrary parameter adjustment [38]	Unspecified	12 users	Users significantly improved their performance with practice over a small number of repeated uses of the system.

Task	User cohort	Sample size	Outcome
A 10 step predefined scenario [34]	Computer science researchers	5 users	Gesture recognition accuracy was 74% for foot gestures, and 79% for hand gestures. User feedback was generally positive, especially noting simpleness of use.
Controlling an operating room table [39]	Unspecified	10 users	The system was graded as an above-average interface.
Controlling an interactive hospital room [40]	Unspecified	18 users	97% voice and 93% gesture accuracy. Users prefer to perform whole tasks using only one type of interaction.
Peg transfer and pointing tasks [28]	Experienced surgeons	5 users	The da Vinci system had the lowest latency, the lowest tremor radius, and the fastest time to task completion when compared to the 3Gear and Mantis systems.
Laparoscopic cholecystectomy procedures [41]	Instructing surgeons (6), operating surgeons (29) and camera assistants (48)	83 users	No negative effects on surgery completion time when using wireless hands-free surgical pointer (WHaSP). The WHaSP was found to be comfortable, easy to use, and easy to control. Further, the WHaSP improved communication effectiveness in the OR.
Voice input to Electronic Health Record (EHR) [42]	Nurses(7) and others (3)	10 users	Voice input took an average of 304.5s per record versus 1459s for keyboard input.
Creating and editing medical reports using voice control [43]	Pathology assistants (48), residents (12) and attending physicians (20)	80 users	Use of voice recognition has led to a marked reduction in report turnaround time (554min to 102min).
Creating and editing radiology reports using voice control [44]	Radiologists	7 users	Reports generated using voice recognition were approximately 24% shorter in length and took 50% longer to dictate than those transcribed conventionally. The reports generated using voice dictation also contained more errors (5.1 errors/report vs 0.12 errors/report).
Creating and editing radiology reports using voice control [45]	Radiologists	2 users	Productivity for one radiologist was calculated at 8.6 reports per hour using voice recognition and 13.3 reports per hour using a transcriptionist.
Voice recognition of sentences [46]	Users with no medical background	8 users	The system performed at a higher rate of classification in command mode than free speech mode (81.6% vs 77.1%). Training the system without background noise improves recognition rate (85.5% vs 77.8%).

Task	User cohort	Sample size	Outcome
Updating anaesthesia record using speech recognition [47]	Doctors and nurses	12 users	Users found it slightly more difficult to update the anaesthesia record by voice, although voice input required significantly less time than traditional input, and almost two times as many medication registrations were made with voice input as compared to without voice input.

Context-sensitive systems

A "context-sensitive system" is a system that is affected by, or reacts to the context it appears in. Whilst the term context-sensitive systems was not found within the corpus, the concept of "context" was found in five of the papers. Several papers investigated the relationship between context and system functionality. Contextual information, such as the focus of attention, can be employed to improve recognition performance [22]. Another method for augmenting recognition accuracy is switching between vocabularies of gestures based upon context as suggested by Wachs *et al.* [48]. This approach could improve performance in gesture-recognition systems, as discrete gesture recognition algorithms are used for smaller gesture subsets. There are a number of contextual cues that should be considered for touchless interaction systems, both from a situational and an individual context. From a situational context, for example, there are specific activities in the OR that a surgeon can be expected to be engaged in, such as focusing on the patient and the key image, as well as remaining proximal to the patient while also considering information in key digital images [20].

2.2.3 Touchless Control Technologies

Available touchless control technologies have improved markedly. At the time of publication of [30], no reliable, mature technology for effective gesture control existed. Researchers have since investigated a broad range of technologies; eye gaze technology (EGT) [49], capacitive floor sensors [34], and inertial orientation sensors [34, 33], colour cameras such as the Canon VC-C4 [20], the Loop pointer [29], MESA SR-31000 time-of-flight (ToF) cameras [21], Siemens Integrated OR System [50], wireless hands-free surgical pointer [41], the Apple iPad [29], Leap Motion Controllers [37, 38, 36], and the Microsoft Kinect ToF) camera G1 [29, 30, 19, 31, 25, 39, 40, 28]. Data generation frequencies ranged from devices with 15fps output, to the Leap motion controller with greater than 100fps output. Chao *et al.* suggests that adopting a stereo camera configuration might improve accuracy for touchless interaction [29]. In addition to advances in sensor hardware, there have also been complementary developments in available software for implementing touchless systems. Many of the papers that utilised the Microsoft Kinect also used specialised software, such as Primesense drivers [29], OpenNI [29, 22], and skeletal tracking software NITE [29]. Studies described in the papers involve using purely gestural commands, and combinations of gesture and voice commands (a form of multi-modal interaction).

The primary modes of human communication are speech, hand and body gestures, facial expressions, and eye gaze [20]. Much of the human-computer interaction literature on these forms of interaction draws motivation from these "natural" modes, rather than the touchless properties of these interaction modalities: "*Gestures are useful for computer interaction since they are the most primary and expressive form of human communication*" [48].

Only one paper, Chao *et al.* directly compared the efficacy of multiple devices. In their paper, the authors compared the Microsoft Kinect, Hillcrest Labs Loop Pointer, and the Apple iPad. Theirs was the only paper to report using either the Hillcrest Labs Loop Pointer or the Apple iPad. Their results reported the Apple iPad to have had the greatest number of participants with prior experience of the device, and the Hillcrest Labs Loop Pointer to have had the least prior experience [29]. They report the Apple iPad to have the greatest usability score, as well as the shortest completion time for a sequence of measurement tasks (mean usability score: 13.5 out of 15; mean completion time: 41.1 seconds), followed by the Hillcrest Labs Loop Pointer (mean usability score: 12.9 out of 15; mean completion time: 51.5 seconds), with the Microsoft Kinect scoring the lowest usability sore and the longest completion time (mean usability score: 9.9 out of 15; mean completion time: 156.7 seconds) [29].

Gesture recognition

Gesture interaction is claimed to be intuitive because users are familiar with communicating with other people by means of gestures [39]. Twenty of the papers discuss the design and implementation of gesture recognition systems. Wachs *et al.* discuss a number of forms of analysis for hand-gesture recognition [48] (Table 2.5, p.22).

Method	**Observation**
Motion	Effective and computationally efficient
Depth	Deemed to be potentially useful for separating hands from the background
Colour	Heads and hands can be located with reasonable accuracy using only their colour
Shape	If the object is clearly differentiated from the background, can reveal object orientation
Appearance	More robust but with higher computational overhead
Multi-cue	A combination of the previous approaches

Table 2.5: Forms of analysis for hand-gesture recognition

Conventional interfaces via gesture or native gestural interfaces Several papers have attempted to apply gestures directly to conventional means of interacting with a computer [21], for example, Rosa and Elizondo [36] implemented a virtual touch-pad suspended in mid-air. However, adhering to existing control design paradigms such as mouse and keyboard control design has been found to be a limitation [19]. Others have created complete gesture sets with purely gestural control in mind [29, 30, 27]. Designing a system with gestures as a requirement rather than replicating touch paradigms would also help to reduce the issues associated with adhering to existing mouse and keyboard control design according to [19]. Other systems have used gesture modalities to provide more direct control within medical procedures. Such a system is FAce MOUSe where a surgeon can control the direction of a laparoscope simply by making the appropriate facial gesture[20].

Gesture set Selecting the appropriate gesture set is a key element in the system design process. The hardware characteristics of input devices in conjunction with the application domain must be considered [30]. There have been a large number of gestures described in the literature, some, however, are more common than others. Table 2.6 (p.23) classifies the gestures encountered into categories; system control, content control, and input, with content control being the largest category. Despite the extent of this list of gestures, O'Hara *et al.* state that limiting the number of gestures benefits ease of use, as well as learnability [26]. Furthermore, limiting the gesture set can enhance reliability and avoid "gesture bleed" (whereby gestures containing similar movements are mistaken for each other by the system).

O'Hara *et al.* discuss expressive richness as *"how to map an increasingly larger set of functional possibilities coherently onto a reliably distinctive gesture vocabulary"*, as well as how to approach transitioning between gestures [27].

System control	**Content control**	**Input**
click [21, 37, 39, 38, 36]	translation [21, 30, 39, 25, 39]	measuring [29, 36]
unlock [27, 23]	rotation [21, 30, 27, 36]	cursor [21, 30, 37, 23, 36]
reset [21]	scrolling [29, 23]	entering a value [37, 38]
windowing [29, 30]	zooming [29, 30, 27, 34, 38, 36]	
set idle [30]	panning [29, 27]	
lock [27, 23, 39]	ROI extraction [30]	
	ROI erasing [30]	
	animating [30]	
	navigation [37, 23, 34, 39, 36]	
	complex 3D [37, 25, 36]	

Table 2.6: Complete list of gestures found in the literature set

Rossol *et al.* designed an interface with the purpose of using gestures designed to be equally efficient to use with bare hands or hand-held tools [38]. In order to minimize complexity and cognitive load, their system used a vocabulary of three gestures for finger or tool-tips, and one gesture for hands. Furthermore, they implemented a means of performing highly precise adjustments by means of tapping gestures on one hand while using the proximal hand for parameter adjustment [38].

Considering fine detail performance, depth cameras can now track fine motions of hands and fingers [28]. Tan *et al.* flagged issues with hand tracking and inconsistent responsiveness as issues, along with their stylistic choice of requiring two hands for gestures [19]. Wachs *et al.* discussed the issue of lexicon size and multi-hand systems where the challenge is to detect and recognise as many hands as possible [48].

Wachs *et al.* flagged system intuitiveness as an issue, i.e., there is no consensus amongst users as to what command a gesture is associated with [48]. Dealing with differences in gesture execution between individuals is a considerable challenge [30]. Wachs *et al.* state that there are many different anthropometric characteristics to be accounted for, in addition to variations in the types and numbers of gestures [48].

The need to adjust continuous parameters may also be a possibility in gestural input - O'Hara et al. use a combination of voice commands for discrete commands and mode changes, and gestures for the control of continuous image parameters [27].

Soutschek *et al.* found that a majority of processing time was spent on acquiring images and preprocessing [21], e.g., resizing. However, as computers have become more powerful, such processing is now easily accomplished allowing for more sophisticated real-time interactions. System performance and user familiarity impact directly on the user experience. Soutschek *et al.* assert that users dislike systems when there is a perceptible delay during use [21].

Unintended gestures and clutching Not all gestures are intended as commands, e.g., pointing out a feature of interest to a colleague [37], and the misinterpretation of such gestures can adversely affect system behaviour and the user experience. The inclusion of a lock and unlock gesture (an example of a clutching mechanism) is essential according to Mauser and Burgert [37]; O'Hara *et al.* used a voice command to lock and unlock the system [26]. The aim is to ignore inadvertent commands: the system should be inactive until hailed by a distinctive action and should be locked using another distinctive action [23]. Deliberate gestures and unintentional movements need to be distinguished from each other. Unintentional movement usually occurs when the user interacts with other people, or is resting their hands [48]. Rossol *et al.* found unintentional finger-or tool-tip movement being interpreted as input, to be a drawback with their system [38]. In order to combat overlapping gestures, they suppressed any recognized gestures that overlapped the previous gesture's time window [38]. Tan *et al.* state that fine movements were the most difficult to deal with [19]. How to define the start and end of a gesture [30, 37] is also a difficult issue. One advantage of depth cameras is the ability to take account of motion in the Z plane to reduce unwanted gestures when returning to idle [30].

Physical issues In practice, gesture recognition works best at particular distances from the sensor [17]. When designing gestures, O'Hara *et al.* looked for ways to facilitate the work of clinicians, while maintaining sterile practices, by restricting movements to the spatial area in front of the torso [27]. Tracking movements relative to the torso may be the most appropriate [16]. Information from the operator's upper limbs and torso can be used to implement the functionality of a mouse-like device [23]. Regarding the 3D gesture zone, *"this zone extends roughly from the waist inferiorly to the shoulders superiorly and from the chest to the limit of the outstretched arms anteriorly and to about 20cm outside each shoulder laterally"* [23]. By using environmental cues for intent, Jacob *et al.* allowed users to perform gestures anywhere in the field of view [22]. However, depth segmentation has required an upper and lower threshold [51], meaning that to use a system the user cannot be too close or too far away from the sensor and must remain within the zone.

Comfort and fatigue When considering the issue of interaction space, Wachs *et al.* ask is it correct to assume that the user and device are static and that the user will be within a standard interaction envelope [48]? Tan *et al.* also say that ample space is required to operate their touchless system [19]. Sufficient operational space is one of the ergonomic factors affecting user comfort. This relates to the issue of fatigue, and the need to avoid intense muscle tension over time. Consideration of both static ("the effort required to maintain a posture for a fixed amount of time"), and dynamic ("the effort required to move a hand through a trajectory") stress respectively is key in promoting user comfort [48].

User training Systems should be easy to integrate into existing ORs with minimal distraction, user training, or human resource effort according to Strickland *et al.* [23]. Rosa *et al.*, state that with moderate training of the user, use of their gesture interface is easier and faster than changing sterile gloves or using an assistant outside the sterile environment to operate the imaging system [36]. In O'Hara *et al.*, the surgical team became familiar with the system through ongoing use of the system, "learning on the job", supported by prompts from the lead surgeon who acted as a champion for the system [27]. Chao *et al.* determined that prior use of a device had a significant impact on task completion time[29], and found that gamers were faster on all devices [29]. However, Ebert *et al.* observed no significant difference between gamers and non-gamers [32].

System training and calibration A gesture classifier typically needs to be trained by providing a large set of sample data. Jacob *et al.* states that it is imperative to use as large a number of users as possible (high-variance training data) [22]. System calibration results in a time cost. For example, Wachs *et al.* reported that the total setup time, including calibration, was circa twenty minutes for their system [20]. Calibration is an additional requirement in some systems that use a Kinect [30]; this too incurs a time cost.

Voice control

Perrakis *et al.* believe that voice control has an important role to play in minimally invasive surgical procedures, allowing the surgeon to take control of the entire OR without breaking sterility or interrupting the surgery; this potentially allows for single-surgeon procedures, resulting in reduced costs and greater use of surgical resource [50]. Mentioned in twenty-nine papers, and discussed on a practical level in twenty papers, voice control has been found to be slower but more accurate than gesture control, and both were reported as slower than traditional interaction methods[40]. Two major issues encountered using voice control are people's accents [22], and ambient noise, with the noise levels in an OR greatly impeding accurate voice control [20].

When designing a grammar for a speech recognition based interface, care must be taken to select words that are easily recognisable for the various users of the system, and sufficiently distinct from each other phonetically, to avoid possible overlap and misrecognition [46]. Strickland *et al.* suggest that the implemented gesture vocabulary does not need to support the full functionality of the PACS system, but rather focus on a subset of the most commonly used functions [23].

Voice control for text input Thirteen papers discussed voice control as a means of inputting text. Given a sufficiently high speech recognition engine confidence score, use of a keyword to activate the system, and that of another keyword to switch between command based and free text modes may allow for a completely hands-free approach [47].

Voice recognition has been described as having functional accuracy rates as high as 99%, however, some studies have shown somewhat lower accuracy rates than human transcription [43]. According to Kang *et al.* the largest benefit of using voice control is a decrease in turnaround time, which results in higher administrative, and clinician satisfaction, whereas the biggest disadvantage is an increased editing burden on clinicians [43]. They report natural dictation at speeds of 160 words per minute [43]. Marukami *et al.* investigated voice recognition input to an Electronic Nursing Record System (ENRS), and compared the time taken to input records to the EHR using voice recognition with that of keyboard input, finding that users were able to input records roughly five times faster using voice recognition (304.5s vs 1459s) [42]. In contrast, Pezzullo *et al.* found that reports generated by means of the voice recognition system were 24% shorter in length and took 50% longer to dictate than those transcribed conventionally [44]. In terms of cost per report, Puzzello *et al.* note that the use of a voice recognition system may result in a 100% increase in dictation costs [44], caused by the significant difference in cost per hour of radiologists compared to transcriptionists [45].

Voice control for discrete commands Voice control is generally deemed good for discrete commands, though it is not considered appropriate for continuous parameter adjustment, for which gestures are better suited [27]. Dictation systems require discrete commands, with Argoty *et al.* proposing two-word commands as more meaningful to the user than single word commands [40]. Nagy *et al.* say that increased command length plays a significant role in improving recognition hit rates [52]. Alapetite found that voice recognition displayed higher accuracy when issuing discrete commands (81.6%) as compared to free speech mode (77.1%) [46]. Use of a voice recognition interface resulted in a significantly higher quality of anaesthesia recording when compared to the traditional typed interface (99% of medications recorded versus 56%), as well as a reduced error rate[47]. In contrast, Pezzullo *et al.* found that their voice recognition system resulted in more errors per report than conventional transcription (5.1 errors/report compared to 0.12 errors/report), and go on to suggest that “radiologists are not good transcriptionists” [44].

Training for voice control Hoyt *et al.* state that the success or failure of voice recognition technology in a hospital is dependent on personal experience, training, technological or logistical factors [53]. To this end, voice recognition vendors may provide “train the trainer” sessions to users with high levels of aptitude (“superusers”) [53]. Rossol *et al.* found that users can significantly improve their performance with practice over a small number of repeated uses [38]. In Kang *et al.* new users undertook a one-hour training session, and subsequent setting up a new voice profile took approximately 10 minutes [43]. Alapetite found that setting up and training a new voice profile took an average of 30 minutes [47].

Fatigue using voice control Marukami *et al.* considered the issue of user fatigue when implementing voice recognition as compared to keyboard input, and gathered user feedback regarding both input modes by means of a questionnaire [42]. Their results indicated that users found that voice input caused less fatigue and was easier to use compared with keyboard operation, despite being inexperienced with voice input [42].

Time of day also seems to impact on performance. Luetmer *et al.* identified an increased error rate in laterality in radiology reports during the evening and overnight shifts (0.154% during the evenings, 0.124% overnight, compared to 0.0372% during the day) [54]. Laterality errors lead to misidentification of the side (left or right) or the locus of pathology or injury. They also found that reports generated using voice control had similar laterality error rates as those generated without using voice control [54].

Performance of voice control Alapetite found that the voice recognition interface led to shorter action queues than the traditional system interface, but users found that it required slightly more concentration and was more difficult to make updates to the anesthesia record using voice recognition input [47]. Pezzullo *et al.* declare that with diminished speech recognition accuracy comes an increase in time spent editing reports, resulting in a decrease in user efficiency and satisfaction [44].

Alapetite observed a speech command must be said in one go, distinctly and without any dysfluency [47]. Perrakis *et al.* found that speaker's accent did not have a significant impact on system accuracy, with functional errors in using the system being approximately the same for non-native and native German speakers [50]. Alapetite found that when a user speech profile was trained in the presence of background noise there was a slight increase in free speech recognition performance (78.2% vs 75.6%), and stated that "background noises have a strong impact on recognition rates" [46].

Other technologies

Nine papers discussed the use of EGT as an interaction modality. Modern eye tracking cameras have the advantage of being very easy to install, calibrate, and use [49].

Two papers discussed inertial-type sensors attached to the users' bodies to capture gesture input. Jalalinya *et al.* stated that advantages of such a system were that the system did not require a direct line of sight for the user, and that it would allow only a designated person (the wearer of the sensor) to interact with the system [34]. Bigdelou *et al.* discussed the hardware issues of inertial orientation sensor-based systems, highlighting issues such as noise and drift [33].

2.2.4 Non-functional Requirements

The most commonly referenced non-functional requirements in the corpus relate to usability and reliability of gesture and voice control. System design should consider real-time interaction, sterility, fatigue, intuitiveness/naturalness, robustness, ease of learning, unencumbered use, and mechanisms for managing unintentional commands [20]. Another requirement identified is the need for low cognitive load, through the use of short, simple, and natural gestures [48]. Natural interaction is taken to include the use of voice and gesture commands [48]. Furthermore, it is suggested that systems should focus on being stable and providing essential functionality rather than attempting to be more powerful and versatile [23]. The system needs to support both coarse and fine-grained system control through careful design of the gesture vocabulary [16], and how these gestures are mapped to control elements in the interface [16]. Reliability is identified as a key non-functional requirement [23], which is impacted in particular by the issue of unintentional gestures. Gesture control interfaces may need to sense the human body position, configuration and movement in order to achieve this [30].

2.2.5 Evaluations

A variety of experimental outcomes are studied in the literature (Table 2.4, p.20) with the accuracy of gesture recognition being the most frequently reported outcome (seven papers). There are a number of criteria that should be considered when evaluating a touchless system [48]. Validation of sensitivity and recall of gestures, precision and positive predictive value, f-measure, likelihood ratio, and recognition accuracy should all be rigorously evaluated using standard, public data sets [48]. Furthermore, usability criteria such as task completion time and subjective workload assessment, as well as user independence should be evaluated. The usability of interfaces is described relative to different standards, which focus on efficiency, effectiveness and user satisfaction [39]. Soutschek *et al.* deem aspects such as classification rate, real-time applicability, usability, intuitiveness and training time as relevant for evaluation [21].

Both quantitative and qualitative assessments of a system should be carried out comparing gesture interaction to other technologies such as voice control and mouse and keyboard [48]. Subjective evaluation by experienced clinicians is important, and are likely to be more insightful than purely technical factor comparisons [29]. This requirement introduces an obstacle in terms of gaining access to potentially large numbers of qualified practitioners, as well as introducing ethical and safety concerns for conducting real-world evaluations. However, a number of the papers have performed evaluations with representative end-user cohorts, for example, Tan *et al.* asked 29 radiologists to evaluate their system for efficacy as well as determining possible advantages and disadvantages [19].

Technical evaluations

In Jacob *et al.*, development and validation involved three steps, lexicon generation, development of gesture recognition software, and validation of the technology [22]. Other authors have evaluated their systems from both technical and subjective (user experience) perspectives[21]. When performing a technical evaluation, data regarding accuracy needs to be collected and analysed [21]. Technical evaluation might focus purely on the accuracy of gesture recognition at the early stages of development, or on task performance at later stages, during which performance time can also be measured. The choice of realistic tasks is important for such evaluations. High accuracy is a requirement for medical implementation, with an accuracy of 95% upwards suggested for use in a medical context [21].

Feasibility

With the advances in technology using touchless technology, the focus is no longer on technical feasibility. Rather, it is important that we understand how systems and their design impact the patterns of behaviour of hospital staff [27]. However, it is claimed that much existing work from a medical background lacks consideration of real-world context. Implementations remain experimental, and work originating from a technology background often suffers from over-simplification and abstraction from real-world medical complexity [16].

User acceptability and satisfaction

A range of methods have been used for performing qualitative evaluations including contextual interviews, individual questionnaires, and subjective satisfaction questionnaires. Wachs *et al.* and Ebert *et al.* made use of subjective questionnaires to determine user satisfaction with the study system [20, 32]. Questionnaires may be used to probe themes such as previous task experience, perceived ease of task performance, task completion time, and overall task satisfaction [20]. Robustness is key to the acceptability of a system; Wachs *et al.* specifically mention robustness for camera sensor and lens characteristics, scene and background details, lighting conditions, and user differences [48].

Overall there is a consensus in the literature that systems should be subject to both technical and qualitative evaluation using public data sets, and demonstrate a very high level of gesture recognition success in order to be appropriate for medical use [21]. There is also a recognised need to minimize unwanted side effects such as accidental gesture recognition.

2.3 Updates to the Literature Review

In February 2021, a second search was performed on each of the databases, using the same methodology, covering a period from January 2016 to February 2021 in order to provide an up-to-date review of the literature. This section presents findings from these sources separately from the initial search results in order to maintain the integrity of the initial analysis and findings. This updated literature search resulted in the identification of twenty unique articles, all of which were obtained. The results of this search are summarised in Figure 2.2 (p.33).

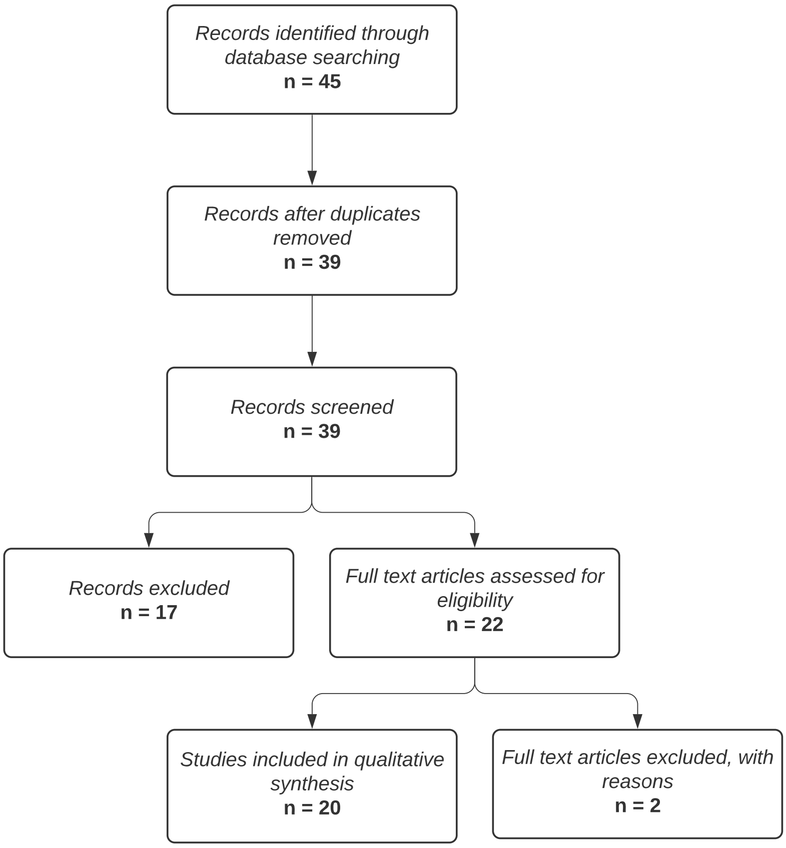


Figure 2.2: Study selection flow diagram (January 2016 - February 2021)

2.3.1 Sterility as a Motivation for using Touchless Control

The need to enhance sterility continues to provide a motivation for touchless control. There is an ongoing need to reduce the potential spread of infectious pathogenic microorganisms that present occupational hazards which can be transmitted by physical contact [55].

Recent work has seen further, and more nuanced investigation of touchless control in the OR, in particular addressing the interaction gap between HP and key imaging. The need to maintain sterility can have a negative impact on HP workflow; there exists an interaction gap between pre-operative planning/image review and intraoperative control of key imaging. Improving intraoperative control of key imaging would allow for additional planning opportunities during surgical procedures [56]. Despite the value of timely and comprehensive access to pertinent imaging in the sterile field, allowing operators to rapidly modify their therapeutic plan through an analytic decision-making process, existing solutions for controlling key imaging in the OR are often time-consuming, inefficient, error-prone, and disrupt the workflow [57]. Currently, HPs may choose to directly control imaging using a mouse and keyboard (which Zaman *et al.* note is not possible without breaking asepsis) or they may instruct an assistant to interact with the imaging on their behalf. While this second option does not break asepsis, if there is a difference in levels of subject matter understanding, there can be unwanted complications and undesirable cognitive and workflow disturbances [56, 58]. This lends further weight to the previously flagged issues with the use of proxies in subsection 2.2.1 (p.15) and subsection 2.2.3 (p.22).

2.3.2 Context-sensitive Systems

Understanding the type of user that will use the touchless system as well as the appropriate operating scenario, can inform the design process. For example, subject expert users, such as a neurointerventionalist manipulating an angiography, may exhibit shorter learning times than other users; the user's experience with the subject matter may be more important than other factors in achieving consistent success with a touchless system [59]. Alternatively, younger users, such as medical students, may exhibit shorter learning times due to greater familiarity with technology in general (ibid.). Contextual enquiries and user interviews can be used to gain an understanding of operating context when developing touchless medical image systems [2].

Operating room

The requirement to maintain asepsis can present challenges when considering how to enable direct control of key imaging. However, due to the ongoing advances in image-guided therapy, direct image control is becoming increasingly valuable in image-guided therapies, resulting in new use cases and requirements for touchless control in the OR. An example of this is the increasing reliance among endovascular specialists on complex imaging datasets to “formulate or to alter their therapeutic plan during procedures” [58]. Furthermore, the HP's hands are often occupied with surgical tasks, such as handling medical instruments; interacting with controls requires these tasks to be interrupted [60].

These challenges result in HPs seeking more effective methods of interacting with key imaging [61]. Touchless interaction devices provide a suitable solution, allowing greater levels of clinician autonomy in carrying out image interaction tasks during surgery [62]; this can lead to greater use of medical images to support decision making and instruction intraoperatively [26].

Technical challenges

The OR can present unexpected challenges when using certain touchless sensors. For example, the heat from powerful halogen lights in an OR can have a negative impact on the performance of sensors that rely on infra-red, such as the Kinect V1 [55]; such an effect could have an impact on clinical work. Furthermore, it may not be possible to position a sensor in an ideal location in the OR. In such instances, devices with short operational ranges, such as the Leap Motion, may not be appropriate [63].

2.3.3 Technologies

It is clear that the release of the Kinect V1 in 2010 (and its associated software development kit (SDK) in 2012) instigated a significant increase in the number of publications looking at touchless interaction in hospitals and ORs, peaking in 2014 (Figure 2.3 (top)). However, since 2015 there has been a marked decline in such publication (Figure 2.3 (bottom)); this may be due to an increased research focus on virtual reality (VR) and augmented reality (AR) applications, both amongst researchers and companies developing touchless technologies (such as Microsoft or Ultraleap) [64]. The high level of publications using similar approaches to designing gesture-controlled medical imaging viewers suggests there is no longer a need to produce novel gesture interfaces, but rather invest greater effort into improving and evaluating usability and intuitiveness [1].

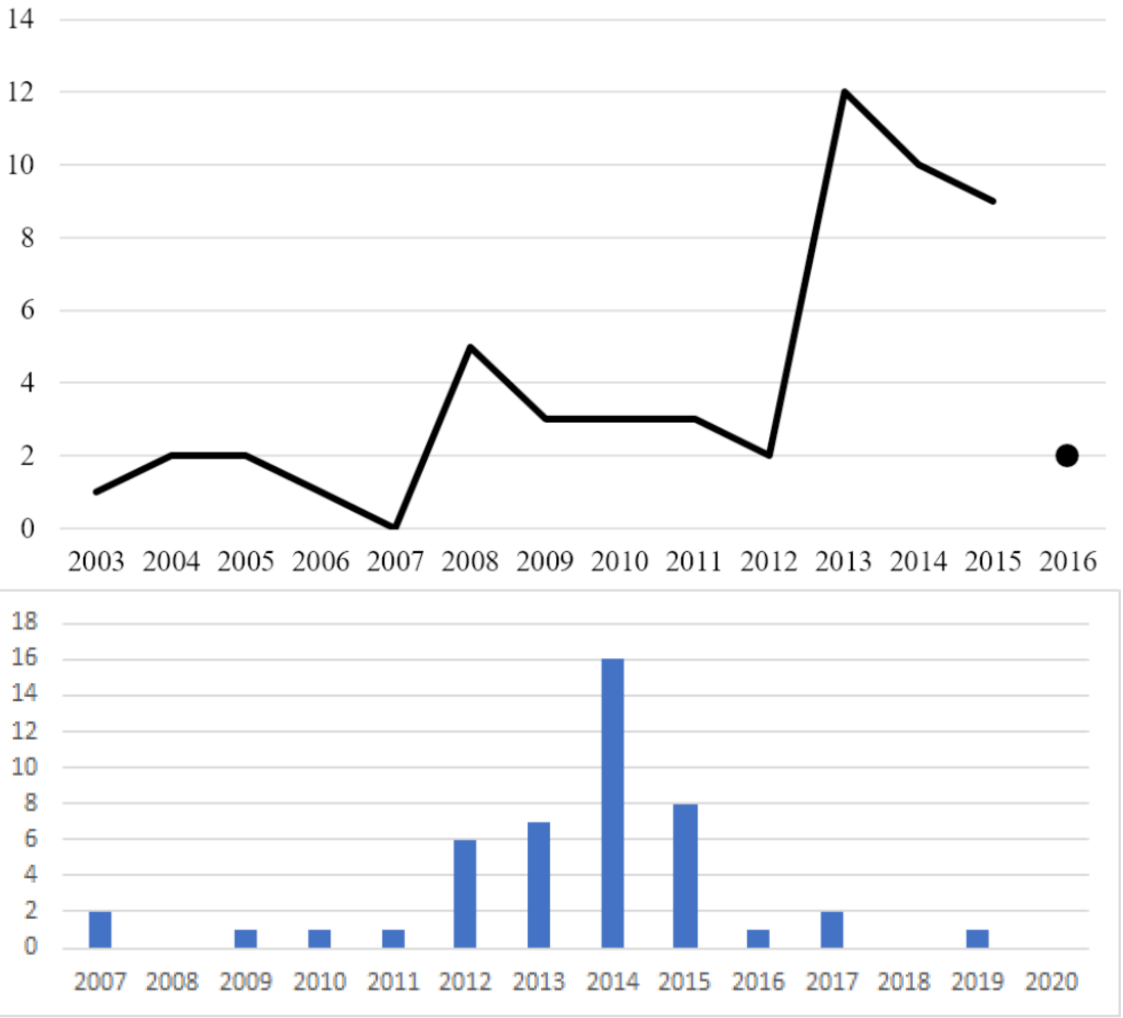


Figure 2.3: Top: The number of published papers dealing with touchless control of medical software in the immediate environment of the OR and IR suite between 2003–2016 (PubMed) [1]. Bottom: The number of published papers dealing with natural user interfaces (NUI)/touchless interfaces in hospitals between 2007–2020 (ACM Digital Library, IEEE Xplore, PubMed, ScienceDirect) [64].

Choosing the gesture lexicon

Recent work has further considered the development of a gesture lexicon, as an important part of developing a gesture-controlled touchless system. When approaching lexicon design, it is recommended to start from an intuitive set of initial gestures. Such a lexicon should aim to be extensive enough to support all required tasks, but not so extensive as to reduce a system's ability to discern between different gestures [55]. Gestures that use a single hand, rather than two-handed or body gestures have been found to be more acceptable with users, despite similar performance levels [62, 55]. Finger-based gestures have been found to be challenging to reliably track from a variety of angles and ranges [65]. An initial gesture lexicon may benefit from using one-handed gestures, especially those that are simple to perform, such as opening the palm, as these are uncomplicated and practical to perform in the OR [66].

It has been found that, during gesture elicitation, users show very low levels of agreement between gestures. This is especially true when a command assigned to a gesture is abstract, e.g., mute, and when the set of commands is very large (>20) [67]. Furthermore, this approach of selecting popular gestures can result in lost context due to the lack of agreement between users [67]. This reduces the value of such approaches when developing gesture lexicons. Despite this lack of agreement, users demonstrate a willingness to accept and learn limited gesture lexicons [55]; this may be due to an increased level of acceptance of advances in technology. Leveraging existing gesture languages that users are likely to be familiar with, may provide an approach to lexicon development, potentially reducing learning difficulty for the user. An example of using social context to inform system design can be seen in Nishihori *et al.*. Their solution leveraged shared knowledge by using *janken*, the Japanese equivalent of "Rock Paper Scissors", as the basis of their gesture lexicon [68]. In Japan, *janken* is ubiquitous, holding importance unique to that societal context; this immediately reduces the learning curve, with users having a high level of familiarity with the gestures required to use the touchless system. The provision of an intraprocedural reference guide for gesture control may reduce the impact of learning the gesture lexicon, which may be desirable among users [69].

The impact of environmental sounds on voice control

Interest in the application of voice control has also continued in the literature. In a hospital environment, voice control input recognition rates may be impacted negatively by environmental noise. Furthermore, systems tested in lab conditions have demonstrated superior recognition rates to those tested in real-world environments; this can be seen when comparing the work of Nathan *et al.* and Hötker *et al.* [70, 71] with those of Alapetite and Perrakis *et al.* [46, 50]. It is important, therefore, to test touchless systems that use voice control in real-world environments, as understanding the impact of authentic environmental noise on a system allows the implementation of appropriate countermeasures [68].

Other technologies

As in the original review of the literature, EGT was discussed, however, compared to the original review of the literature, details regarding EGT in terms of use and calibration requirements were more closely explored.

EGTs can only provide a coarse indication of intent, meaning they are most suitable in applications where fine levels of accuracy are not required. Examples of such applications include selecting a viewport or switching application focus across several monitors [60, 1]. Systems that make use of EGT need to support calibration; over time EGTs can suffer from significant statistical degradation of accuracy, which can lead to performance issues; a calibration step should be performed every twenty-four days to maintain acceptable levels of performance [72].

Frau *et al.* propose combining AR with facial recognition to automatically load patient information, supporting hands-free interaction [73], though this approach may encounter resistance due to data protection concerns.

2.3.4 Feedback

Providing appropriate feedback to the user in a timely manner can allow HPs more easily master touchless control of medical images [2]. This feedback can come in several forms, from on-screen textual notifications [60, 65], to real-time visualisation of the user's hands (such as in Figure 4.8, p.91). Unfortunately, though such feedback, visual or auditory, is beneficial in reducing user confusion, it is not often provided [1].

Incorporating feedback elements into a novel touchless system in a clinical setting can enhance user acceptance when using a touchless system [74]. This can be achieved by providing feedback that informs the user regarding the system's interpretation of their intention, such as notifications of successful commands. However, it can also be achieved by informing the user of the benefits of touchless control in terms of benefits when managing the risk of cross-contamination as compared to touch-based systems [55]; this is increasingly true in a post-COVID-19 context.

2.3.5 Clutching

For those papers which involved the design/evaluation of a touchless system, clutching and its role in preventing unintentional input continues to be seen as an important issue, echoing the results presented in subsection 2.2.3 (p.22).

When using touchless interaction, involuntary input has the potential to "lead to decisions based on wrong information or demand extra time to revert the system to the desired state" [75]. Unintentional input leading to involuntary manipulation of supporting information and imaging should be prevented; this requires the inclusion of a clutching mechanism [75]. The inclusion of clutching mechanisms, such as in Zaman *et al.*'s smart shoe, has also been found to enhance the user experience with touchless systems.

2.3.6 Non-functional Requirements

Three papers compared touchless interaction with mouse and keyboard control. For touchless systems to be successfully adopted, their performance should be comparable to traditional systems to avoid user frustration. Such comparable performance has been reported, with Nishihori *et al.* and Dmytriw *et al.*, reporting similar task times for the same operation when comparing touchless control and mouse and keyboard [59, 68], and Oshiro *et al.* reporting shorter task times when using touchless control, due to no longer needing to replace surgical gloves [66].

2.4 Discussion

One limiting feature identified across papers containing experimental results was the persistent use of a small sample size during user testing of the systems, and no system was tested in more than one hospital. Ten of the studies were conducted using non-medical personnel. Those papers that performed experimental work in hospital environments may have been constrained by access to hospital staff, whose time may be difficult to obtain. Whilst these sample sizes are sufficient for early-stage prototyping and investigating feasibility and acceptability, they are too limited to determine if the solutions proposed are appropriate for wide-scale deployment. No studies reported outcomes relating to effects on contamination. Ultimately, outcome-focused evaluations showing reduced levels of bio-contamination will be required to promote widespread adoption of these technologies.

2.4.1 Standardising Evaluations

Effective and systematic protocols for evaluation of touchless technologies should be established for a range of medical environments. In particular, producing a set of standardised tasks for a particular context and user cohort would support experiments comparing time-on-task measures for different designs and interaction modalities, and also studies of outcomes based on cross-contamination (e.g., comparing a touchless to a touch-based design). Critically, it would also allow for meta-analysis across different studies. Further standardisation of a gesture set (or set of voice commands for speech) to support a particular task would allow comparison of different implementations with regard to recognition speed and accuracy, again opening up the possibility of meta-analysis. Reliable comparisons of different hardware technologies would require experimental control of the tasks, gesture set, and user cohort, and thus would ideally be conducted as part of the same experiment.

2.4.2 Beyond the Operating Room

While several OR and interventional radiology use cases have been explored, there has been no systematic consideration of the use of touchless systems in other contexts within a hospital. This is an opportunity for future work as pathogens can be spread anywhere in a hospital. However, with this opportunity comes additional considerations and restrictions. For example, in the context of voice control, audio feedback to the user is sometimes useful. However, this feedback may not be possible within some parts of the hospital environment, or at certain times (such as during the night) as it may be important to minimise ambient noise levels.

2.4.3 Contextual Cues and Clutching

A number of contextual cues were also considered in the papers, including the use of gaze as a cue. Several papers examined eye-gaze technology for touchless control of systems. While gaze-based control is challenging to implement, determining if a user's gaze is directed at the system is by comparison, a relatively simple task. This provides the potential for unencumbered gesture use of a system where gaze is used as a contextual cue to process or ignore gestures (clutching). Similarly, voice control can be accurate and well suited for discrete commands for clutching. Strickland *et al.* identified system activation in a crowded operative space as being the primary source of issues during use of their system [23]. Ultimately, multi-modal systems combining speech and gesture, gaze and gesture, or gaze and speech, while technically challenging, may have the best chance of success.

2.4.4 Implementation

Regarding implementing gesture recognition in systems, we see in the literature a gradual move from implementing conventional interaction methods using gestures, to designing systems with gestures in mind from the ground up. Advancements in technology provide emerging opportunities in terms of potential accuracy of both depth and colour inputs. Developments such as the second generation of Microsoft Kinect, which was not used in any of the papers investigated, is a sufficiently significant improvement when compared to the first generation of such technologies that one can reasonably anticipate a noticeable improvement in performance in possible applications.

2.4.5 Best Practice

As touchless control advances towards being a viable option for use in the hospital environment, it is appropriate to consider best practice in the design and evaluation of these systems. While different in scope and focus, several of the findings presented in this research are supported by the recent independently conducted review of touchless interaction in the OR and interventional radiology by Mewes *et al.*[1]. Major themes in their analysis echo the conclusions presented in this research regarding recent improvements in the feasibility of touchless control, the need for improved evaluations, the need to improve usability, including issues surrounding accuracy and unintended gestures, and the potential of multimodal interaction to address some of the practical difficulties in making these systems appropriate for deployment. With regard to best practice, these findings support careful consideration of usability in the design of touchless systems, using multimodal input to support clutching, using realistic tasks, and conducting larger studies with representative healthcare workers.

2.5 Summary

It is clear that real-world touchless medical image systems are now feasible due to ongoing progress in areas such as hardware capability, and understanding of the problem space.

Although there have been an increasing number of studies performed with clinical participants, deeper investigation of HP requirements would be beneficial in order to bring further understanding to the problem space. Chapter 3 (p.45) addresses this topic through qualitative analysis of the needs of HPs in the context of existing PACS as well as their needs and attitudes to touchless medical image systems.

There remain multiple open issues concerning the design of such systems, such as developing appropriate touchless lexicons, and understanding appropriate clutching mechanisms. These will be investigated in Chapter 4 (p.75) and Chapter 5 (p.115) respectively, looking at the process of developing touchless medical imaging systems (TMIS), focusing on issues identified in this review of literature, as well as the HP experience with TMIS.

By bringing together these investigations of HP needs and experiences, insights into designing TMIS, and HP experiences using a prototype TMIS, the issues identified in this chapter can be studied in relation to one another, and be brought together and evaluated as a whole.

Chapter 3

A Qualitative Analysis of the Needs and Experiences of Hospital-based Clinicians when Accessing Medical Imaging

The literature review presented in Chapter 2 (p.11) reveals the need to develop a better understanding of the needs and experiences of hospital-based HPs when accessing medical imaging. Having a better understanding of the context of use is vital for identifying user requirements for the design of real-world systems for touchless control of medical imaging. In order to bring insight to this area, a qualitative study was conducted with a cohort of practicing HPs regarding their experiences with existing PACS, as well as probing their attitudes to touchless control of PACS; the results of this study are presented in this chapter.

3.1 Introduction

3.1.1 Motivation

PACS has become a key element of the digital hospital workflow, alongside radiology information systems (RIS), dictation systems, preliminary report systems, and EHR [76]. Technology improvements that support clinical excellence and service to the community are more likely to be accepted by clinicians [77], and high levels of satisfaction with PACS were reported following it's introduction [78].

Increasingly, hospital PACS have been integrated with a variety of EHR systems [79], including national systems such as the National Image Management Information System (NIMIS) in Ireland, which supports distributed access to medical imaging. While these integrated solutions bring many benefits, including significantly reduced time for radiologists to access clinical data in the EMR [79], such integration requires adherence to industry standards in order to allow various systems to exchange data effectively with each other. Despite the evolution of these standards, there remain considerable challenges in the use of these systems. Furthermore, trade-offs are often made between the potential to collaborate and share images vs. the response time and image resolution.

Now that digital imaging has become embedded in clinical practice, it is timely to examine clinician experiences with PACS, in order to inform future development of these systems. Recent technological developments have opened possibilities for a wider range of user interaction modalities for the operation of computer systems. Voice and gesture-based interactions, in particular, have been proposed as having applications in interacting with digital imaging, particularly during surgery and interventional radiology [14]. HP attitudes towards touchless interaction with digital imaging are also explored in this work.

3.1.2 Research Questions

The following research questions are addressed in this chapter:

- **RQ1:** How do health professionals (HP) use current PACS?
- **RQ2:** What challenges do HPs encounter when using PACS?
- **RQ3:** What are HP attitudes to touchless interaction with PACS?

3.2 Background

A brief overview of previous relevant work on the adoption and usage of PACS is presented in this section.

The introduction of PACS to hospital environments has had a significant, positive impact on clinician workflows. According to Lederman *et al*, PACS has workflow influences beyond automation, having impact also on radiologist, technologist, and clerical staff productivity, report turnaround time, and on communication between clinician and radiologist [77].

Technology improvements that bring clinical excellence and service to the community (which Lederman *et al* flag as key factors for public hospitals [77]), PACS in particular, are more likely to be met with acceptance by clinicians. Bryan *et al.* reported that PACS users exhibited a very high level of satisfaction with the software [78]. According to Reiner *et al.*, PACS resulted in an approximate twenty-minute reduction in time required for intraoperative angiography, as well as leading to a perceived reduction in stress amongst surgical staff (both surgeons and nurses) [80]. PACS has enabled clinicians to 'enhance' images after they have been captured by introducing tools such as image windowing and zooming to the clinical workflow [80]. Surgeons in particular, place great value on the improvements to time management and efficiency brought about by the introduction of PACS [80]. The increasing availability of digital imaging via distributive technologies has lead to an overall reduction in the frequency of consultations between radiologists and clinicians [80], with the radiologist's report becoming an even more vital part of the clinical workflow. As the technology has matured, PACS has become increasingly commonplace in hospitals, with many younger hospital staff members never having experienced film-based workflows. The maturity of PACS, the impact of PACS on the clinical workflow (notably in surgery and radiography), as well as the increasing availability of novel interaction technologies is a key motivation for this work.

Traditional film-based workflows have a number of drawbacks. One of the most significant issues with film is the tracking and management of the location of physical patient film records [81]. File clerks describe searching the hospital filing system for film records, which would eventually be reported upon, often a week or more after the actual imaging procedure occurred [81]. It is reported that up to 25% of all patient examinations were never formally reported on due to films being “spirited away” by residents and medical faculty before they had been read and reported on [81]. According to Hayt *et al.*, the issue of physical film records being lost had such a pervasive impact that the introduction of PACS and the immediate access it allowed to any needed examination generated “a positive attitude and determination to work with the system by all of the staff” [81].

There are significant benefits to adopting a digital workflow incorporating PACS as compared to an analogue film workflow. In the opinion of Siegel and Reiner, the most significant of these benefits has been the ability to re-engineer the overall workflow, allowing for the reduction of many inefficiencies and errors that existed in diagnostic imaging departments [82]. For example, prior to PACS the process of requesting, obtaining, reporting, and transcribing a conventional chest radiograph required 59 steps, compared to the PACS-based workflow that reduced the process to 9 steps [82]. Srinivasan *et al.* reported that the introduction of a digital radiology system improved clinician satisfaction and workflow increased the rate of clinician image review [83].

The combination of speech recognition systems and PACS has proven to be a powerful pairing. Trumm *et al.* reported that, in the context of PACS/RIS, speech recognition systems enable radiologists to produce written reports without the assistance of a transcriptionist, potentially improving hospital-wide report availability [84]. Hayt *et al.* reported that the introduction of a digital voice recognition system had a significant positive impact on report turnaround times, resulting in 86% (as compared with 3% prior to PACS) of examinations being reported on within 12 hours of being performed in Elmhurst Hospital in New York [81, 85].

Hayt *et al.* reported that the introduction of PACS resulted in a significant drop in the number of unreported examinations, from 25% to less than 1% [85].

Srinivasan *et al.* report significant financial benefits attributed to using digital imaging systems as compared to analogue film. In the case of UA Davis Health System, a reduction of 73.4% of film printing and a 50.3% reduction in file clerk time resulting in annual savings of $1,001,452, as well as freeing a significant volume of storage space (valued at $2,018,320). Chen *et al.* note that institutions operating without PACS are now at a competitive disadvantage [86]. This is especially true as PACS is now sufficiently affordable to be within the reach of smaller institutions and more cost-sensitive groups [86]. According to Chen *et al.*, what they refer to as the "commoditisation of PACS" has resulted in decreased cost, improved interoperability, easier replacement, increased customer choice, and improved vendors' response to their customers [86].

Since the introduction of PACS, clinicians have become accustomed to reading radiology reports [81], resulting in a significant reduction in the amount of radiologist/clinician consultation time [85, 83]. However, it has also been noted that radiologist/clinician dialogues, which were common prior to PACS, often resulted in a better understanding of the clinical problem, which could result in a more meaningful final report [85].

Another potential issue with the introduction of digital imaging workflows is the shift in expectations regarding the radiology department's deliverables, with the reduced report turnaround times leading clinicians to expect more real-time access to these reports [85]. "When the urgency of a study is not appropriately managed, patient care decisions may be unnecessarily delayed, with possible adverse outcomes" [76].

Improvements in technologies such as PACS have meant that the evolution of digital workflows is inevitable [76]. The ability to decouple digital workflows facilitates improvements to workflow efficiency and workforce utilisation. Previously, the workflows of the technologists and radiologists were tightly linked, which helped ensure the proper prioritization of case reviews. However, it also makes it more difficult for radiologists to 'see' the complete state of the imaging for particular patients, obscuring looming problems until they become crises [76].

There are a number of factors that should be considered when implementing PACS. Hospital-specific resources, capabilities and the use of PACS are strongly interrelated. The involvement of multidisciplinary teams consisting of physicians, technicians, and engineers is a critical success factor when applying improvements to PACS [87]. The degree of PACS integration will impact the level and extent to which it is used, as well as how it is incorporated into clinical practice [88]. Proper integration of PACS into a specific hospital setting is essential for its success, and without this, the potential gains of PACS are unlikely to be realised [82].

3.3 Methods

In order to investigate the experiences of health professionals with PACS, a qualitative research study based on semi-structured interviews was used. Interviews were conducted with clinicians in a range of roles, between 27 February 2019 and 9 December 2020. Research ethics approval was obtained from the relevant institutional ethics review committee (Figure B.1, Figure B.2, Appendix C, Appendix D, Appendix E).

3.3.1 Sampling, Recruitment, and Consent

Participants were approached using a combination of personal and professional contacts, as well as from a hospital visit. The enrollment criteria for the study were that participants were healthcare professionals, and had the experience of using PACS as part of their normal duties. Participants representing a range of perspectives, and participants with a high level of PACS experience were sought through purposive sampling. The sample size was deemed sufficient once saturation was reached.

Potential participants were contacted directly and were provided with a short summary of the objectives of the study to gauge interest and relevance. Those who consented were given the choice between face-to-face interviews and over-the-phone interviews. Snowball sampling was used to recruit further participants. All individuals that expressed an interest in taking part in the study were provided with a time, date, and (in the case of face-to-interviews) a location for their interview. They were also provided with a participant information sheet and a consent form. No form of incentive was provided and none of the contacted individuals declined to take part.

3.3.2 Participants

35 participants from five countries were interviewed (Ireland, UK, UAE, USA, and Australia) with interviewees being predominantly from Ireland. The sample included clinicians of various levels of seniority, and from a range of roles. All interviewees were hospital-based healthcare professionals. Role, speciality, and experience for the participants are summarized in Table 3.2 p.55).

Role	**Number of participants**	**Speciality**	**Years of experience with PACS**
Intern	3	3 General	3 = 0-4
Senior House Officer/Resident	11	6 General 1 Orthopaedics 1 Paediatrics 1 Obstetrics & Gynaecology 1 Anaesthetics 1 Surgical	10 = 0-4 1 = 5-9
Specialist Registrar/Fellow	6	2 Colorectal 1 Orthopaedics 1 Neurosurgery 1 Pain 1 Unspecified	1 = 5-9 5 = 10-14
Consultant/Attending	6	3 Radiology 2 Surgical 1 Rheumatology	4 = 5-9 1 = 10-14 1 = 20+
Clinical Specialist	2	1 M.R.I. 2 C.T.	2 = 5-9
Radiographer	6	4 Radiography 2 Ultrasound 1 C.T. 1 M.R.I.	4 = 5-9 2 = 10-14
Radiation Safety Officer for Clinical Specialists	1	1 General	1 = 5-9

Table 3.1: Participant characteristics

3.3.3 Interview Design and Content

A review of the existing literature provided the groundwork for the semi-structured interviews. The interviews aim to explore the role of PACS in the workflow of each clinician, investigating their requirements for PACS, the perceived importance of sterility when interacting with PACS, as well as their attitudes toward touchless interaction with PACS.

Interviews were performed both in–person (either in the hospital context or in private settings) and remotely (due to participant location, which could be outside of Ireland, and, ultimately, restrictions caused by the COVID-19 pandemic).

The study presented here forms part of a wider project that explores new technologies for user interaction with PACS, and aims to inform the future development of touchless clinical systems.

3.3.4 Analysis

Interviews were audio-recorded and transcribed verbatim, with each participant assigned a unique ID code for anonymity. Transcriptions were analysed thematically. Coding was performed using the Framework Method described by Gale *et al.* [89]. The Framework Method was chosen as a suitable tool for the analysis of the interview data due to the multi-disciplinary nature of the research, and is particularly suitable for the analysis of semi-structured interviews. An experienced qualitative healthcare researcher provided guidance and feedback for the analysis. A CAQDAS (Computer Assisted Qualitative Data Analysis Software) was used to facilitate the qualitative data analysis. This tool was chosen due to its established position as a qualitative data analysis tool, providing robust tools for viewing data, applying codes and themes, and generating reports to bring insights to the data; further, being a cloud hosted service, the CAQDAS enabled collaborative analysis of data, codes, and themes.

Sitting within the family of analysis methods often referred to as thematic analysis qualitative content analysis, the Framework Method provides a structured approach similar to that Braun and Clarke have proposed [90, 91, 92, 93]. As with other thematic analysis approaches, the approach proposed by Gale *et al.* seeks to capture patterns in data. A benefit of such TA-based approaches is that it allows for theoretical flexibility, as it lacks an inbuilt guiding theory [93].

The method of constant comparison was used, and following coding and initial analysis, key themes were identified and developed from the transcripts (Figure 3.1 p.53). Discussion sessions were then conducted to refine themes and sub-themes, and the transcripts were revisited until agreement was reached.

Figure 3.1: Procedure for analysis of qualitative data

- **1. Transcription** A verbatim transcription is needed [94]. A combination of cloud–based A.I. audio transcription software and manual review was used to convert interview recordings to time–coded scripts.

- **2. Familiarisation with the interview** Becoming familiar with the whole interview is a vital stage in interpretation [94]. This familiarity was achieved through reviewing the interview transcriptions alongside the original audio.

- **3. Coding** The researcher reads the transcript line by line and applies a label, i.e., a 'code', that describes what they have interpreted as important. For example, in the line “In the light of COVID it would be more important, particularly in a hospital that has seen a lot of COVID, and the terminals are almost never wiped down.”, the code 'sterility and COVID' could be applied as the participant is discussing how COVID-19 has made sterility even more important to their workflow.

- **4. Developing an analytical framework** Following the review of several transcripts, researchers decide on a set of codes for all subsequent transcripts. Codes can be grouped together into categories, forming a working analytical framework. The analytical framework is never 'final' until the last transcript has been coded [94]. In this research, the category 'Problems' contains multiple codes, including 'Understanding the software', 'Data security', and 'Lack of functionality'.

- **5. Applying the framework** The working analytical framework is applied to subsequent transcripts.

- **6. Charting the data into the framework matrix** Charting the data involves summarizing the data by category from each transcript. The CAQDAS provided reporting functionality that allowed the data be exported to spreadsheets.

- **7. Interpreting the data** Through discussion with other members of the research team, characteristics of and differences between the data are identified. If the data are rich enough, the findings can explain reasons for emergence of phenomena, predict how an organisation is likely to respond to a situation, and identify areas that are not functioning well [94].

3.4 Results

The findings for each theme are presented under the following headings: adoption and evolution of PACS, locations and roles, tasks and features, workflow, performance issues, training, and touchless interaction and sterility.

3.4.1 Adoption and Evolution of PACS

HPs report that the introduction of PACS had a dramatic impact on the clinicians' working day, bringing a newfound convenience to the clinical workflow. However, despite the fundamental nature of PACS in the clinical workflow, multiple HPs commented that they had observed no appreciable improvements in system functionality over time. If anything, it is noted that PACS has become less usable over time for the typical user.

SREG4: "What stands out the most is that in 12 years, there's been very little change."

Though there are multiple providers of PACS solutions, HPs reported finding little difference between the various offerings.

Consultant Rheumatologist (CR1): "They all appear the same to me, to be honest."

Impact on work practices Participants are positive about the beneficial impact of PACS on their work practices. PC. 1: "I could easily see an X-ray and give my expertise based on an X-ray. That could only be done through PACS. It's convenient. It allows you to outsource the expertise."

Analytic functionality integrated with PACS is also mentioned, and these capabilities are likely to increase in the future [95]. PC 1: "I use engines of artificial intelligence to actually get a better diagnosis . . . you can make it more precise, faster, cheaper, more efficient, etc."

Whilst many HPs use PACS, the extent of use can vary greatly between different users. The majority of users report that they would use PACS to review basic imagery and review reports prepared by radiologists (the radiologist's report is considered the definitive interpretation of a scan).

3.4.2 Locations and Roles

HPs report three primary locations where they would use PACS: dedicated hospital workstations specific to the user (e.g., a radiologist's office), at a shared hospital workstation/desktop (e.g., in the nurses' office, on the ward, or in the clinic), and in the operating theatre. Consultant radiologists also report having access to a PACS system from home to facilitate on-call work during the night. Previously, if a clinician was on-call, they could be required to travel to the hospital in the middle of the night, so the ability to read scans remotely is seen as a significant improvement. CR 1: "Generally consultants who work in hospitals where there are no radiology trainees will have PACS at home for on-call, so if we're called in the middle of the night to read a scan we can read it at home."

Users report that gaining access to computers to use PACS can be a challenge on shared hardware as there is not enough hardware for the number of users. One user states that they use PACS (SHO 8) "Wherever I can get a computer." In terms of computing power, many HPs report a lack of dedicated PACS hardware. Often, they access PACS software on general-purpose desktops that are frequently described as slow and not powerful enough for their needs.

# of uses per day	Number of participants
<1	3 (10%)
1–5	6 (19%)
6+	22 (71%)

Table 3.2: Frequency of PACS usage by participant

3.4.3 Tasks and Features

Imaging

HPs report basic navigation and management tasks as being common activities, such as searching for and opening scans; deleting images if they are not wanted was also referred to, though much less frequently. Deletion of unwanted or flawed imaging is reported as being performed in the radiology department, generally at the time the scan is being performed.

There is a strong emphasis on efficiency; HPs refer to a ‘red dot system’, where images in a sequence are marked with a red dot by the radiologist to identify images of significance, allowing faster identification of key images when reviewing the report. Use of dynamic features to play a sequence of images is also reported.

The lack of online access to scans performed in some hospitals is a clear source of frustration for certain HPs; when a scan is required in that scenario the imaging and reporting need to be saved to optical disc, sent to the relevant hospital, and then uploaded into the local PACS. As patients often have different identifying numbers at each hospital they visit, difficulties are encountered when trying to determine if a patient has had previous scans due to the lack of uniqueness of their identifying number.

One radiographer notes that sometimes image uploading can go wrong, with images ending up in incorrect storage locations. Radiographer 5: “Sometimes there’s multiple names or slight changes. Like, we’ll say, saved into the wrong folders”.

Despite the high level of functionality that PACS offers, it is reported that “for the most part [it’s] basic functions, basic functions”(PC 1). Many HPs report using very limited subsets of PACS features. HPs from the radiology department report using a wider range of PACS features.

It became apparent from the interview reviews that various feature sets are shared by HPs in the same role. Further, it is clear that the feature set a clinician would use can change if they alter role, e.g., one HP reported previously using inverting a lot in a different role, but not using the feature at all now in their current role.

Use of the measurement feature is an example of where HPs fell into one of two use categories, either reporting having no use for the feature, or having significant use for it. Not all measurement occurs on PACS itself, with some being performed directly on the scanner before the imaging is added to PACS. The need to perform a measurement can be situational, depending on the information a scan contains.

Use of the PACS suite in multidisciplinary team meetings Several HPs referred to multidisciplinary meetings (MDTs) and reported that a more extensive range of tools is employed during those meetings. It is during the multidisciplinary meetings that various specialities come together to plan treatment for a patient [96]. Medical imaging is a key tool in the MDT that allows clinicians to leverage their combined expertise, resulting in more informed diagnosis.

SHO 1: "So obviously PACS is a pivotal part of that meeting. It's loaded up on a big screen and everybody is looking at it, things are being measured, we're looking for lymph nodes, things like that. But it's the radiologist that's operating it, we're all just viewing."

PACS as a tool to enhance imaging during post-processing HPs report that they need to adjust the contrast of scans when the scan exposure is less than ideal, with a number of reasons being cited.

HPs note that human error, differences between the image displayed on the scanner versus the image that is saved to PACS, or even the patient's physical size can impact the exposure of a scan being not quite right. In these scenarios adjusting the contrast allows the HP to better interpret the scan. Some HPs note that they perform this contrast adjustment directly on the scanner before sending the image to PACS.

Zooming an image is relatively commonly reported as being useful, especially for investigating subtle features such as certain fractures. Usage of mono-inversion (flipping black to white, and white to black) is heavily task-related, e.g., chest x-rays are frequently inverted, while other forms of imaging are not. Inverting is described as very helpful for optimising image contrast.

Labelling and annotating

One HP notes that annotating can be used to record information that would otherwise be lost, such as the state of the patient during the scan, e.g., if the patient was in a state of expiration (breathed out). Such annotations allow HPs to understand and make allowances for potential shortcomings in the imaging.

Radiation Safety Officer: "...might not be evident on the x-ray. You just write the patient was expiration." However, labelling and annotating images are not commonly reported as being used.

3.4.4 Workflow

One HP reports that sometimes the emergency department would order scans but then not collect the scan. Every scan must be viewed by the person who ordered it, and must be recorded as having been viewed. If a scan is not viewed by the ordering clinician, additional workload is created for other members of staff in resolving this issue. This can be due to issues arising from the structure of the organisation or by errors entering the ordering clinician. SHO 9: “Sometimes the primary consultant is incorrect. So, scans that are ordered get attributed to the wrong department.”

Of note, is the reported simplicity of the average clinician’s PACS workflow. Though PACS supports a wide range of functions, the vast majority of interactions will only use a small subset of those functions. This large feature set is reported as having led to difficulties using PACS, with HPs reporting a desire for simplification of the user interface.

SREG4: “We’re just trying to log in, find the patient, look at a report, look at an image. Having two dozen different options is not really useful for most of us.”

PACS is often used in busy environments

HPs report that they frequently use PACS in busy environments with high levels of noise and staff movement. Further, it is reported that multiple people would be performing the same role on a given day, sharing the PACS human-machine interface (HMI) between multiple HPs (with each HP having their own login). HPs note that rather than there being a lack of space, there is an abundance of people.

PACS as a part of the operating theatre workflow

In the context of the operating theatre, it is noted that there is a circulating scrub nurse who can be directed to use PACS. However, it is also noted that this is very inefficient and inconvenient as often the user being instructed may not be familiar with PACS and may not fully grasp the intent of the surgeon.

SHO 1: “It’s such a pain to try and direct a circulating scrub nurse on how to get to the exact point, scroll to the exact slice that you want.”

The impact of clinical governance and data protection on HPs' use of PACS

There is a need to maintain a balance between workflow efficiency and clinical data governance. In order to protect patient information, PACS systems log the user session out after a period of inactivity. Some HPs report this period as being as short as 5 minutes. All PACS logout automatically, including PACS in the operating theatre. HPs report that in the operating theatre the mouse needs to be 'jiggled' periodically by an unscrubbed staff member to keep the session active. If the session becomes inactive it will log out, and then the correct login details need to be entered and the session logged back in. In the context of the operating theatre, it is noted that this involves an additional staff member knowing the clinicians log in details. After logging back in, the correct image needs to be located and displayed correctly. HPs describe this process as being a significant frustration.

SHO 8: "That's really frustrating because we'll have the imaging up, everything logged in and on the image that we want, and then as soon as the computer goes to sleep it logs out."

Some HPs report that the PACS they have access to does not allow multiple simultaneous dialogue windows. Similarly, it is reported that switching patients, and between RIS and PACS for the same patient would completely close one application in order to open the other rather than switching between them. This is deemed to be very frustrating for the HPs. The lack of system integration can have a significant impact on user efficiency as, in order to move between different functionalities, they must sacrifice their existing progress through the interface.

CS 1: "...you have their information up in the RIS and you're like 'oh God, did they have a scan?', you have to go to click on the imaging part, and it closes that down completely. It doesn't even minimise the window, it actually just closes it."

PACS in the WHO surgical safety checklist

In the context of the operating theatre, there is a specific WHO checklist that aims to decrease errors and adverse events [97]. It defines the workflow for multiple stages of a surgery, including ensuring that essential imaging is displayed, and thus PACS plays an important role in adhering to the checklist.

3.4.5 Performance Issues and Their Impact on User Attitudes

While not universal, many HPs report various performance issues when using PACS. The primary complaint is with respect to speed, with many HPs reporting that PACS would run very slowly, with an adverse impact on their workflow, both from the delay it causes, and thereby creating a reluctance to consult the imaging.

SREG4: "Everyone should be looking at the images . . . but we are discouraged from doing that because the system is so incredibly slow."

3.4.6 Perceived Shortcomings in User Training for PACS

A significant number of HPs report having never received any formal training in PACS, with most learning through observing their colleagues use the system. Many HPs report a lack of understanding of PACS, saying that there are features of the software that they do not know how to use and therefore simply ignore.

Intern 3: "Pretty much all of my PACS knowledge is self-taught, or like is taught by someone else on the go. I feel like I could function more quickly and efficiently if I had had more formal training."

One HP reports that though PACS training is available, most people do not take advantage of the training. It was noted that many HPs are unaware of any available training.

SREG 2: "I think people don't possibly take those opportunities up because they're like 'oh, I've used that before' . . . So I think that perhaps it's not that there isn't access to opportunity in training in PACS but more so that people don't avail of them properly."

When asked if they feel that training would be beneficial, most HPs respond in the affirmative, though a small minority say they do not think that training would be beneficial, or would only be beneficial for more junior staff. Some HPs suggest that a basic level of training should be provided when first being introduced to a hospital's PACS, with further, more advanced, training being provided some time after (between a month and a year). Some HPs suggest that short refresher courses would also be beneficial.

SHO3: "...you don't really understand what you're going to be doing, and then it would probably be good to train you in your first week and then maybe after a month to touch base with people again."

It is reported that the scope of PACS software functionality can make it difficult to learn in a single training session. HPs expressed a strong appetite for ongoing micro-learning.

CR2: "But you know you are bombarded at the beginning, it can do this this this and you're only going to retain a very small amount of it. To be honest what you're trying to do is the very basics, how do I call up the image and how do I report. But there are lots of other things you can do that most of us never use because you know you might be told on the first day but that's no good. You really need to kind of to be retrained on it."

It is suggested that PACS training could be incorporated as a component of other forms of training.

SHO3: "maybe as part of the surgical training or the medical training."

It is also suggested that PACS should be sufficiently intuitive as to not require face-to-face training. Instead, the majority of functions should be intuitive to use, with more advanced features available in additional views.

SREG4: "It shouldn't really be required. So, a PACS system is so fundamental that training should either be able to be almost non-existent, then it's so straightforward as you're logging in for the first time, that your training is like a tutorial when you open up a new app on your phone."

The impact of poor user interface design on user understanding of the software

Many HPs interviewed express a feeling that they lacked a proper understanding of PACS, opining that with a greater understanding they would be more efficient in using PACS, and would be able to leverage a greater number of the available tools to achieve more. This lack of understanding is attributed to a combination of an unfriendly user interface and a lack of training. HPs reported that they would be more likely to use PACS if they felt more familiar with it.

Multiple HPs report that PACS is unclear and unintuitive to use, suffering from complex UI and poor UX. Even basic and essential features such as comparing images are described as being difficult to use, with this difficulty being attributed to a lack of understanding of the software. HPs report adopting manual solutions that lack the efficiency of built-in tools to get around their difficulties in using PACS.

3.4.7 Touchless Interaction and Sterility

The perceived lack of importance of sterility amongst PACS users

Most HPs indicate that sterility is not an important concern to them when using PACS. This is because they would have an opportunity to wash their hands between using PACS and interacting with a patient. Those users with dedicated workstations (such as radiologists) report that they are not concerned with cross-contamination. CR 2: “No, because essentially it’s my own grime. . . I don’t tend to use anyone else’s so I’m not worried about that at all.”

SREG 6: “In the light of COVID it would be more important, particularly in a hospital that has seen a lot of COVID, and the terminals are almost never wiped down.”

HPs who use PACS in the OR, however, report that sterility is especially important to them in that context.

User attitudes to the potential of touchless interaction with PACS

Most HPs react positively to the introduction of touchless interaction to PACS. In particular, HPs are easily able to envision using voice-controlled shortcuts to speed up their workflow. HPs note that many tasks in PACS take a large number of mouse clicks to achieve, with each click potentially taking a significant time to respond, and minimising clicks is seen as an important benefit.

HPs also express positivity toward the concept of using a touchless system during aseptic procedures, or at times when the HPs hands might be busy. While touchless interaction might not be ideal for all environments in a hospital, there are a number of situations where access to touchless interaction of PACS is appealing to HPs.

SHO 2: “I suppose if the technology was there it would be fine. You’d just need to make sure the theatres would just have to be set up properly so that there’s actually room for someone to move and make gestures, cause often theatres can be so tightly packed surrounding the computer that it would be hard to maintain sterility and perform gestures.”

Social context is also seen as affecting the acceptability of voice and gesture-based interfaces. PC 1: “In the operating room we also work with head gestures like nodding and things like that. It was just, you look almost stupid. So people would make fun of you.”

PACS and touchless interaction in the operating theatre

HPs report that PACS use in the operating theatre is not a given, but rather is situation dependent. HPs note that when PACS is used in theatre it is very important. The operation is planned around one or more key images, which are displayed from the start until the end of the operation. HPs in this study cohort reported that any interaction with PACS during a procedure is generally unplanned, with all imaging having been reviewed in advance of the operation.

SREG 1: “Typically you’d have a screen up in theatre and you’d have one key image up that you’re interested in. . . . Then you can plan around it, just to refer back to just so you can know if your screws are in the correct orientation or something like that.”

HPs respond favourably to the idea of touchless interaction with PACS in the operating theatre. SHO 3: “It would be really nice to be able to control the images yourself while you’re scrubbed in theatre wearing your gloves and your mask and not having to ask a nurse, a scrub nurse, ’scroll up, scroll down, go back, go to this’.”

HPs report that they would be happy to adopt a robust touchless system, especially in a scrubbed environment, noting that voice dictation has already been successfully adopted in their hospitals.

The preference of HPs for voice control

Many HPs express an interest in being able to use voice control to interact with PACS. Most of them can visualise leveraging voice commands to accelerate their workflow. Voice dictation is already commonplace in hospitals, so the use of voice as an interaction method is already established. Conversely, gesture interaction is more difficult for people to visualise being used in a clinical setting. Gesture-based interactions are not currently used in hospitals, resulting in greater learning effort for gesture as compared to voice control due to unfamiliarity with this mode.

SHO 11: "I don't think there would be a downside... for me if you pull up simple commands, for example when comparing chest x-rays, if you have a command like 'show me the [unclear] x-ray of the chest', I mean that would be so useful rather than trying to trawl through all the investigations and find the last one."

HPs report that they would be happy with some specific training for voice control if it improves performance. CS 1: "If you are trained to use specific words, or if you have to put it in chart number this and then it recognises that you're searching for a patient."

Based on feedback from HPs, there exists a desire for effective operational shortcuts within PACS. Many HPs express a desire to issue commands such as 'display all images for patient X'. Aside from the convenience of being a verbal command, the greatest benefit of this instruction is that it reduces the complexity of performing the task, combining several actions into one.

Several HPs raised a potential issue regarding noise levels and the feasibility of voice control in the operating theatre. There can be many people talking at once, and in some scenarios, music playing too. They contend that a voice control system would need to be very robust to background noise to be of use in a real-world environment. SHO 8: "It's very noisy. And there's always beeps and machines going, and a lot of people talking. So it would need to account for that.... If we need to use it while we're scrubbed, it tends to be, you know, an unplanned use."

An additional issue that was identified for voice control is data protection and privacy. While this issue is only raised by PC 1, with regulations such as GDPR protecting data privacy within the EU, any touchless technology needs to be designed in such a way as to properly protect patient information.

PC 1: “Sometimes, for example, voice interface there can be a little privacy issues where you need to be aware of your surroundings. You may not want to give information that you could otherwise just type.”

Adoption of PACS technologies

Overall, most HPs within the cohort express a positive attitude to adopting a touchless PACS. One HP asserts that new technologies in the hospital are often adopted from the bottom up, with junior doctors trying the technology first and word of the technology trickling up through the ranks until either the technology is adopted or fails to gain traction.

Intern 2: “You’ll have people starting to use it in the hospital and buy in from a couple of people, then that gets around... Generally it’ll be more junior doctors and if they find that it’s more efficient that’ll catch on ... If you had a negative experience the first time you might be less inclined to try it again, but if you had feedback from other people that it was very easy to use that would change your practice.”

A number of HPs expressed a reservation towards the adoption of touchless PACS, reporting that if there is not a clear benefit to their day-to-day work, then they would be unlikely to adopt the technology.

Some HPs note that they would like to have the option of both touchless interaction and conventional interaction methods, at least initially. This would be the case for seated interaction with PACS, touchless interactions such as gestures are unlikely to be used due to the convenience of the mouse and keyboard. However, combining voice control with mouse and keyboard is suggested as being a beneficial combination in the context of workflow efficiency.

3.4.8 PACS - The Radiologist's Perspective

Radiologists are the largest single user group of PACS. Therefore, it is important to understand the perspective of radiologists when considering the design and use of PACS. The following sections report on findings in this study related to the use of PACS tools by radiologists.

Measurement

Radiologists report that the measurement tool with PACS is of use, especially when considering a tumour or feature that needs dimensional analysis in a scan. The need to perform such a measurement can be situational, depending on what a scan contains.

CR 2: "It depends on what you see. If you see a tumour you obviously have to measure that. If you see something that looks like it could be some blood you might take a measurement of its attenuation value to see if it is blood."

Comparing and reconstructing images bring powerful modes of interacting with imagery

Comparing a patient's current scan against previous imaging is reported as being a task of very high importance across multiple specialities and levels of user. Comparing up to sixteen simultaneous images facilitates monitoring for changes in a patient's condition which is deemed to be a critical part of patient care, in particular for patients with chronic conditions.

A number of HPs comment on the impact of their display window setup on their workflow: some prefer a single image be displayed as large as possible, and others prefer multiple images to be displayed together. HPs report that PACS itself has various task-specific views built-in, such as ortho-bone, or long windows.

Only consultant radiologists report reconstructing scans, making it a good example of a PACS feature that is of great value to a specific subset of users but is unused by other users.

The central role of radiology reports in the clinical workflow

HPs in Ireland report the implementation of a nationwide system, NIMIS that is used in conjunction with PACS; NIMIS has come to replace a number of the functions of previous PACS setups, most notably the reporting element.

Although the content of the raw imaging is considered important, HPs note that often the radiologist report is of greater importance, with some saying that they would only open the scans if they need to check something in the report. As the report is considered the authoritative interpretation of the scan, more junior HPs in particular, report a preference for relying on the report rather than trying to interpret the scan themselves.

HPs report that all scans are officially interpreted by radiologists, with results being available to view, generally by the next day. Many of the HPs report that, in the interest of speed, for more straightforward scans, such as chest x-rays, they would often make an initial assessment of the scan before the radiologist report is available. HPs note that often, time is of the essence, and there is an emphasis placed on treating as many patients as soon as possible.

Integration of PACS with other systems

While there has been progress toward greater levels of integration between PACS and other EHR systems, HPs report that there are still significant shortcomings in this area. HPs note that this reduces their ability to transfer skills between hospitals, often having to learn how to use unfamiliar software to perform familiar tasks.

SREG4: “Once you’re in the PACS, even you’re using an EHR, they tend not to be well integrated . . . Even in private hospitals, which have relatively well integrated electronic records, the PACS tends not to work with it.”

HPs note that they tend to move between applications based on the strengths of each system e.g., between PACS and NIMIS.

Performance issues and their impact on user attitudes

Slow image loading is highlighted as a significant hindrance to the radiology workflow.

CR 2: “If I try and look at a C.T. that was done in [the hospital] four years ago you literally try and retrieve it and come back hours later, see if it’s come up. So there’s no instant for anything old, there’s no instant. You can’t compare immediately, which is a big problem.”

Consultant radiologists report that accessing PACS remotely using the Homelink system is considerably slower than on-site access, with one HP speculating that on-site systems being prioritised by the central server. It is noted that performance using remote access suffered most during regular working hours (8:00 - 17:00 GMT) to the extent that one consultant radiologist says that during work hours you can just forget trying to access PACS remotely.

CR 1: “So the main one that comes to mind is the slowness of the system, particularly when pulling up old images. . . . I don’t have it crash too much, you know. It occasionally crashes when you ask it, but that’s usually when you ask it to do too much, you’ve too much going on in the background, say you would pull up 10 C.T.s and reconstruct them and it just says enough is enough.”

Perceived shortcomings in user training for PACS

Experienced radiologists report having received little training in the use of PACS. It is noted that they have developed their facility with the software over time through trial and error. One radiologist suggested that this approach might be superior to formal training, due to, what they describe as, the unfocused nature of training days where the training may be too broad, failing to focus closely on the real-world issues and requirements of users. A need for repeat training to refresh user skills is suggested by another radiologist.

CR 2: “When we started the system I think we got maybe half an hour. The problem with that is that at the beginning you get used to real basic stuff which is a puzzle to you. . . . But all the extra kind of things it can do, I’ve never used them because I’ve never had any training in them. . . . You really need to be retrained.”

Radiologists echoed the opinions of other HPs regarding a lack of usability in PACS interfaces. Radiologists use more advanced feature sets of PACS than most users; a reasonable conclusion can be made, that the issues with interface design of PACS effects users of all levels of expertise.

User attitudes to the potential of touchless interaction with PACS

Radiologists indicate positive attitudes to the possibilities of touchless interaction with PACS. In particular, the potential for improvements to workflow, such as reducing the number of clicks, is flagged as important. CR 3: "Even just as a radiologist, minimising the number of clicks, especially when the system is slow, if there are any steps that you can take that would negate having to use keyboards or presets would be excellent."

3.5 Discussion

3.5.1 Summary of Results

This qualitative study explores the experiences and preferences of HPs when interacting with digital imaging, it is clear that for many HPs, PACS is an important part of their workflow, and there are significant issues with existing PACS. For many HPs there is a clear sense that they feel that they lack knowledge of PACS features, and that additional training and improved usability would be of benefit. Usage of the various PACS features is generally aligned with a HP's current role, with many users using only essential functionality. HPs express overall positivity to the adoption of touchless interaction with PACS, with most user's expressing a preference for voice control.

3.5.2 Comparison with Existing Literature

In this section the results generated in this qualitative study are reviewed in the context of existing studies on the use of PACS.

Tasks and features The ability for multiple users to access images simultaneously once they are stored on PACS is highlighted in our interviews and is reported as a benefit by van de Wetering *et al.* [98].

Similar to the accounts in the study interviews, Fridell *et al.* note the superior ability to display 3-dimensional reconstruction as an important feature enabled by the advent of new technology [99].

The HP cohort in this study favour accessing the radiology report alongside the image, but say that it is not always easy to do in current systems. So too, Top's results show that the majority of users consider the ability to place radiology reports alongside imaging to be useful [100]. Top observes that reporting times decreased by 25% after the introduction of PACS [100]. Although this study did not formally investigate this particular question, the cohort HPs at interview report greater efficiency following the implementation of PACS.

The importance of having a reliable user experience while using PACS is raised in a recent study by Roseland et al (2019) [101] who report that a "stable system with predictable behaviour" that minimises "repetitive non-value-added work", supports "interoperability" and with "near-instantaneous load times" are key requirements in any new PACS system.

Workflow Cohort HPs commented on the lack of consistent patient identifiers between hospitals, noting that this created difficulties for finding previous scans for a patient. In common with several countries, there is no national identity number program currently in place in the Irish healthcare system, even though legislation has been passed to enable this [102]. By contrast, many countries such as the Swedish healthcare system make use of a unique national identifier to identify patients. The primary purpose of such an electronic system is to promote the medical care of individual patients, as well as facilitating their effective management over time [103]. In addition to effective patient management, the use of a unique national identifier also allows "medical data to be used for educational purposes, research and quality assurance schemes" [103]. Currently, it is not easy to compare two scans from a patient in Ireland if the scans are performed in different institutions. Since these scans will have different identifiers, the system interprets these records as originating from different individuals, if the records are not already linked.

Cohort HPs report that errors have occurred when storing images to PACS, with images being stored in incorrect record locations. This type of error is recognised in the work of van de Wetering *et al.*, who advises that it is vital to perform a check to verify that images have been correctly uploaded from PACS to the correct patient directory [98].

Echoing the desire for greater efficiency in the user interface in our study, Gale *et al.* write of the large number of “clicks on the mouse” required, which results in a loss of efficiency for the user [94]. HPs expressed a desire for a significantly simplified user interface, stating that the current complexity of PACS interfaces is a limitation.

Integration of PACS with other information systems The advantages of integration between PACS and other EHR systems are clear. According to Cohen *et al.*, patient clinical information is not always available at the point of care, instead being stored locally where it is created [104]. There are clear shortcomings to this situation in terms of patient care. “The current medical system needs to be integrated, secured, and available to health professionals and patients” [105]. A higher degree of system integration can help overcome the shortcomings of standalone PACS solutions, such as managing user authentication, access control, eliminating inconsistent patient identities, and local audit trail recording. Industry standards such as DICOM and HL7 were key to enabling integration of diverse data sources [104]. It is clear from our study that while progress toward integration has been made, further work is required on improving and standardising the user experience.

Perceived shortcomings in user training for PACS Consistent with this study, Top reports that half of the responders have no training to use PACS, with half of those responders reporting that no training had been offered [100].

Cox reports that users who receive formal training say they find that PACS workstations are easy to use, albeit sometimes “fiddly” [106]. This reinforces Top’s report that there can be significant differences between hospitals, potentially due to differences in types of PACS software used, or levels of user training and experience [100].

Difficulties using PACS can lead to a requirement for external expertise; this is echoed in the work of Fridell *et al.*, who say “If these technicians cannot solve the problem, the vendor’s technicians are called in. This makes the technology more distant to the radiographer, just as it makes the entire solution more complex than before.” [107].

Adoption Supporting the potential for voice interaction noted in this study, Langer reports that speech recognition has a significant impact on productivity (up to 70%)

for production of radiology reports, concluding that the adoption of PACS or speech recognition, or both, improve report turnaround time [108]. This result is echoed by Lepanto *et al.*, who reports significant decreases in dictation turnaround times one year after PACS implementation across several sites [109].

In contrast to earlier work, PACS is seen by the participants in this study as a stable technology and an integral part of the hospital workflow, but the introduction of new PACS software is seen as challenging and potentially disruptive. Thus, difficulties encountered in earlier PACS implementations [107] may arise again as new software or features is introduced. Paré and Trudel note that "merely deciding to adopt PACS does not guarantee success; effective PACS implementation is also necessary." [110].

3.5.3 Strengths and Limitations

This study seeks to address a number of gaps in the existing literature by exploring HP experiences with current PACS installations, as well as their attitudes to touchless interaction with PACS. The semi-structured qualitative approach to elicit themes provides rich insight into routine usage by HPs. The study explores the experiences of a range of stakeholders to provide an overview of the experiences and needs of health professionals using PACS in hospital environments.

The main limitation of this study is that participants were primarily recruited from Irish hospitals. Users in other countries may have different experiences with PACS, and different information systems used in conjunction with it. However, given the general agreement of findings from this study with existing research, it appears that these differences do not appear to have a significant effect on the findings. While the overall sample size is limited by the availability of clinicians, saturation was reached in the analysis of interviews, with later interviews confirming issues already identified in the analysis. A larger sample might allow greater insights by demographic to emerge, however, there the sample under investigation did not display significant differences in response by role or experience.

3.6 Conclusions

This work investigates the real-world needs and requirements of HPs when using PACS through interviews and qualitative analysis. Insights into how HPs use PACS, as well as the challenges they encounter, are presented.

There is a clear appetite amongst HPs for significant improvements to existing PACS. Whilst some of the changes highlighted fall outside the scope of this study, e.g., poor server performance or increased levels of user training, other improvements can be introduced through consideration of how users interact with PACS. As workflows continue to evolve, it is of value to consider novel interactions with PACS, such as those enabled by touchless interaction technologies. The operating theatre is the location where the benefits of touchless interaction could be of most significance. Currently, it is arduous for clinicians to interact with PACS when scrubbed up. Touchless interaction, whether voice control or gesture-based, would allow HPs direct control of PACS and would provide greater access to advanced imaging to surgeons in the operating theatre.

The hospital environment presents a range of technical challenges for any touchless interaction system. The hospital is a loud, busy environment where both voice control and gesture-based recognition will experience challenges. It is also an environment where operational efficiency is a requirement. In order for a technology to be successfully adopted by HPs, it must deliver a high-quality user experience and must avoid impeding HPs in their day-to-day tasks.

For PACS developers, there are a number of key points. Firstly, though PACS provides a powerful set of features, it suffers from poor usability. PACS interfaces are overly complex for the average user to effectively discriminate and exploit the set of features most useful to them. This study suggests that improvements to usability would enable more effective use of PACS. For example, users should be able to view a simplified interface that presents the tools most relevant to their role.

Secondly, PACS would benefit from additional interaction mechanisms, especially voice commands. There is a strong appetite among clinicians for voice commands that could simplify and speed up their workflow. For example, the ability to verbally instruct a

PACS to display all images for a patient, including previous imaging, is mentioned by a number of clinicians as a functionality they would highly value.

In the context of investigating alternative interaction mechanisms (such as touchless interaction), from HP feedback, there is a need to investigate multiple modalities. HPs expressed interest in, and enthusiasm for, different input techniques, envisaging using gesture and/or voice input. Furthermore, HP workflow and work environments may benefit from enabling different touchless modalities.

HPs expressed a strong need for any touchless system to deliver reliable, efficient performance. Performance challenges such as slow recognition rates or unintended touchless inputs would be significant issues for medical users. Furthermore, HPs may be reviewing patient image records in a ward context; unintentional input resulting in displaying information for the wrong patient would represent a data privacy risk. To bring insight to these areas, touchless clutching mechanisms will be investigated in general in Chapter 5 (p.115), and in particular as a means of preventing unintended inputs when using touchless medical image systems.

Chapter 4

Exploring Touchless Interaction through the Development of Prototype Systems

In this chapter, the process of designing two unique prototype touchless control systems is presented with the goal of identifying major steps and decisions in the design and development of touchless interaction systems. The first prototype system was developed using the Kinect V2; providing insights into touchless interaction generally. The second prototype system was then developed using the much improved Kinect Azure DK; this second system integrated the learnings from the first prototype and the thematic outcomes from the interviews of HPs regarding their needs (presented in Chapter 3, p.45). The second prototype further provides a testbed for exploring the user experiences of clinicians operating a touchless system for medical imaging, which will be employed for work presented in the next chapter. The process of developing such touchless systems through combining findings from existing literature, feedback from clinicians, and the experience of designing two touchless systems is discussed.

4.1 Introduction

4.1.1 Motivations

Numerous decisions must be made when considering the design of a touchless medical imaging system, ranging from hardware, software, interface design, system usability, and performance. Understanding the use context as well as the targeted user cohort is essential when approaching the decision process. There is much to be learned through the process of developing prototype systems in this respect.

4.1.2 Contributions

The following contributions are made in this chapter:

- Context is developed for developing touchless interaction systems through a review of published topics including natural user interfaces, the role of technology, the importance of use context, and the roles of user feedback and interface latching (clutching).

- The development of a first prototype based on the Kinect V2 is described, exploring the development of touchless interaction in the context of imaging generally.

- The development of a second prototype using the Azure Kinect DK is described, exploring specific elements of touchless control of medical imaging specifically, such as active zone control, natural language elements, and the role of clutching.

4.2 Background

Natural User Interfaces

The role of touchless interactions in natural user interfaces has been frequently discussed, with much research claiming that gesture-based interactions can deliver more natural modes of interacting with computers [111]. However, as O'Hara *et al.* observe, naturalness is not conferred by technology; rather, it is the ways that people make the actions they perform with the technology "apposite, appropriate, or fitting to the particular social setting and their particular community" that confer naturalness [112]. According to O'Hara *et al.*, "the naturalness of how a technology might be interacted with lies not in the physical form of that technology, nor in any predefined interface (natural or otherwise) but in how that form and the interface in question meld with the practices of the community that uses them."

Previous research on NUIs has tended toward either of two approaches. One approach takes the view that interfaces and interactions need to echo the user's real-world actions. Saffer says "the best, most natural designs, then, are those that match the behavior of the system to the gesture humans might actually do to enable that behavior" [113]. Jacob *et al.* describe this as making HCIs "more like interacting with the real non-digital world" [114]. There is an assumption in this approach that existing communicative gestures are pointers toward common or universal "natural interactions". A potential shortcoming of this representational approach is the lack of social context, where natural interactions are not constituted by situational context but are brought to it [112]. Meanwhile, other research, such as Norman, criticises this approach [115] preferring instead to focus on the interface as a source of explanation for naturalness, usability, intuitiveness, and learnability. Both approaches risk the loss of opportunities for obtaining a better understanding of what can be achieved with touchless technologies.

O'Hara *et al.* state that, rather, in situ and embodied aspects of interaction with touchless technologies should be a more fundamental element of our understanding of naturalness, moving from a simply representational concern to an interactional one [112]. In this approach, the social basis of the individual experience is emphasized. Naturalness becomes not something to be represented but an "occasioned property" of action, something that is "actively produced and managed together by people in particular places and at particular occasions" [112].

4.2.1 Understanding Touchless Interaction

The user's experience of interacting with touchless interaction interfaces diverges from touch-based systems in several ways.

Unlike touch-based interactions, touchless interactions can be performed across a range of proximities to the surface of a system, extending from centimetres to several metres. With a touch-based system, a user is required to be co-proximate to the surface they are touching. Further, the surface must be accessible, and within reach. This limitation can be removed using touchless interaction, where the distance at which the touchless interaction takes place is dependent on the touchless technology being employed.

One important motivation for using touchless interaction is the prevention of cross-contamination. When using touch-based interfaces, contact with the surface interface results in a transfer of material from the user to the device and vice versa. Touchless interactions, in contrast, avoid this transfer. This attribute is of especial benefit where the management of biological cross-contamination is important, such as in a hospital setting.

Touchless interaction avoids the need to apply momentum and pressure to the surface being touched. This reduces the amount of "movement, damage, erosion, and attrition" of the surface. This can be beneficial in extending the operational lifespan of heavily used equipment.

With touch-based interactions, movement is "bound and constrained by the shape and properties" of the surface being touched. In contrast, touchless interaction technologies do not constrict movement in this way, allowing for greater freedom.

O'Hara *et al.*'s properties of touchlessness framework can provide useful grounding when approaching touchless interaction design [112]:

Properties of touchlessness

Social context + context of use

According to Rico and Brewster, social acceptability should be considered when designing gesture-based interfaces [116]. The social acceptability of an individual's actions is determined by bringing together information about the current environment with their existing knowledge [117]. It is through action that people create shared meanings with others; these meanings create a "common ground" that allows a socially organized, understood, and coordinated experience to be formed from individual perception [118]. The

Use Case	Description
Wake up	Waking (clutch) the system via touchless input
Validate identity	Authenticating the user and provide access to the medical image database
Select patient/image set	Selecting the patient/image set using touchless input, such as voice command or gesture interaction
2D image operation	Windowing, zooming, browsing/comparing, and measuring via multimodal input both beside the operating table and in front of the medical screen
3D image operation	Rotating, zooming, measuring, selection, and hiding via multimodal input both beside the operating table and in front of the medical screen

Table 4.1: Use cases for touchless interaction with medical imaging systems in the OR [2]

importance of the social meaning of action can also be seen in the OR. The act of not touching the mouse during an aseptic procedure is more than simply avoiding contact with the surface of the mouse device; rather, it demonstrates the "ongoing commitment to the unchallengeable delineation between sterile and non-sterile; to making that delineation "real" by "doing"" [112].

Merleau-Ponty describes the difference between the objective body and the lived body [119, 120]:

- **Objective body** how the user's bodily actions might be described by a third person in terms of abstract muscular performance.
- **Lived body** how the user experiences the world through their own bodily actions.

Larssen *et al.* emphasised the centrality of the body when understanding the potential for action [121]. By shifting perspective to that of the user, it is possible to achieve significant insight into what human-machine interaction can entail. This, in turn, can lead to innovative approaches to touchless interaction systems [112].

4.2.2 Touchless Technologies

The advent of mature, easily accessible touchless technology devices such as the Microsoft Kinect, amongst others, resulted in a significant increase in the number of publications dealing with touchless interaction in the OR [1]. These technologies have had a profound impact on the collective imagination, resulting in the development of interaction paradigms, beyond the mouse and keyboard [112]. Developments are not limited just to vision techniques, but also such technologies as natural language interfaces [112].

These novel technologies offer intuitive interface modalities that do not require specialist techniques for communicating with computers [112]. By combining technology and the material world, our interactions with computers can be configured in new and meaningful ways. Furthermore, these technological developments are not limited just to a healthcare setting, but to many other applications, such as ATMs, vending machines, and learning devices [122]. Touchless applications prove especially valuable in a during- and post-COVID-19 world for managing cross-contamination risk.

Figure 4.1: Body tracking using 2D video and MediaPipe

Depth camera technologies have seen significant advancements in recent years, with the advent of hardware technologies such as the Azure Kinect DK, and software solutions for body tracking such as MediaPipe representing a new maturity in the field (Figure 4.1, p.80). Devices such as the Ultraleap (formerly Leap Motion Controller), Microsoft Kinect, and the Apple's Face ID have brought depth camera technology and gesture control into people's homes and offices. The Microsoft Kinect is one of the best selling pieces of consumer hardware of all time, having sold more than thirty-five million units, and holds the Guinness World Record for the fastest-selling consumer electronics device. Though

Microsoft has discontinued the original Kinect, it nevertheless brought touchless interaction to the average user's living room and into the public consciousness. Though modern Kinect devices target industrial applications, virtual reality offerings such as the Vive Index or the Oculus Rift have continued to attract public attention.

Touchless sensing

The use of touchless sensing devices has become increasingly common during the COVID-19 pandemic, with touchless sensing gel dispensers, and car park ticket devices being a common sight in many venues (Figure 4.2, p.81). However, such interactions have long been common in applications such as automatic doors or automatic lights. Due to the pervasive nature of these technologies, they have become a part of many people's daily experiences.

Figure 4.2: Automatic gel dispensers, already present in hospital environments, have become a common sight in non-medical contexts in recent times due to COVID-19.

Voice recognition

Entirely voice-controlled products such as Google Home, Amazon Echo, and Apple Homepod have brought voice control into people's everyday lives, from the kitchen to the nursery. The concept of talking to a computer is a step change from being the stuff of sci-fi into the real world. This is especially so when provided in a format that is affordable for the average consumer. Over the 2017 Christmas period, Amazon Alexa was the most downloaded app on the Apple App Store, which points to increased user adoption of the voice-controlled Amazon Echo smart speakers [123]. Further evidence of this growth can be seen in the rapid increase in Alexa enabled devices between 2019 and 2020, with the number of Alexa enabled devices doubling from 100 million in 2019 [124]. Technologies such as Siri and Google Assistant have brought sophisticated voice control to hundreds of millions of smartphone users, and with the rise of smart homes, dedicated devices such as Google Home and Amazon Echo are bringing voice control into the home. The accuracy of voice recognition continues to improve through ongoing investment in these platforms.

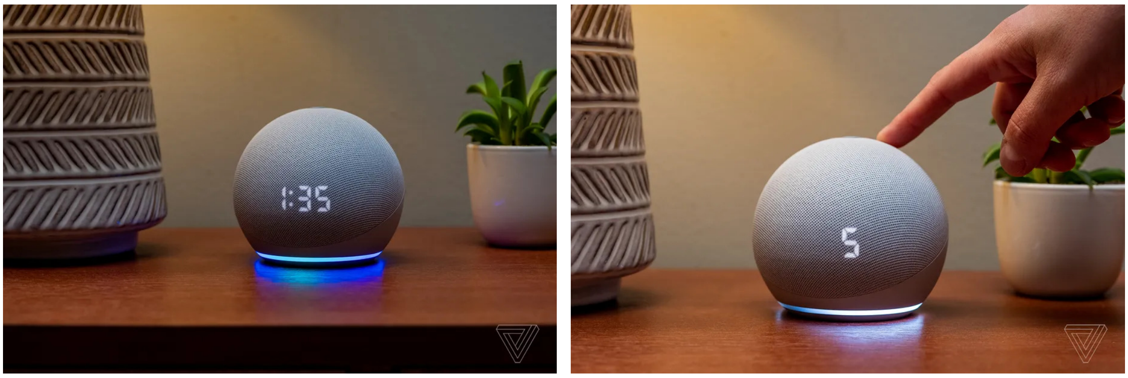


Figure 4.3: The Amazon Echo supports both voice control and touch input. An accessible price point, useful functionality, and appealing design have made such touchless devices common. ©The Verge 2020.

Facial recognition

Facial unlock technology is most commonly found on mobile phones and computer devices; since 2017, Apple devices such as iPhones and iPads have used Face ID as a biometric security feature, removing the need to enter a password or touch a fingerprint scanner. Microsoft introduced Windows Hello with Windows 10, which provides facial recognition for authentication on compatible devices, removing the need to enter a password or touch a fingerprint scanner. Biometric solutions, such as face recognition, can provide high levels of security, provided appropriate encryption solutions are used. Face ID, found in iPhones and iPads, use a combination of an RGB camera, an IR camera, and a dot projector to accurately recognise faces reliability (Figure 4.4, p.83). The resultant biometric signature is encrypted on the device for security. This combination of camera elements provides much higher levels of reliability than a system that relies on just RGB data, due to the additional depth information provided by the IR camera and dot array.

Figure 4.4: Apple Face ID authentication functions by projecting and analysing thousands of IR dots to create a depth map of the user's face sensors. ©iFixit 2017. Reproduced with permission.

A significant disadvantage of using facial biometrics is the need to maintain a database of users' facial data. User acceptance for the storing of this data may be a challenge when implementing this technology. There is also a data governance issue in terms of providing sufficient protection for this data. Unlike a password, which can be reset, a biometric signature, once taken, cannot. Touchless authentication provides various security mechanisms that have a low barrier to entry for the user. Applied correctly, touchless authentication could enhance both security and the user experience.

Personal devices

Body-mounted devices are another option for touchless interaction. One type of such a body-mounted device is an inertial sensor. These devices receive input from combinations of motion sensors (accelerometers), rotation sensors (gyroscopes), and, potentially, electromyographic sensors (e.g., in the Myo Armband (Thalmic Labs, now discontinued)). This information is used to continuously calculate using dead reckoning, the position, orientation, and the velocity of the device without the need for external reference. Another type of body-mounted device is the smart glove, such as in Figure 4.5 (p.84). These gloves use arrays of stretch sensors, embedded into a fabric layer, to detect hand and finger movements in real-time. Body-mounted sensors can enable silent, touchless control, without the need for a sensor to maintain line of sight with the user. This helps to isolate the touchless control from the environment. The primary disadvantage of such devices is that their use is limited to a single user. If there is a need for concurrent users in a clinical setting, a separate sensor must be provided for each user. If a single device is to be shared by multiple users, then there is a need for correct sterilization practices between changeovers. Without such practices, body-mounted sensors can become cross-contamination sites. Additionally, due to its mobile nature, there is an increased risk of the hardware being misplaced, which could leave users unable to interact with systems.

Figure 4.5: Tyndall's Smart Glove provides an easy-to-wear solution that can allow the user to interact with both digital and physical devices [125]. Reproduced with permission.

4.2.3 Designing Touchless Interfaces for Hospital Environments

To date, the majority of research has focused on the use of touchless control within the OR, with little consideration for application more generally in a hospital setting. The primary reason for this focus on the OR is that asepsis, which is a significant benefit of touchless interaction, is key in a surgical environment. The thoughtful design of touchless interaction also has the potential to improve HP work, especially in aseptic contexts. Touchless control has the potential to return direct control of key imaging systems to scrubbed users, such as the surgeon in the sterile field.

In this chapter, the Microsoft Kinect V2 and the Azure Kinect DK are used to establish the potential of touchless interaction in a hospital environment. Touchless interaction technology is most effective when directly integrated into hospital devices. Ideally, the relevant software SDKs would be available to hardware manufacturers who would integrate the touchless interaction framework into their products. However, an alternative approach could be the provision of customisable software and hardware that can be installed by the hospital. Such software would act as an interaction layer on top of existing software platforms. This "bolt on" approach would, however, be less functional than a fully integrated touchless interaction system.

In order to design an acceptable and usable touchless medical image system, it is important to give consideration to the efficiency, effectiveness, and learnability of the system's HCI components in the appropriate context, such as the surgical context considered by Hui *et al.* [2]. According to Norman, "good design takes care, planning, thought, and an understanding of how people behave" [126]. Preece *et al.* note good interaction design should produce a product that is "easy to learn, effective to use and provide an enjoyable user experience" [127].

When designing touchless interactions for any system, there are primary considerations that should be considered:

Naturalness Interactions with a system must feel natural for a user. If an interaction feels difficult, strained, or the user is concerned that they appear foolish, the user is unlikely to use it. However, if the user finds an interaction natural, acceptable, and easy, they are much more likely to continue using it.

Learnability For most users, touchless interaction is novel. Some interactions must be learned, e.g., the “Ok Google” command to trigger Google Assistant. It is essential that the use of these new commands and interactions are easy to use, do not increase cognitive load, and should take no more than a small number of use cycles to learn.

Convenience A touchless system needs to be convenient to use. A user should not have to go to additional effort to use it. Methods such as always-on listening, which is used by devices such as Google Home, provide this convenience; merely saying the keyword or phrase at any point in time allows the user to interact with the system.

Functionality As with any interaction mechanics considerations, it is essential that a touchless interaction system be functionally relevant. The system must deliver clear value to users or else they will not be vested in using it.

Efficiency

Healthcare professionals in hospital environments are often extremely busy; they require the technology they use to be as efficient as possible. If a technology is perceived as preventing the healthcare professional from completing their tasks both efficiently and to their satisfaction, that technology is unlikely to be adopted where an alternative exists. For touchless interaction technology to be successful, it must therefore exhibit comparable levels of efficiency compared to existing interaction technologies, such as the mouse and keyboard.

Authentication

A common application of touchless interaction in hospitals is for user authentication, such as with NFC or RFID security badges. By holding a badge in front of a scanner, a door can be unlocked, provided the user has sufficient access privilege. Badge based touchless authentication provides several key benefits. They are generally very straightforward for users to use, with a limited learning effort. However, badge-based systems only offer basic levels of security as there is often no verification that the badge is being used by the intended user, a badge can be misplaced, given to another user, or stolen. Biometric solutions, such as facial recognition, can provide much higher levels of security provided appropriate hardware solutions are used.

Touchless authentication facilitates various levels of enhanced security that has a low barrier of use for the user. Applied correctly, touchless authentication could enhance both security and the user experience.

Frameworks for the design of touchless interaction

It is useful to combine an understanding of touchless interaction with structured approaches such as the framework presented by Karam (Figure 4.6, p.87). This framework is tree-structured, consisting of four primary categories; application domain, enabling technology, system response, and gestures. Each primary category is comprised of several sub-categories and each sub-category contains various parameters. Parameters represent variables that can be configured and compared between different systems, e.g., how many gestures are included in a lexicon. Though Karam's framework is presented in the context of gesture-based interaction, it may be useful to extend and adapt it to include alternative touchless modalities, such as voice control.

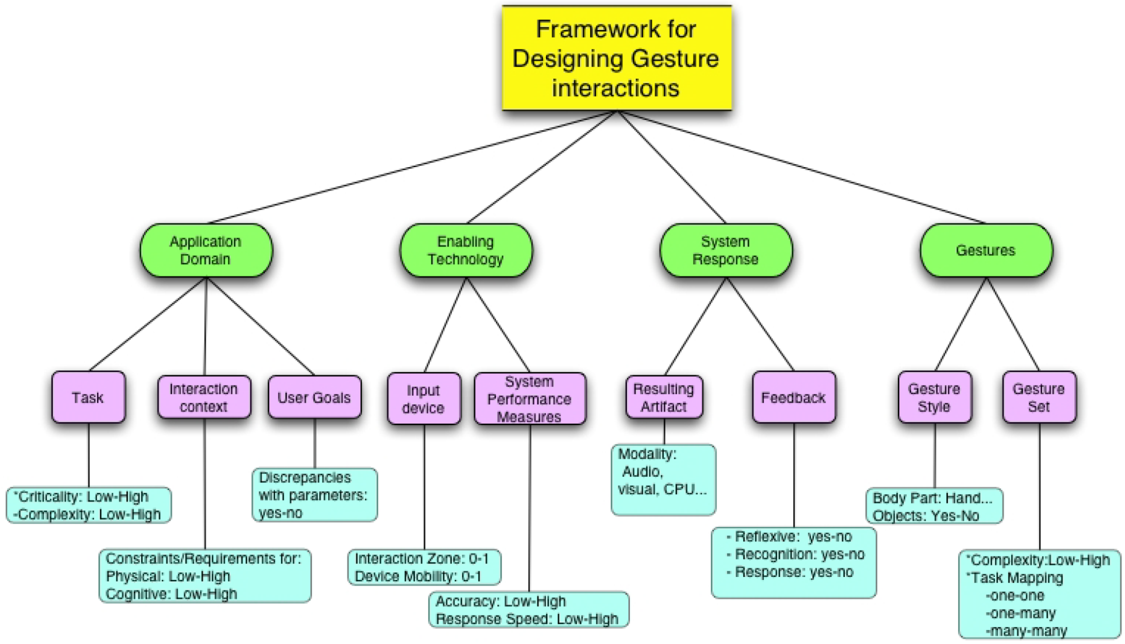


Figure 4.6: Karam's proposed framework for designing gesture interactions. The top row contains four primary categories, each with various sub-categories [128]

In Karam's framework, the application domain assembles information regarding applicable technologies, tasks, scenarios, and contexts to deliver insights that can guide design. An example of understanding the application domain can be seen in Hui *et al.* whereby through performing task observation and evaluating environmental constraints, the researchers were able to determine a number of design challenges when developing a touchless interface for use in the OR [2]. These challenges included; the limited physical space available for touchless interaction in both pre-operative and intraoperative scenarios, and the potential for noise disturbances in the OR. Combining these constraints with an understanding of specific user behaviours, such as surgeons' hands remaining at their chest to avoid accidental contamination [65], can inform the design process. Hassan *et al.* present another example of understanding the application domain through understanding the interaction of the environment with the sensor technology being used. In their work, they discovered that the Kinect V1 camera's performance, which depended on an IR camera for depth data, was being affected by the heat of the OR's halogen lights [55]. Investigating the application domain can help avoid performance issues which would adversely impact the system behaviour and on the user experience.

Nguyen presents a useful framework for the development of touchless NUIs (Figure 4.7, p.89). O'Hara *et al.* state that it is important to consider how touchless systems provide benefit through reconfiguration of existing work practices, thereby changing how the world is experienced [112]. Understanding the context of use of such systems is central to this consideration. Nguyen presents a framework for building touchless NUIs with an emphasis on providing user feedback [129]. Though this framework focuses on gesture interaction, it can be expanded to incorporate other touchless modalities.

Clutching

Unintended input is revealed as an issue with touchless control of medical imaging in subsection 2.2.3 (p.22), with HPs noting potential issues such as data governance in the qualitative study presented in Chapter 3 (p.45). One way of reducing the likelihood of unintentional input to a sensing system is to employ a clutching mechanism [130], to help the system determine a person's intention to address it (i.e., to direct their actions

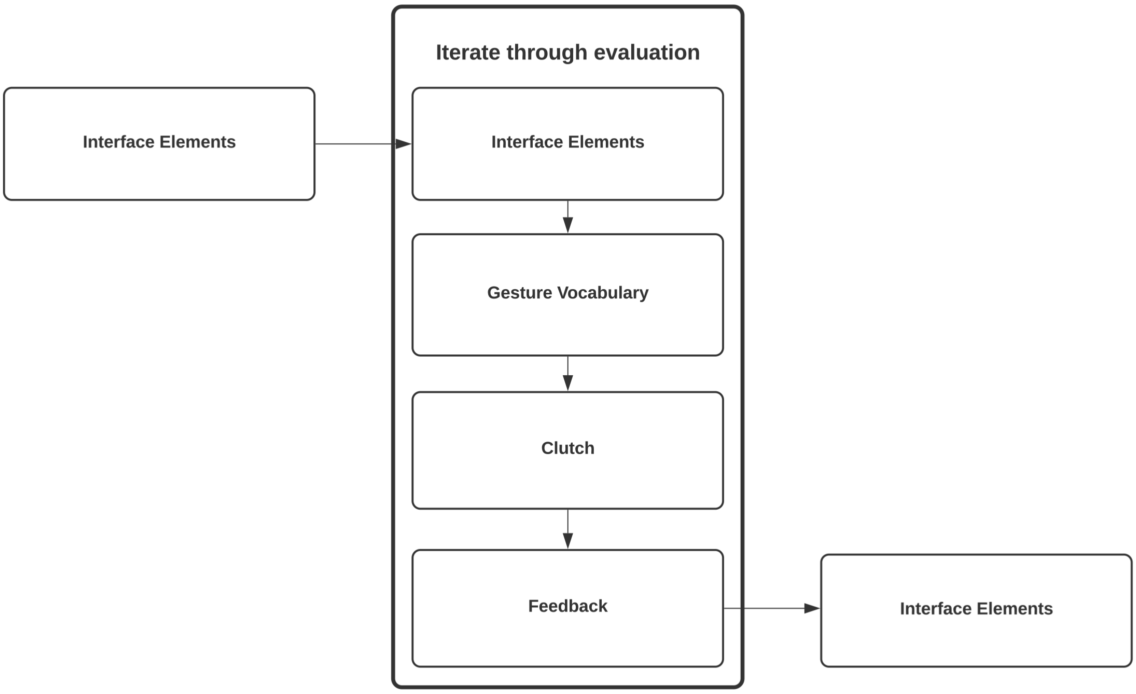


Figure 4.7: Nguyen's framework for building touchless NUIs [129]

towards it [131]). Clutching mechanisms may require purposeful actions that act as a 'mode switch' or indicate intent. A system may also indirectly infer intent from a set of constraints and contextual cues. These actions and cues help a sensing system to segment input, identifying patterns in input data that correspond to intentional interactions.

Clutching (latch/unlatch mechanics that prevent accidental interaction with a touchless system) is noted as being of importance for most graphical user interfaces (GUI) [132], especially those with touchless interaction. According to Woźniak *et al.*, clutching is a fundamental operation, with most GUIs requiring a clutching mechanism [132]. The framework presented by Nguyen (4.7 (p.89) explicitly includes clutching in the design process. In a post-WIMP ("window, icon, menu, pointing device") world, clutching is of increasing importance; it is a feature that people do not even think about. When they say "Alexa, what's the weather today?", they are using the clutching command, "Alexa". Clutching is both vital and, frequently, invisible in successful touchless operation. It reduces inadvertent interaction that would result in high levels of error and user frustration. User frustration can result in poor adoption of technologies, and hence robust clutching mechanics are essential for touchless interaction with medical image viewers such as PACS to be possible. The field of touchless interaction has reached maturity, and standardised gesture/task sets will eventually be defined. Despite its fundamental nature, clutching, however, remains an unresolved issue.

A central element of the research presented in this chapter is the investigation of the design of touchless clutching for medical environments. Based on previous qualitative research investigating the needs of clinicians, as well as a review of the existing literature, it is clear that there are many situations in hospital use that would favour the ability to clutch a touchless system. Situations such as a clinician needing to turn away from a system, or a workspace becoming busy, or even the need to fully lock a system for data security all contribute to this need.

Clutching techniques

The intent to interact can also be inferred by how or where users perform actions. Baudel *et al.* [133] described the use of an "active zone", as an area of space where sensed movements are treated as an intentional input. This can limit the available input space significantly and, if clearly delineated, can help users visualise where gestures will be reliably sensed. Alternatively, information about body posture and gaze can be used to infer an intention to interact [134]. For example, Jacob *et al.* [135] used information about head and body orientation to determine when a user was intending to interact with an MRI system. While this reduced the false-positive gesture rate in their study, O'Hara *et al.* [136] suggested those contextual posture cues may be misleading in other usage scenarios (e.g., when looking at images with others).

According to Hatscher and Hansen, "Without a clutching mechanism, involuntary manipulation of supporting information can happen quite easily. This might either lead to decisions based on wrong information or demand extra time to revert the system to the desired state" [75]. This is undesirable behaviour in any system, and is especially to be avoided in medical imaging systems, where patient outcomes and safety may be affected by such issues.

User Feedback

Following the framework presented by Nguyen (Figure 4.7, p.89), as well as the results presented in subsection 2.3.4 (p.38), providing appropriate feedback is important to ensuring a successful user experience with touchless interaction systems.

In order for users to feel confident when using a touchless system, providing feedback through methods, such as; on-screen text as in Bockhacker *et al.* [65], or by rich visual feedback, such as that presented by Chiang *et al.* in Figure 4.8 (p.91), can provide reassurance for the user when using novel touchless interaction modalities.

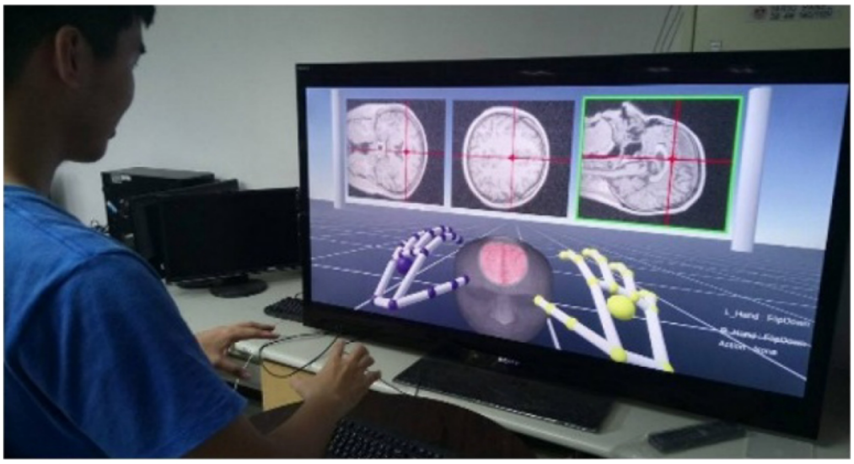


Figure 4.8: Touchless interaction environment showing visual feedback [74]. By showing a digital representation of their hands to the user, there is a clear visual cue when the system recognises a user action.

Karam discusses three stages of feedback; **reflexive**, which informs the user about the current state of the processor, **recognition**, which occurs at the end of the processing stage and informs the user which touchless command has been recognised, and **response**, which indicates that a task is complete [128]. Feedback that raises user awareness of the value of touchless interaction has also been flagged as valuable to adoption. Making users aware of the risks of cross-infection can enhance the acceptance of a touchless system in clinical practice [55], especially in the context of highly infectious diseases such as COVID-19.

One feature of touchless interaction that presents a potential limitation to the technology is the difficulty in providing haptic feedback. This can limit the opportunity for fine-tuning and refining user manipulations in the moment. In contrast, touch-based interaction presents immediate and familiar tactile feedback to the user. The satisfying click of a button lets the user know that they have successfully sent a command to a system. Though some research looks to solve this issue through approaches such as ultrasonic tactile feedback ([137, 138]), it remains a limitation with many touchless technologies.

4.3 Investigating Touchless Control through an Initial Prototype System

4.3.1 Motivation

In order to gain a deeper understanding of the processes, decisions, and challenges associated with developing a touchless interaction system, there is considerable value to developing prototype systems. Doing so brings nuance and understanding of the design challenges. The act of developing a system both answers and raises questions about the development process and associated design decisions. Furthermore, developing such a prototype facilitates testing with pilot users. A pilot study is key to developing an understanding of user needs and experiences with such a system.

4.3.2 Design

Background research was performed to gain an initial understanding of existing interfaces. Both PACS and DICOM image viewers were studied, in order to build an initial understanding of the functionality provided by these systems and how they operate.

Various DICOM viewers revealed the relative prominence given to the measurement feature in existing viewers; this can be seen in Figure 4.10 (p.94).

In Figure 4.11 (p.94), it can be seen that PACS contain similar visual elements to the DICOM viewers in Figure 4.9 (p.93). However, further inspection revealed a greater functional complexity to PACS. While a DICOM viewer serves simply as an imaging viewer, a PACS generally contains an imaging viewer as a part of a more extensive software package. PACS delivers greater functionality for accessing remote patient scans and reports as compared to DICOM viewers.

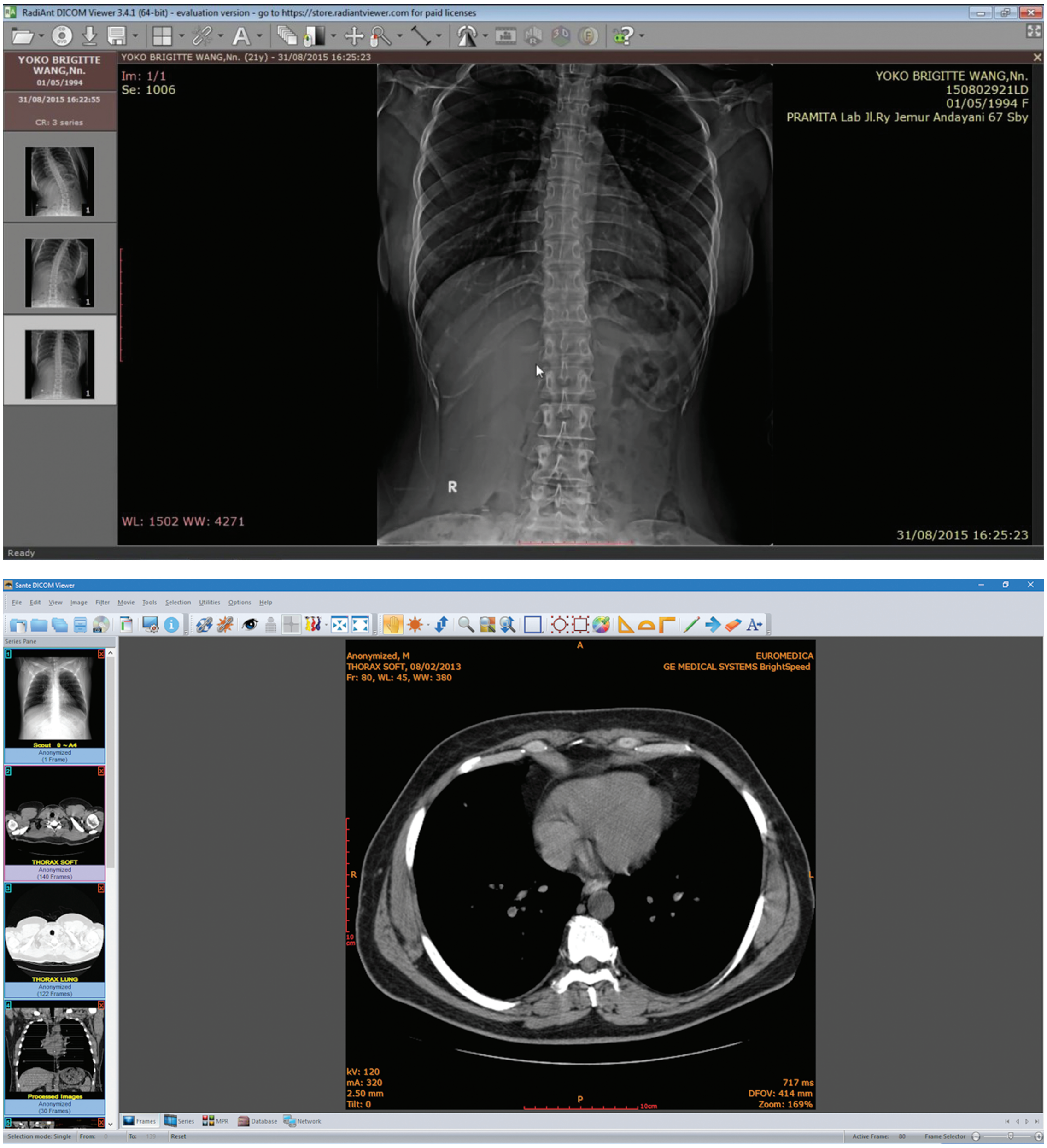
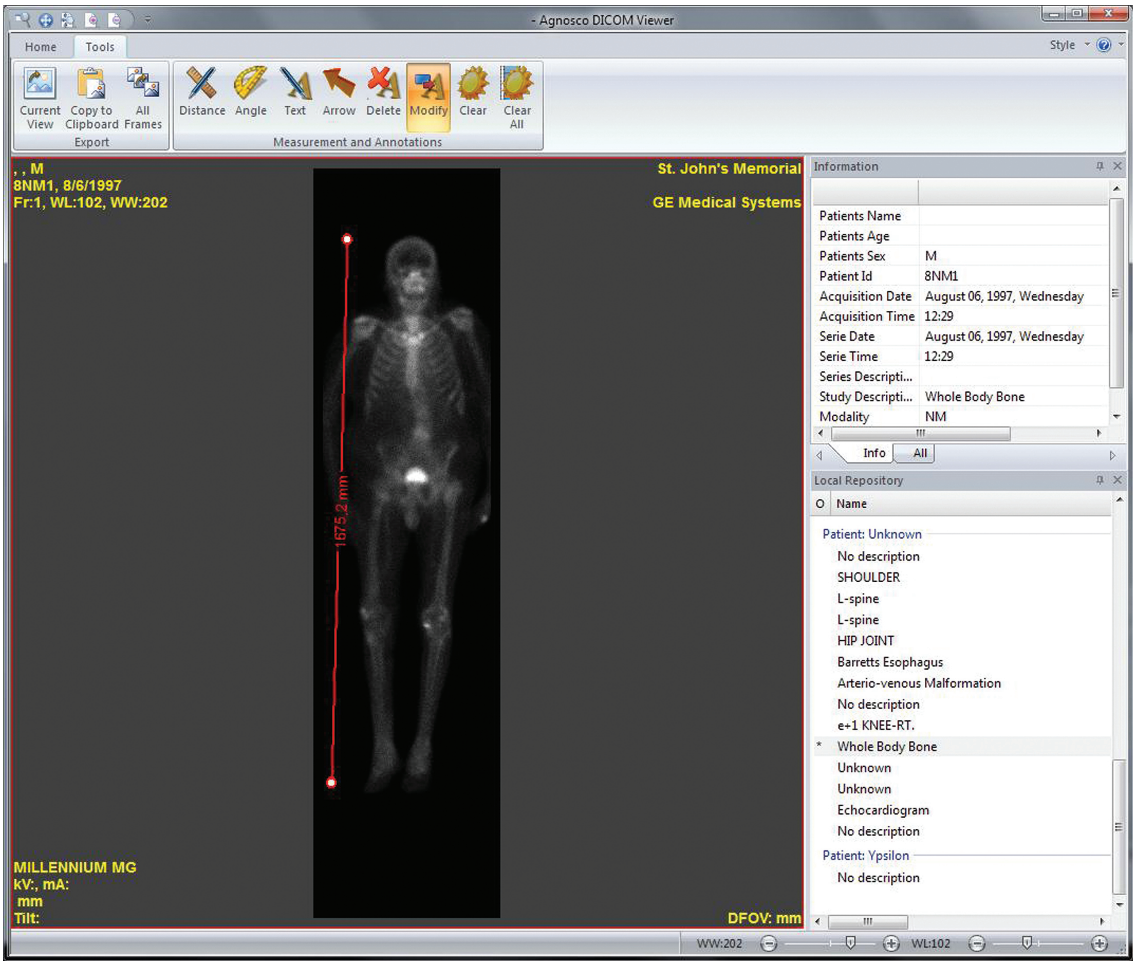


Figure 4.9: Top: RadiAnt DICOM Viewer 3.4.1. The toolbar along the top of the interface allows for quick access to key functionality. Bottom: Sante DICOM Viewer. Compared to the RadiAnt DICOM Viewer interface, the Santa DICOM Viewer interface is more visually cluttered and less user-friendly.



Patients Name	
Patients Age	
Patients Sex	M
Patient Id	8NM1
Acquisition Date	August 06, 1997, Wednesday
Acquisition Time	12:29
Serie Date	August 06, 1997, Wednesday
Serie Time	12:29
Series Descripti...	
Study Descripti...	Whole Body Bone
Modality	NM

Patient: Unknown
No description
SHOULDER
L-spine
L-spine
HIP JOINT
Barretts Esophagus
Arterio-venous Malformation
No description
e+1 KNEE-RT.
* Whole Body Bone
Unknown
Unknown
Echocardiogram
No description
Patient: Ypsilon
No description

Figure 4.10: Agnosco DICOM Viewer's measurement mode. The measurement value is difficult to read due to its small size and rotated 90° to the plane of view; larger horizontal text would be preferable.

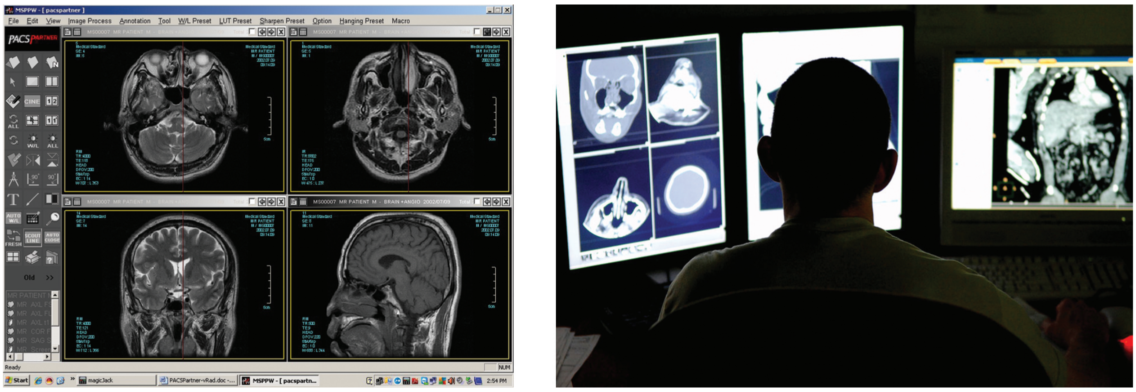


Figure 4.11: Left: PACSPartner user interface showing four views of a brain scan. Right: A typical setup for a radiologist using PACS. Radiologists generally use PACS in darkened rooms in order to optimise contrast as much as possible.

User interface

A prototype interface was designed and implemented to simulate a basic image viewer, without attempting to replicate the functionality of a PACS. The decision was made to exclude certain functionalities, such as viewing multiple images at once, in this design. Though such features are important elements of real-world systems, the goal at this point was to create a relatively simple system to build an understanding of basic touchless control of medical imaging. The ability to switch between touchless input and mouse and keyboard input was included in the design, as well as support for a limited set of interactions (Table 4.2, p.97). The primary goal when considering the design of this first prototype was an investigation of touchless control, with a view to understanding how this compared to traditional mouse and keyboard input.

One advantage of a touchless interaction interface is that it requires fewer on-screen controls. As a result, the application content, i.e., the medical images, can be given higher visual priority, which results in a clearer, less confusing interface with fewer distractions for the user. The application was designed to run full-screen with the medical image prominently displayed in the central area (Figure 4.12, p.96).

Touchless lexicon

An initial gesture set was developed to provide the control set in Table 4.2 (p.97). The tasks were chosen to represent several key image navigation functions, as well as rotation and measurement, which are commonly used functions in existing systems. A mechanism for converting hand positions in 3D space to 2D screen coordinates was also designed to allow for real-time measurement between selected points in MRI imagery. Gestures were trained using the Visual Gesture Builder (VGB) Gesture Wizard (Figure A.1, p.212), and were tested for performance using the Visual Gesture Builder Viewer (Figure 4.15, p.100).

Though the system was initially designed to only use gesture input, following initial testing, it was decided that voice control should also be included. This would allow multi-modal input using voice commands to select the system mode, and gesture control to perform continuous interactions, such as rotation. A basic vocabulary of targeted

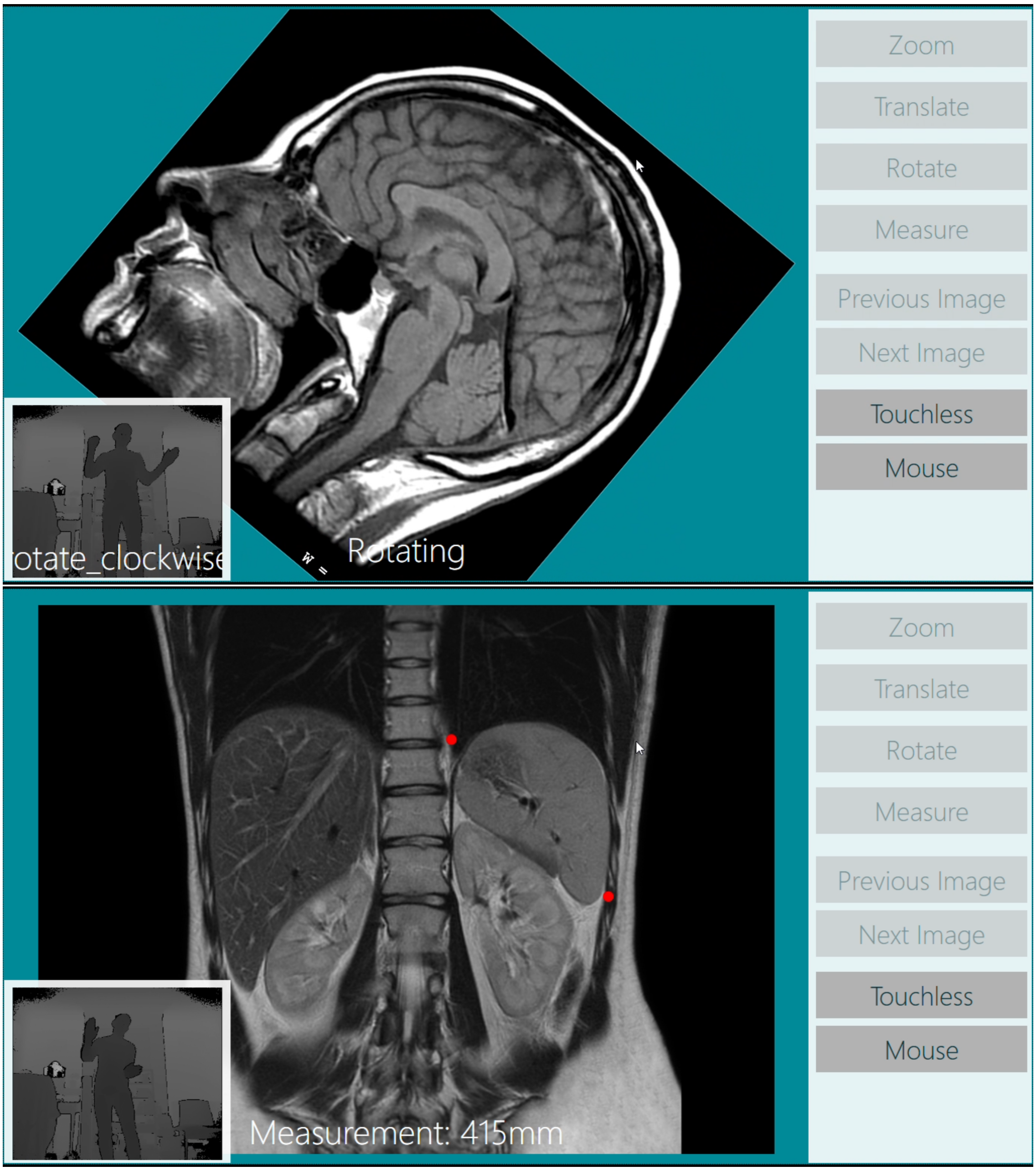


Figure 4.12: Top: The interface of the custom software. The lower left corner displays the live feed from the Kinect depth camera. The controls on the right-hand side allow mouse and keyboard control of the software. The majority of the interface is used to display the medical image the user is interacting with. The currently detected action (rotate_clockwise) is named in the bottom left, and the action chosen using voice command is shown at the centre bottom (Rotating).
Bottom: Performing a measurement using the touchless interface. Each of the user's hands is mapped to one of the red dots, and the distance between these two points is displayed at the bottom of the screen.

Interaction	**Type**
Zoom	in, out
Translate	up, down, left, right
Rotate	clockwise, counter-clockwise
Measure	—
Navigate	next, previous

Table 4.2: A gesture interaction was trained for each supported image control function

keywords was designed and contained the commands shown in Table 4.3 (p.97); this was stored in a key phrase dictionary using XML.

Command type	**Command**	**Keywords**	**Note**
Navigate	Next	'next'	used to navigate to the next image
	Previous	'previous', 'back'	used to navigate to the previous image
System mode	Measure	'measure'	used to set the system to measurement mode where measurement tools are shown to the user, and other gestures are not accepted
	Gesture	'gesture'	used to set the system to gesture mode, listening for gesture commands and hiding measurement tools
Zoom	Zoom	'zoom', 'zooming'	
Rotate	Rotate	'rotate'	
Translate	Translate Up	'translate up', 'pan up', 'move up'	
	Translate Down	'translate down', 'pan down', 'move down'	
	Translate Vertical	'translate vertical', 'pan vertical', 'move vertical'	users were advised not to use this command due to poor level of robust distinction between up and down translation gestures
	Translate Left	'translate left', 'pan left', 'move left'	
	Translate Right	'translate right', 'pan right', 'move right'	
	Translate Horizontal	'translate horizontal', 'pan horizontal', 'move horizontal'	users were advised not to use this command due to poor level of robust distinction between left and right translation gestures

Table 4.3: Following initial testing, voice control was added to the system.

User feedback

A key consideration when designing a touchless interface is how best to provide feedback to the user. In this case, the decision was taken to design the interface with a live depth camera image feed presented in the bottom left corner with the currently detected user pose name displayed on top of the depth image. This feature was included to reassure users that the system was “seeing” them and also allowed the user to ensure they are positioned correctly in the sensor frame. The currently selected gesture is displayed in the bottom centre of the screen, or, when in measurement mode, the current measurement value is shown. This allows users to verify that the system was set to their desired gesture mode. This system was informally tested to gain insights into use and performance challenges of using touchless interaction with DICOM viewer/PACS.

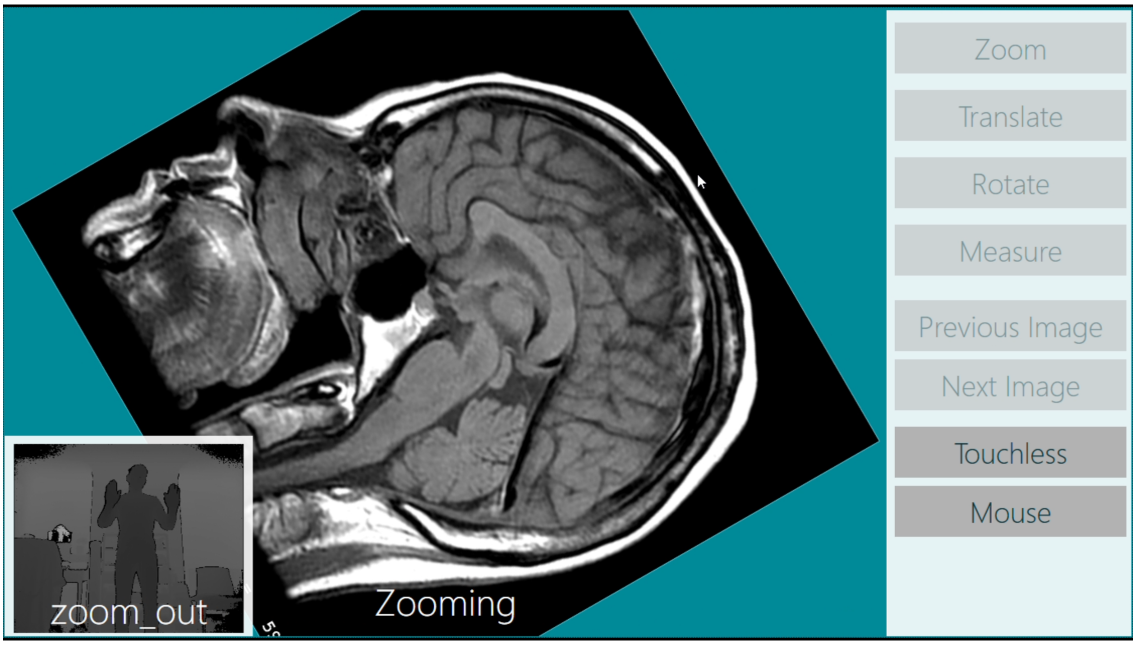


Figure 4.13: Primary interface showing dominant medical image placement. Feedback to the user in the bottom left provides information about the current interaction type. Buttons on the right allow for switching between mouse and keyboard and touchless control.

4.3.3 Implementation

Hardware

The Kinect V2 camera was used to enable touchless sensing (Figure 4.14, p.99). At the time of development, this represented the most current version of the Kinect camera. This sensor provided a combination of colour and depth imaging, as well as microphone input, and an official Microsoft SDK to support the development process.

The Kinect sensor was mounted on a tripod, which allowed for easy re-positioning.

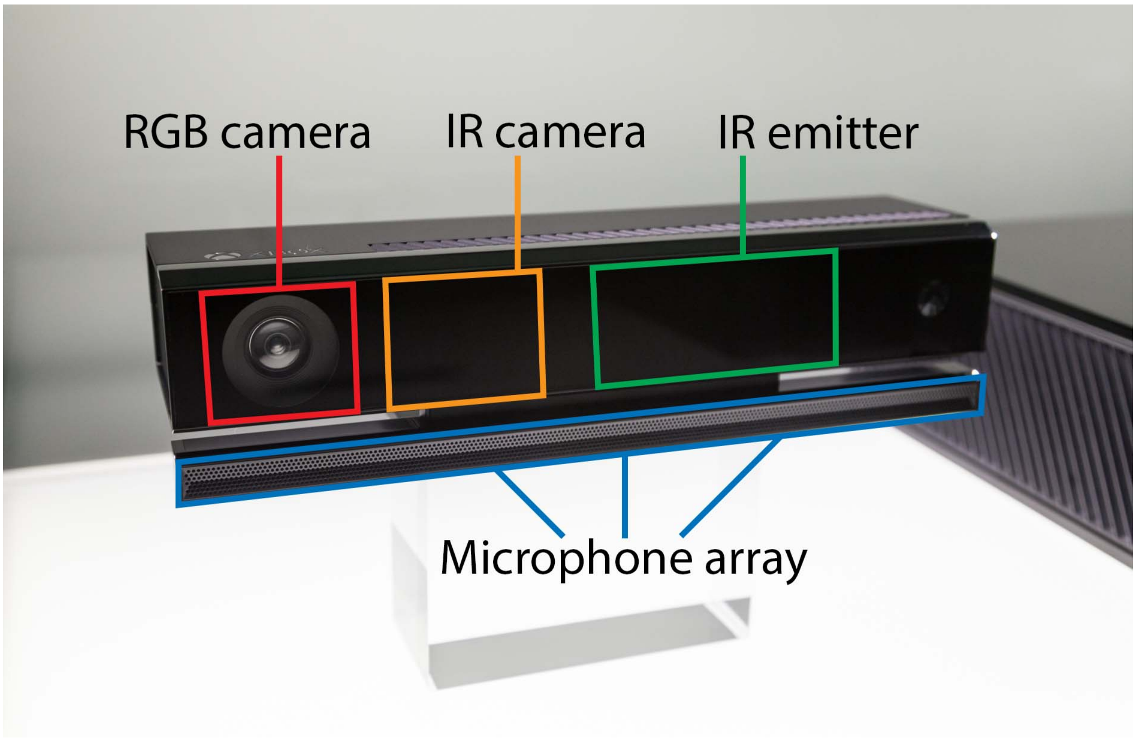


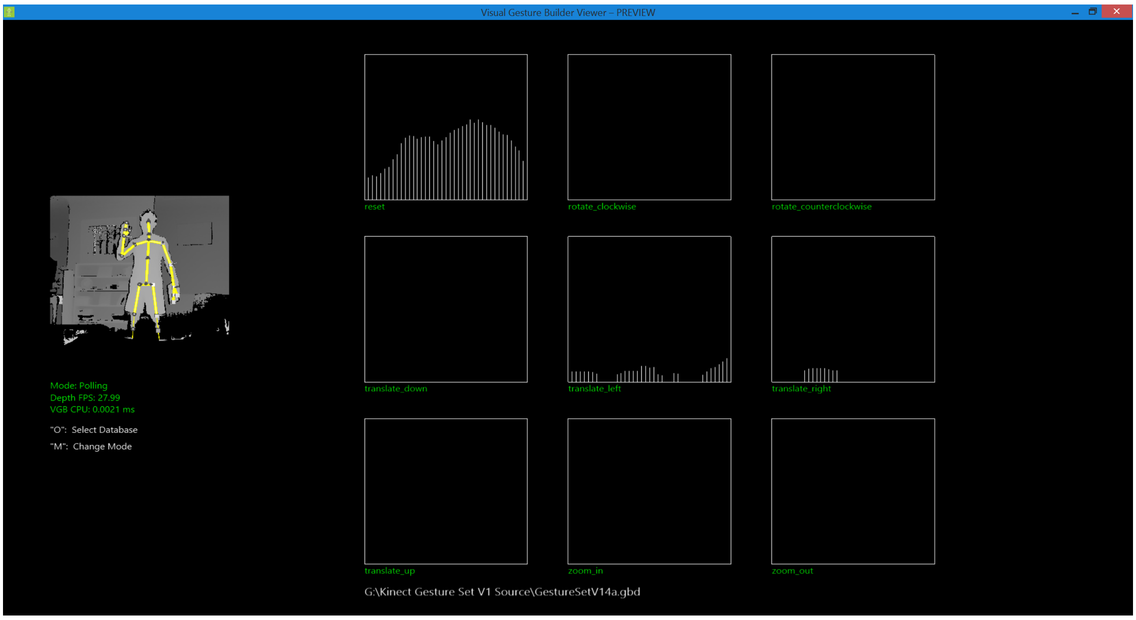
Figure 4.14: Anatomy of the the Kinect V2

Software

Visual Gesture Builder VGB (Figure A.4, p.214) is an application provided by Microsoft for tagging gesture training sets. At the time of conducting this study, it was available as a preview (or beta) application. It allows for the creation of a solution file, referred to as GestureSetV1 in this case, that can contain any number of gesture projects. Each gesture is inserted into the solution file as a separate project as a .vgbclip file. the VGB Gesture Wizard was used (Figure A.1, p.212) to create the gesture projects for the first prototype system.

Each gesture was populated with multiple .xef files containing training data for that gesture. This was performed in order to improve the robustness of gesture detection. Upon selecting a .xef file, the infrared detector feed and the 3D depth view, as well as any detected skeletons are displayed in the playback area. The researcher then reviewed each training set, tagging each time the gesture was performed.

All of the resultant training data was analysed and compiled into a Gesture Database file (.gdb) in order to test a set of gestures. This file could then be tested using the built-in VGB Viewer (Figure 4.15, p.100). All gestures in the .gdb file were displayed simultaneously by the application as a real-time history of recognition confidence values between zero and one displayed as a vertical bar. This application proved to be exceptionally useful. It allowed testing of whether or not a gesture was being recognised with high confidence values, and also provided an efficient means of visualising any overlap between gestures. This process was used during the early development stages to make gesture design decisions. A good example of this is the rotation gesture. Originally the rotation gesture involved moving both hands around each other. However, this was found to result in very low levels of recognition confidence. As a result, the gesture was redesigned to comprise of a raised fist with the palm of the other hand rotating like the hand of a clock. This gesture proved to be by far the most robust gesture for the system. In total, seventeen separate gesture databases were compiled during the development of the gesture set to support all required input tasks.



Mode: Polling
Depth FPS: 27.99
VGB CPU: 0.0021 ms

G:\Kinect Gesture Set V1 Source\GestureSetV14a.gbd

Figure 4.15: The Visual Gesture Builder Viewer provided an efficient means of testing the compiled gesture database, showing gesture recognition confidence values over time.

4.3.4 Evaluation

Publicly available medical images were chosen for inclusion in the prototype image viewer. These are useful repositories that facilitate testing using real-world imaging. The issue of patient privacy is eliminated during experimentation by using publicly available anonymised resources to create a test system. Pilot phase testing was performed informally with a small number of non-medical users. This allowed for some general observations and feedback to be gathered. The system was developed iteratively (reflecting Moran and Carroll's observation that design proceeds by iterative development [139]), with the most significant change to the touchless element of the system being the inclusion of voice commands to switch between different functions. This changed the touchless experience from a single modality to a multi-modal experience. Though this introduced an additional operational step for interactions, it afforded users greater control of the interface. This addition was especially important due to the relatively poor gesture recognition performance of the Kinect V2. Without the additional voice command, the system would suffer from significant levels of accidental gesture input, resulting in user frustration.

Users of the pilot system reported a significant difference in experience between inputs that predictably worked first time versus those they needed to try more than once. It was clear that emphasis should be given to using the most robust gestures possible when designing a touchless system in order to improve the user experience. Users also reported that they liked using direct input, such as measurement, where there was a one-to-one relationship between their gesture and the on-screen input. For continuous tasks, such as adjusting zoom, or increasing brightness, such one-to-one gestures may be preferable to give users an increased sense of agency. From this pilot study's results, when designing a TMIS it is important to establish a core set of robust interactions, focusing on delivering the most important functionality at a high level of reliability.

4.3.5 Discussion

The process of developing this first prototype touchless system provided several learnings, and revealed a number of challenges associated with the design process.

It was clear that certain interactions paired especially well with touchless control. An example of this was in performing the measurement of image features. Feedback from users indicated that using each hand as the endpoints of a virtual ruler was very effective. Two major contributing factors were identified for this positive user impression. The first was the appropriate metaphor of the hands as endpoints; users found this easy to understand and very intuitive. The second factor was the robust performance of the interaction; measurement was the most performant gesture input, which boosted user satisfaction through a sense of agency. It can be observed that touchless interactions need to combine a level of intuitiveness with a high level of performance.

Another insight from this initial investigation was the value of feedback to the user. Presenting the user with the system's "view" of the world allowed users to ensure they were in the correct position and could be "seen" by the system. By showing users the sensor's depth view it was possible to provide this feedback while avoiding any issues with the users feeling self-conscious.

The unfinished nature of VGB (as preview software) presented several technical challenges; issues such as project files breaking with no appropriate repair options meant that often gestures would need to be re-imported from scratch. While this is undesirable, nevertheless, these software platforms proved very useful in generating the prototype system.

Limitations

The Kinect V2 did not provide a sufficiently advanced hardware platform. As a result, it was found to suffer from performance limitations in terms of robust gesture detection. This significantly affected the choice of gestures used. Rather than choosing gestures that would feel most natural for the user, gestures were instead chosen for acceptable detection rates and minimal false-positive recognition. Technology should, as far as possible, support the design, rather than the design having to adapt to the technology.

4.4 Exploring Clutching and the HP Experience through a Second Prototype System

4.4.1 Motivation

The process of developing the first prototype provided valuable insights into approaching the design of a touchless system. A second system was designed and implemented using the Azure Kinect DK. This allowed learnings from existing literature, from the first prototype system, and feedback from HPs be integrated, along with significantly updated hardware and software tools to develop this experimental setup. The results of user testing using this second system are presented in Chapter 5 (p.115).

One significant challenge encountered with the first prototype was due to performance robustness using the Kinect V2. The aim of this research is to investigate three areas; understanding the user experience of HPs when using a touchless system, investigating multi-modal touchless interaction, and comparing several clutching modalities and their impact on the user experience. To support these aims, it was desirable, therefore, to improve robustness and suppress the user's awareness of hardware limitations as much as possible.

Following from the results of interviewing clinicians regarding their usage and requirements for PACS (presented in Chapter 3, p.45), it is clear that there is a strong interest in robust voice control of PACS. In order to support this requirement, the second prototype built on the initial software, providing improved performance and expanded voice control functionality. The recently released Microsoft Kinect Azure DK (released in 2019, Figure 4.21, p.110) is utilised, in combination with the computational capabilities of the Microsoft Azure Cloud. The Kinect Azure DK has much more robust skeleton tracking performance (due to higher quality sensors and updated software) (Figure 4.16, p.104), as well as significantly improved voice performance (due to having seven omnidirectional microphones and access to natural language processing through the Azure Cloud). It is hoped that by leveraging this more recent generation of hardware it will be possible to overcome a number of the performance limitations of the first prototype based on the Kinect V2 (which was released in 2012).

4.4.2 Design

Different goals were being pursued when designing the graphical user interfaces (GUI) for the two touchless systems. The goal for the first prototype was to investigate touchless interaction, focusing on gesture and voice interaction, in comparison to the mouse and keyboard. In contrast, the goal with the second system was to replicate the sense of using a real-world PACS and enable an investigation of touchless clutching specifically.

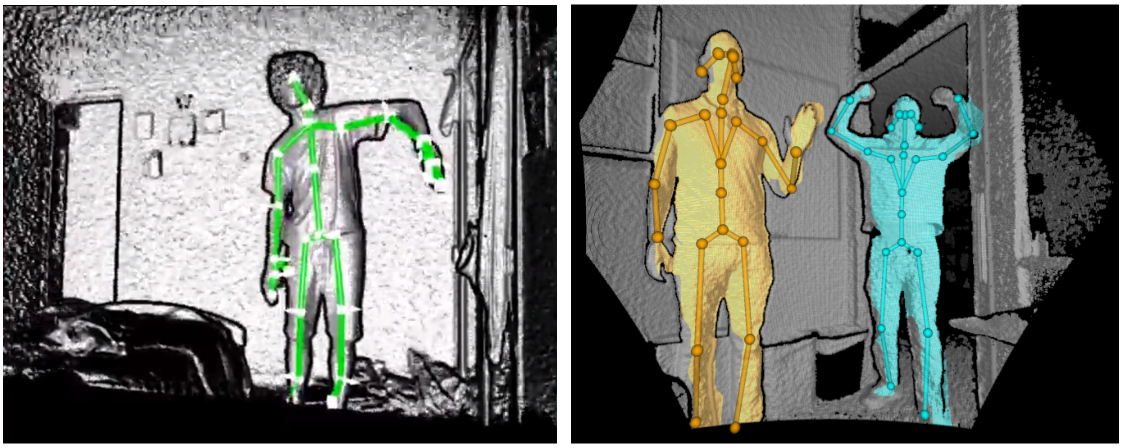


Figure 4.16: Left: Skeleton tracking with the Kinect V2,©2015 asCii88, Right: Skeleton tracking with the Azure Kinect DK, ©2021 Microsoft. The higher resolution sensors in the Azure Kinect DK allow for a considerably less noisy image, as well as superior skeleton tracking fidelity and body masking.

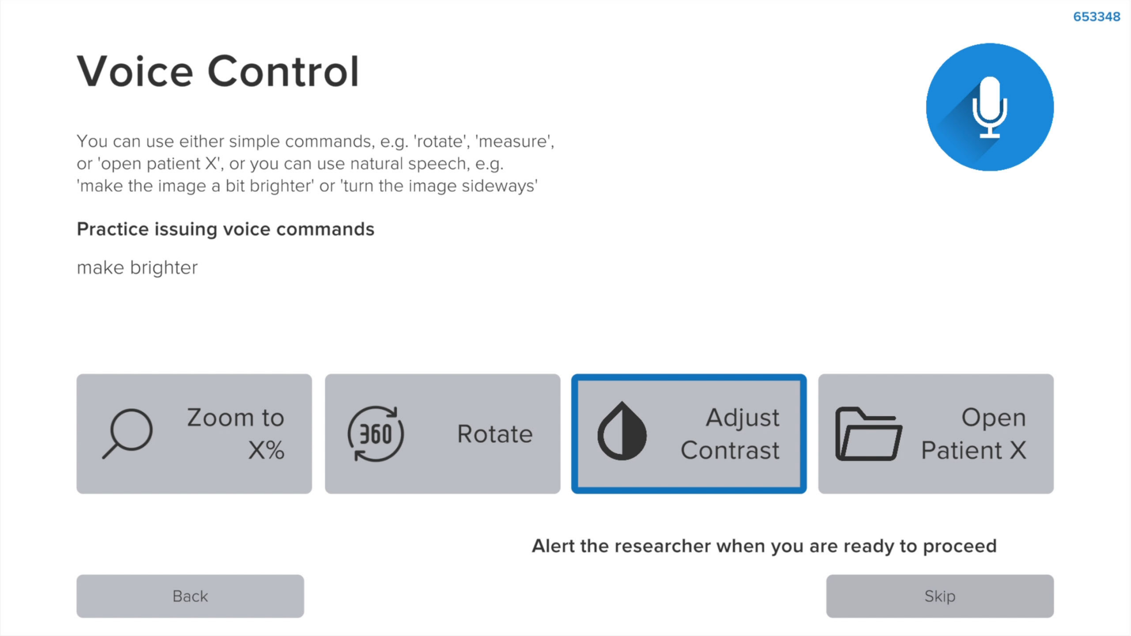
As a result of these differing goals, the design process for each system was quite different and resulted in significant differences between the two interfaces (Figure 4.12 (p.96) vs. Figure 4.18 (p.106)). This emphasis on creating a system that more accurately represented a real PACS required a greater knowledge of PACS, including key on-screen elements and an understanding of the file structure of DICOM files. Understanding which operational features are most commonly used by many HPs, such as magnification, windowing, and viewing reports was required also.

User interface

The design for the second prototype system considered several design goals; viz. integration of training content, support for experimentation with HPs, integration of PACS functionality, use of authentic DICOM files, and the inclusion of a reporting view.

User training content A training module was incorporated into the design of the user interface. This aimed to allow users to train to satisfaction on all novel aspects of the system, including touchless interaction, the clutching modalities, and the experimental GUI itself. As users would be using the system for a single study session, the decision was made to present all training materials in a single educational experience, both demonstrating functionality/touchless inputs and allowing the user to practice for themselves.

Figure 4.17 (p.105) shows the training interface for allowing users to practice voice control; in this view, voice commands would highlight the detected command as visual feedback to the user of a successful interaction. Furthermore, the detected speech is displayed to the user in real-time; this allows users better understand when commands did not work due to misrecognition.


653348

Voice Control

You can use either simple commands, e.g. 'rotate', 'measure', or 'open patient X', or you can use natural speech, e.g. 'make the image a bit brighter' or 'turn the image sideways'

Practice issuing voice commands

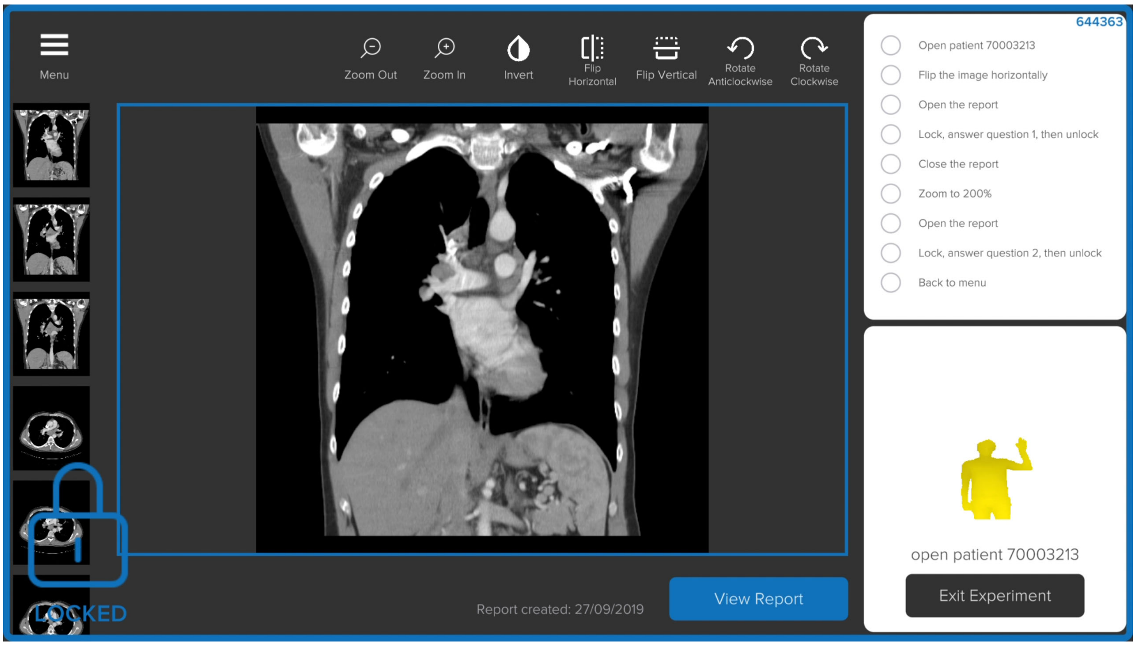
make brighter

Alert the researcher when you are ready to proceed

Figure 4.17: General touchless control user training covered the essentials of using the system with voice and gesture control.

PACS functionality In order to more closely mimic real-world HP experiences, several common PACS functionalities were included (Figure 4.18, p.106). These were chosen by reviewing published literature and reviewing HP feedback in Chapter 3 (p.45); this approach ensured that commonly used PACS features were included in the prototype.

DICOM files To further enhance the accuracy of the prototype system, support for DICOM files was included. DICOM files differ from regular imaging files; two significant differences include their extensive use of embedded patient metadata and their inclusion of multiple image “slices”. The Unity tool Simple DICOM Loader (Kompath) was used to support the loading of the DICOM library used in the user experiment presented in Chapter 5 (p.115).



644363

○ Open patient 70003213
○ Flip the image horizontally
○ Open the report
○ Lock, answer question 1, then unlock
○ Close the report
○ Zoom to 200%
○ Open the report
○ Lock, answer question 2, then unlock
○ Back to menu

open patient 70003213

Report created: 27/09/2019

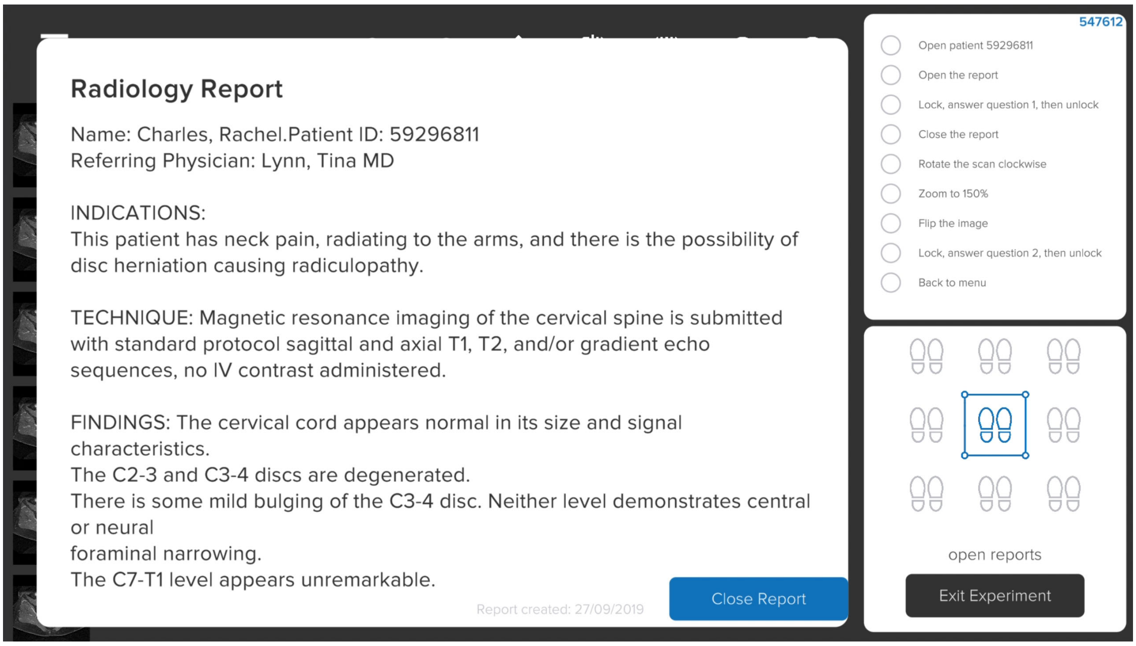
Figure 4.18: The second touchless interface, showing a gesture clutch event. Several standard PACS controls can be seen at the top of the interface. Image slices are shown along the left of the interface. The current image slice is shown in the main central area of the interface. Along the right of the screen are experimental elements including a task list and touchless interaction feedback.

Reporting view A finding from interviewing HPs was the significance of the radiologist's report for many HPs when reviewing patient scans with PACS. An interface was designed to allow HPs easily view reports associated with displayed images (Figure 4.19, p.107).

Reports were loaded dynamically for each displayed image. In order to manage the functional scope of the touchless system, reports were designed to be a single page in length to avoid the need to scroll vertically, or swipe between pages.

Touchless lexicon

In the second prototype system, a touchless lexicon was developed that incorporated four touchless modalities; gesture control, voice control, gaze clutching, and active zone clutching. This approach followed the existing frameworks proposed by researchers such as Karam.



Radiology Report

Name: Charles, Rachel.Patient ID: 59296811
Referring Physician: Lynn, Tina MD

INDICATIONS:
This patient has neck pain, radiating to the arms, and there is the possibility of disc herniation causing radiculopathy.

TECHNIQUE: Magnetic resonance imaging of the cervical spine is submitted with standard protocol sagittal and axial T1, T2, and/or gradient echo sequences, no IV contrast administered.

FINDINGS: The cervical cord appears normal in its size and signal characteristics.
The C2-3 and C3-4 discs are degenerated.
There is some mild bulging of the C3-4 disc. Neither level demonstrates central or neural
foraminal narrowing.
The C7-T1 level appears unremarkable.

Report created: 27/09/2019

547612

- ○ Open patient 59296811
- ○ Open the report
- ○ Lock, answer question 1, then unlock
- ○ Close the report
- ○ Rotate the scan clockwise
- ○ Zoom to 150%
- ○ Flip the image
- ○ Lock, answer question 2, then unlock
- ○ Back to menu

Figure 4.19: The report screen. Users answered off-screen questions based on the report.

Gestures In order to develop a gesture lexicon, numerous gestures were tested to determine their performance in the context of the prototype system. They were reviewed for successful detection rates, as well as for false positives, and false negatives. The decision was made to limit the gesture set in order to reduce the learning effort for HPs who would have limited time to become familiar with the system. As a result, a set of highly performant gestures was chosen, such as raising a hand above shoulder height for key experiment tasks, such as clutching and viewing reports. Gestures that were deemed less reliable, such as touching an elbow, were assigned to less frequently required tasks, such as inverting images.

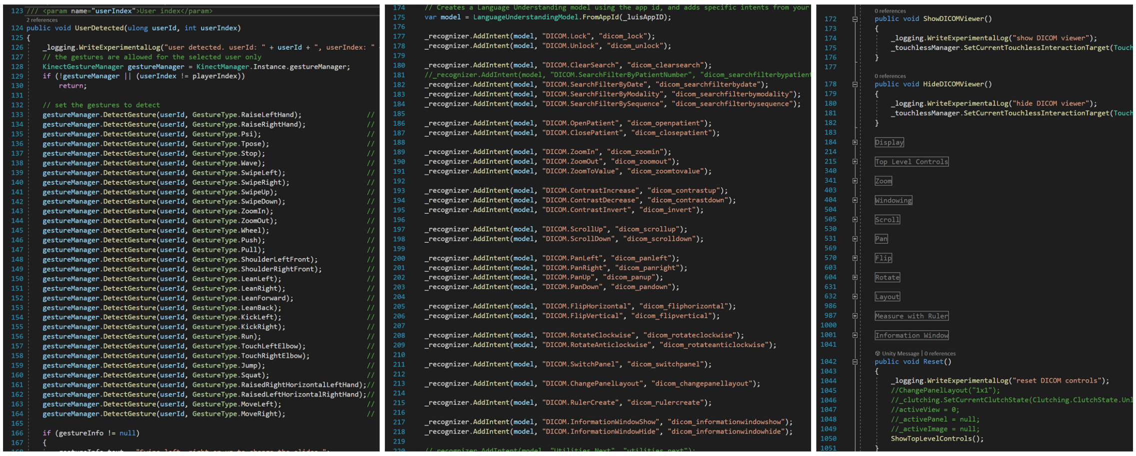
Active zone clutching A position-based active zone modality was designed to use a single rectangular active zone. To accommodate the various locations in which experimentation would be performed, the active zone could be configured in terms of length and width, as well as the location of the zone relative to the sensor. Tape was used to mark the area on the floor to immediately convey to the user where the system's active zone was located.

Clutching

A significant difference between the two prototype systems developed in this study was the inclusion of touchless clutching mechanisms in the second system. Though the first prototype included keywords to change modality, there was no clutching mechanic available to lock or unlock the system. Results obtained from the first prototype system, observing the challenges users faced, as well as feedback in HP interviews regarding their needs for a touchless system, served to highlight the importance of providing clutching. To this end, the second system focused on the implementation and testing of various clutching mechanisms.

Interactions/features

Multiple features and interactions were developed for the second prototype system (Figure 4.20, p.108). Features and interactions were tested for user experience and performance, e.g., gestures were tested for false–negative and false–positive recognition and critical functions were assigned the most robust gestures.



```
/// <param name="userIndex">User index</param>
2 references
public void UserDetected(ulong userId, int userIndex)
{
    _logging.WriteExperimentalLog("user detected. userId: " + userId + ", userIndex: "
    // the gestures are allowed for the selected user only
    KinectGestureManager gestureManager = KinectManager.Instance.gestureManager;
    if (!gestureManager || (userIndex != playerIndex))
        return;

    // set the gestures to detect
    gestureManager.DetectGesture(userId, GestureType.RaiseLeftHand);                  //
    gestureManager.DetectGesture(userId, GestureType.RaiseRightHand);                 //
    gestureManager.DetectGesture(userId, GestureType.Psi);                            //
    gestureManager.DetectGesture(userId, GestureType.Tpose);                          //
    gestureManager.DetectGesture(userId, GestureType.Stop);                           //
    gestureManager.DetectGesture(userId, GestureType.Wave);                           //
    gestureManager.DetectGesture(userId, GestureType.SwipeLeft);                      //
    gestureManager.DetectGesture(userId, GestureType.SwipeRight);                     //
    gestureManager.DetectGesture(userId, GestureType.SwipeUp);                        //
    gestureManager.DetectGesture(userId, GestureType.SwipeDown);                      //
    gestureManager.DetectGesture(userId, GestureType.ZoomIn);                         //
    gestureManager.DetectGesture(userId, GestureType.ZoomOut);                        //
    gestureManager.DetectGesture(userId, GestureType.Wheel);                          //
    gestureManager.DetectGesture(userId, GestureType.Push);                           //
    gestureManager.DetectGesture(userId, GestureType.Pull);                           //
    gestureManager.DetectGesture(userId, GestureType.ShoulderLeftFront);              //
    gestureManager.DetectGesture(userId, GestureType.ShoulderRightFront);             //
    gestureManager.DetectGesture(userId, GestureType.LeanLeft);                       //
    gestureManager.DetectGesture(userId, GestureType.LeanRight);                      //
    gestureManager.DetectGesture(userId, GestureType.LeanForward);                    //
    gestureManager.DetectGesture(userId, GestureType.LeanBack);                       //
    gestureManager.DetectGesture(userId, GestureType.KickLeft);                       //
    gestureManager.DetectGesture(userId, GestureType.KickRight);                      //
    gestureManager.DetectGesture(userId, GestureType.Run);                            //
    gestureManager.DetectGesture(userId, GestureType.TouchLeftElbow);                 //
    gestureManager.DetectGesture(userId, GestureType.TouchRightElbow);                //
    gestureManager.DetectGesture(userId, GestureType.Jump);                           //
    gestureManager.DetectGesture(userId, GestureType.Squat);                          //
    gestureManager.DetectGesture(userId, GestureType.RaisedRightHorizontalLeftHand);//
    gestureManager.DetectGesture(userId, GestureType.RaisedLeftHorizontalRightHand);//
    gestureManager.DetectGesture(userId, GestureType.MoveLeft);                       //
    gestureManager.DetectGesture(userId, GestureType.MoveRight);                      //

    if (gestureInfo != null)
    {
```

```
    // Creates a Language Understanding model using the app id, and adds specific intents from your
    var model = LanguageUnderstandingModel.FromAppId(_luisAppID);

    _recognizer.AddIntent(model, "DICOM.Lock", "dicom_lock");
    _recognizer.AddIntent(model, "DICOM.Unlock", "dicom_unlock");

    _recognizer.AddIntent(model, "DICOM.ClearSearch", "dicom_clearsearch");
    //_recognizer.AddIntent(model, "DICOM.SearchFilterByPatientNumber", "dicom_searchfilterbypatient
    _recognizer.AddIntent(model, "DICOM.SearchFilterByDate", "dicom_searchfilterbydate");
    _recognizer.AddIntent(model, "DICOM.SearchFilterByModality", "dicom_searchfilterbymodality");
    _recognizer.AddIntent(model, "DICOM.SearchFilterBySequence", "dicom_searchfilterbysequence");

    _recognizer.AddIntent(model, "DICOM.OpenPatient", "dicom_openpatient");
    _recognizer.AddIntent(model, "DICOM.ClosePatient", "dicom_closepatient");

    _recognizer.AddIntent(model, "DICOM.ZoomIn", "dicom_zoomin");
    _recognizer.AddIntent(model, "DICOM.ZoomOut", "dicom_zoomout");
    _recognizer.AddIntent(model, "DICOM.ZoomToValue", "dicom_zoomtovalue");

    _recognizer.AddIntent(model, "DICOM.ContrastIncrease", "dicom_contrastup");
    _recognizer.AddIntent(model, "DICOM.ContrastDecrease", "dicom_contrastdown");
    _recognizer.AddIntent(model, "DICOM.ContrastInvert", "dicom_invert");

    _recognizer.AddIntent(model, "DICOM.ScrollUp", "dicom_scrollup");
    _recognizer.AddIntent(model, "DICOM.ScrollDown", "dicom_scrolldown");

    _recognizer.AddIntent(model, "DICOM.PanLeft", "dicom_panleft");
    _recognizer.AddIntent(model, "DICOM.PanRight", "dicom_panright");
    _recognizer.AddIntent(model, "DICOM.PanUp", "dicom_panup");
    _recognizer.AddIntent(model, "DICOM.PanDown", "dicom_pandown");

    _recognizer.AddIntent(model, "DICOM.FlipHorizontal", "dicom_fliphorizontal");
    _recognizer.AddIntent(model, "DICOM.FlipVertical", "dicom_flipvertical");

    _recognizer.AddIntent(model, "DICOM.RotateClockwise", "dicom_rotateclockwise");
    _recognizer.AddIntent(model, "DICOM.RotateAnticlockwise", "dicom_rotateanticlockwise");

    _recognizer.AddIntent(model, "DICOM.SwitchPanel", "dicom_switchpanel");

    _recognizer.AddIntent(model, "DICOM.ChangePanelLayout", "dicom_changepanellayout");

    _recognizer.AddIntent(model, "DICOM.RulerCreate", "dicom_rulercreate");

    _recognizer.AddIntent(model, "DICOM.InformationWindowShow", "dicom_informationwindowshow");
    _recognizer.AddIntent(model, "DICOM.InformationWindowHide", "dicom_informationwindowhide");
```

```
0 references
public void ShowDICOMViewer()
{
    _logging.WriteExperimentalLog("show DICOM viewer");
    _touchlessManager.SetCurrentTouchlessInteractionTarget(Touch
}

0 references
public void HideDICOMViewer()
{
    _logging.WriteExperimentalLog("hide DICOM viewer");
    _touchlessManager.SetCurrentTouchlessInteractionTarget(Touch
}

Display

Top Level Controls

Zoom

Windowing

Scroll

Pan

Flip

Rotate

Layout

Measure with Ruler

Information Window

Unity Message | 0 references
public void Reset()
{
    _logging.WriteExperimentalLog("reset DICOM controls");
    //ChangePanelLayout("1x1");
    //_clutching.SetCurrentClutchState(Clutching.ClutchState.Unl
    //activeView = 0;
    //_activePanel = null;
    //_activeImage = null;
    ShowTopLevelControls();
}
```

Figure 4.20: Left: though the system could identify thirty-two distinct gestures, both static and dynamic, only a small set of the most reliable gestures were used during user experiments. Middle: an intent that correlated to each voice command trained using LUIS.ai was added to an intent recognizer. Right: In order to manage the complexity of the system, various support functions were not required during experiments, e.g., changing panel layout.

In order to focus the system on robust clutching performance, which was the focus of user experiments, the overall feature and interaction set was reduced to ensure robust system performance. The set of features and interactions supported in the experiment are shown in Table 4.4 (p.109).

Feature	**Supported interactions**
Open patient	Voice command, e.g. 'open patient X', or 'open X'
Zoom in/out	Voice command, e.g. 'zoom to X', or 'magnify to X percent'
Window/invert	Voice command, e.g. 'invert image', or 'invert' Gesture command, i.e., touch right elbow
Flip vertical/horizontal	Voice command, e.g. 'flip the image', or 'flip vertical'
Rotate clockwise/anticlockwise	Voice command, e.g. 'rotate clockwise', or 'turn the image sideways'
Open/close report	Voice command, e.g. 'open report', or 'close report' Gesture command (raise left hand above shoulder height)
Lock/unlock	Voice command, e.g. 'lock', 'pause it there', 'unlock', or 'let's go' Gesture command (right hand raised above shoulder height to toggle clutch state) Gaze (head angle no greater than 30°from the display) Active zone
Exit to menu	Voice command, e.g. 'exit to menu', or 'back to menu'

Table 4.4: Features/interactions supported in the prototype system. Though multiple additional input gestures were developed, e.g., continuous window adjustment using hand swipes, a reduced set was included in the final system to prioritize performance.

4.4.3 Implementation

Hardware

The Azure Kinect DK was selected for the touchless sensor. This represented a significant improvement from the Kinect V2 used in the first prototype. Though the Azure Kinect DK requires a more powerful PC to run, it delivers robust performance and functionality not available with the Kinect V2. The Azure Kinect DK delivers higher resolution in both RGB and depth-sensing cameras. This increased resolution, combined with updated software from Microsoft allows for more reliable skeleton tracking. This feature enabled more robust gesture detection and was a primary reason for changing to the newer sensor given the limitations of the Kinect V2.

The inclusion of an omnidirectional microphone array (as compared to the Kinect V2's front-facing stereo cameras), combined with Language Understanding (LUIS) enabled modern natural language processing, would allow the use of either discrete commands or continuous natural language.

The Azure Kinect DK device is considerably smaller than previous generations of Kinect (though it remains larger than alternative sensors such as Intel's RealSense cameras). This smaller size allows deployment in areas where larger sensors may be a limitation, such as in the context of the radiologist's office/desk.

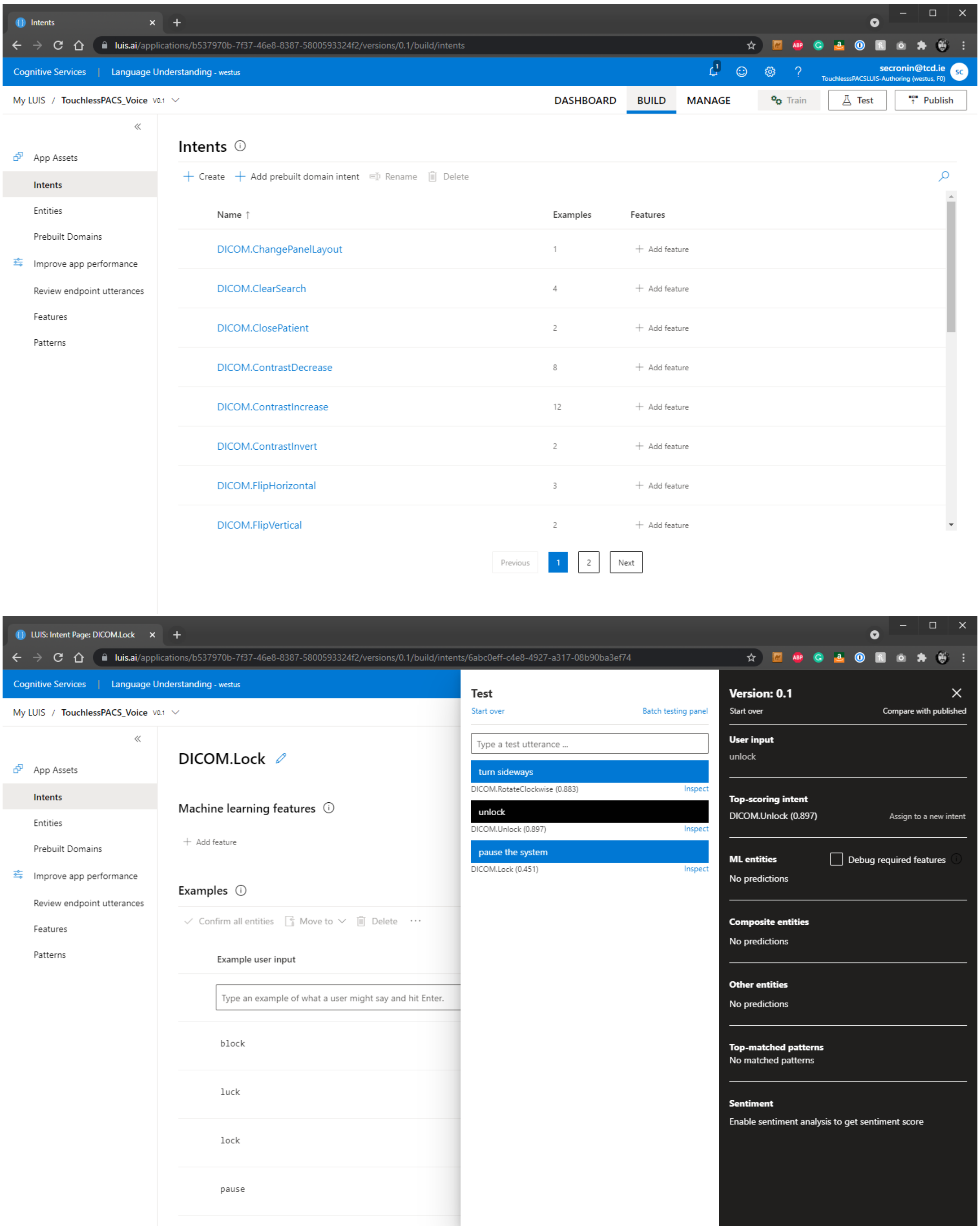
Figure 4.21: The Azure Kinect DK.

Software

Microsoft's LUIS cloud platform was integrated into the prototype system design to provide natural language processing. User commands were converted from speech-to-text locally before being sent to LUIS for processing. If confidence was achieved that a particular intent was selected, LUIS returned that intent, along with metadata such as the confidence value. An intent was developed for each function (Figure 4.22 (top), p.111). Intents were made up of exemplar user phrases, such as "lock the system", "pause". This process was performed iteratively, incorporating additional user phrases during the development process.

Each time intents were updated, the LUIS model was trained, tested (Figure 4.22 (bottom), p.111), and published. Once published, improvements to voice performance were immediately available to the prototype system. This is useful in a scenario where multiple systems were connected to LUIS, as they would all have immediate access to any published changes.

Microsoft Cognitive Services (MCS) was used to enable face and gaze detection. Figure 4.23 (p.112) shows the web interface for managing Azure Cloud resources, such as Cognitive Services. Colour frames were provided to MCS for processing, along with specifications for which features to report, such as head/face details, age, and gender. These



Intents

Name ↑	Examples	Features
DICOM.ChangePanelLayout	1	+ Add feature
DICOM.ClearSearch	4	+ Add feature
DICOM.ClosePatient	2	+ Add feature
DICOM.ContrastDecrease	8	+ Add feature
DICOM.ContrastIncrease	12	+ Add feature
DICOM.ContrastInvert	2	+ Add feature
DICOM.FlipHorizontal	3	+ Add feature
DICOM.FlipVertical	2	+ Add feature

DICOM.Lock

Machine learning features

Examples

Example user input

Type an example of what a user might say and hit Enter.

block

luck

lock

pause

Test

Type a test utterance ...

turn sideways
DICOM.RotateClockwise (0.883)

unlock
DICOM.Unlock (0.897)

pause the system
DICOM.Lock (0.451)

Version: 0.1

User input
unlock

Top-scoring intent
DICOM.Unlock (0.897)

ML entities ☐ Debug required features
No predictions

Composite entities
No predictions

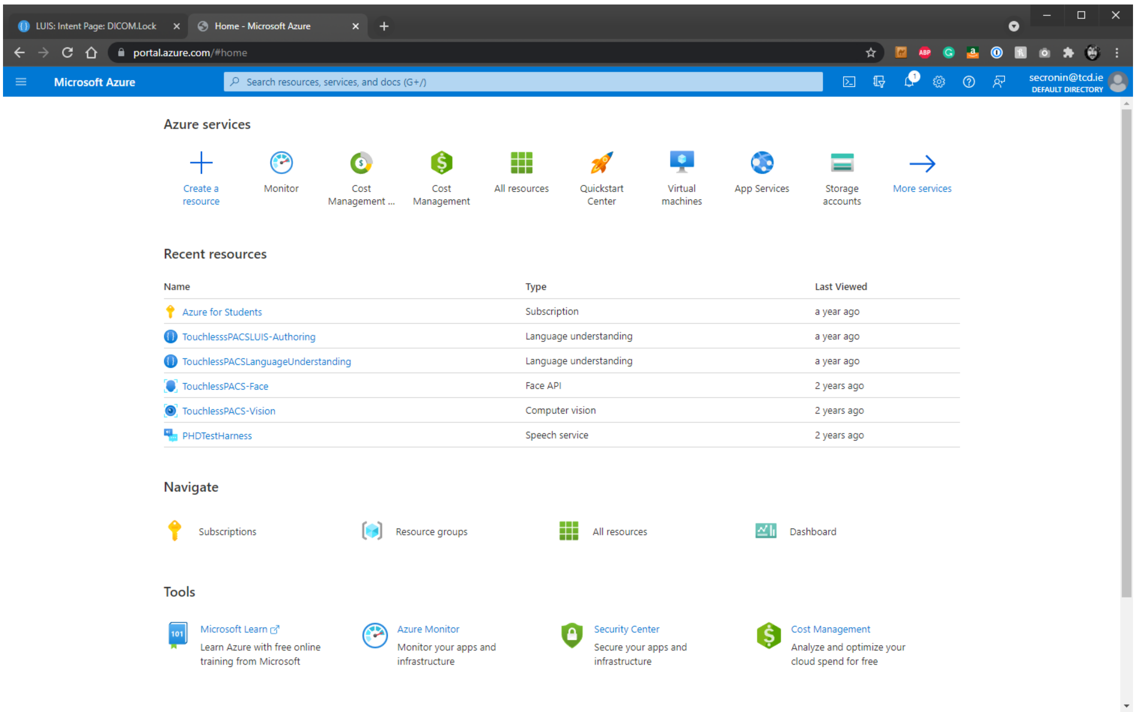
Other entities
No predictions

Top-matched patterns
No matched patterns

Sentiment
Enable sentiment analysis to get sentiment score

Figure 4.22: Top: The LUIS interface showing a list of intents developed for the experimental system. Each intent is linked to a single command, e.g., "DICOM.OpenPatient". Bottom: Intents and their user inputs were developed using an iterative process. User input examples were expanded during testing to incorporate exemplar user command phrases. LUIS reports which intent is detected, along with values such as confidence level. Different inputs can be tested before publishing a LUIS build.

features were returned in the form of text-based metadata.



Name	Type	Last Viewed
Azure for Students	Subscription	a year ago
TouchlesssPACSLUIS-Authoring	Language understanding	a year ago
TouchlessPACSLanguageUnderstanding	Language understanding	a year ago
TouchlessPACS-Face	Face API	2 years ago
TouchlessPACS-Vision	Computer vision	2 years ago
PHDTestHarness	Speech service	2 years ago

Microsoft Learn: Learn Azure with free online training from Microsoft

Azure Monitor: Monitor your apps and infrastructure

Security Center: Secure your apps and infrastructure

Cost Management: Analyze and optimize your cloud spend for free

Figure 4.23: The Microsoft Azure Portal shows which Azure resources have been created. These resources can be interdependent, e.g. face recognition relies on computer vision.

4.4.4 Evaluation

Gentile *et al.* contend that it is essential that the study and testing of touchless gestures should be performed under real-world conditions, in appropriate social contexts, and with authentic users [140] as this can lead to better outcomes [141]. This requirement for authentic environments is made clear by Mewes *et al.* [1], who highlights the difference in voice recognition rates achieved by Nathan *et al.* [70] and Hötker *et al.* [71] under lab conditions than Alapetite [47] or Perrakis *et al.* [50] in a real OR environment. Relying on controlled environment results, risks failing to capture the "actual" system performance and user experience.

Evaluation of the second prototype is presented in Chapter 5 (p.115).

4.4.5 Discussion

The process of developing two separate touchless systems provided the opportunity to gain useful insights that would not have been possible having developed a single system. This is due to several key differences between the two systems, both in terms of their objectives and their implementations. These insights are a function of the technology and development process rather than of user testing, which is covered in Chapter 5 (p.115).

A major difference between the two systems was that of hardware choice, with the Azure Kinect DK improving on the performance of the Kinect V2 in all meaningful areas. Two particular functional differences between the sensors had the greatest impact on the design process; skeleton tracking performance, and voice recognition performance. The difference in performance was due to a combination of significant improvements to the sensor hardware, such as the camera resolution, onboard processing power, an omnidirectional microphone array, as well as improvements in the supporting software, such as an improved body tracking SDK, and availability of the Microsoft Azure-powered facial recognition and natural language processing services. The improvement in body-tracking resulted in a significant improvement in gesture detection performance. This made it possible to design a gesture lexicon that did not rely on keyword activation. This reduced the number of operational steps required for a user to perform a gesture command while maintaining system performance. The older Kinect V2 included a linear array of four front-facing microphones. These enabled the use of a local dictionary of keywords for voice control. In contrast, the inclusion of an omnidirectional array of microphones in the Azure Kinect DK, as well as easily integrated, cloud-based natural language cloud processing through LUIS (Figure 4.22, p.111), allowed for a more advanced implementation of voice control.

4.5 Conclusions

In this chapter, the process of developing two prototype touchless systems, as well as the challenges associated with the associated design process is presented. Developing prototype systems using different touchless sensors (Kinect V2 vs. Azure Kinect DK) brought insights to the impact of sensor selection on the design and development of TMIS. Work presented in Chapter 2 (p.11) and Chapter 3 (p.45) revealed the need to investigate the design of the touchless lexicon, user feedback, and clutching mechanics (to prevent unintended touchless input); these issues were explored in this chapter, integrating the design frameworks presented by Nguyen and Karam [129, 128]. This work can help to inform the development of future TMIS, especially regarding the design of TMIS that rely on input modalities such as gesture or voice control.

Chapter 5 (p.115) explores HP experiences with a prototype TMIS. This will explore the impact of design decisions described in this chapter, especially those addressing the issues of clutching, and touchless lexicon selection and implementation.

Chapter 6 (p.149) goes on to characterise the design process of systems for touchless navigation of medical imaging, building on insights and requirements emerging from the investigation of the design and implementation process for TMIS presented in this chapter. This codifies the learnings from the development process into a framework for future developers and researchers to use.

Chapter 5

Prototype study: An investigation of clutching mechanisms for touchless navigation of medical imaging systems

In this chapter, HP experiences with using a prototype touchless system are investigated, aiming to build on the insights gained through the qualitative analysis presented in Chapter 3 (p.45). When interviewed, HPs expressed a desire for greater levels of touchless interaction, especially in a post-COVID-19 workplace. However, the generally positive attitude towards touchless interaction, and the use of different interaction modalities might change following the experience of using such a system, and it would furthermore be important to explore the nature of this experience to inform the development of future systems. The problem of unintended input, and the variety of possible mechanisms to address it have also been identified as an important aspect of the design and user experience. Thus, this investigation aimed to explore HP's user experience with a prototype TMIS with a particular focus on clutching.

5.1 Introduction

5.1.1 Motivation

The "Midas Touch" problem is an important usability concern for touchless user interfaces. In this problem, a system mistakenly treats a user's unrelated actions as intent to interact, i.e., false-positive recognition [142]. As discussed in Chapter 2 (p.11), sensing systems often have a large but ambiguous input range [143, 144] and their always-on nature can lead to incidental actions (e.g., speech as part of a conversation, or hand movements as part of another activity) being mistakenly acted upon. Such unintentional input can cause user frustration and operational error that may require additional actions to be taken to undo its effects [130].

For medical imaging systems, such unintentional input could be disruptive and may even go unnoticed, since the HP's attention is often focused on their patients, and on other tasks. Hand gestures are frequently used by health professionals to communicate during medical procedures, and so being able to distinguish between interactive and non-interactive gestures is crucial [16, 136, 17]. A specific focus for this investigation is the use of four clutching techniques in the context of a touchless PACS interface with speech and gesture input. Emphasis is placed on a comparison of four clutch modalities for a touchless PACS interface: a (un)lock gesture, a (un)lock speech phrase, gaze, and active zone.

This study explores these clutching mechanisms with an ecologically valid sample of hospital-based healthcare professionals (n=34). Quantitative interaction measures and qualitative analysis of interviews with the clinicians provides rich insight into the usability of these clutching methods and their suitability for clinical use. The findings presented in this dissertation also identify benefits, challenges, and limitations associated with integrating touchless interaction into medical settings. This is the first detailed investigation of clutching in this context, and represents a contribution towards the deployment of usable and effective touchless PACS interfaces, which have the potential to transform digital clinical workflows.

5.1.2 Contributions

The following contributions are made in this chapter:

- **Comparison of four clutch mechanisms** Perform a direct comparison study of four clutch mechanisms for a touchless PACS interface: an (un)lock gesture, an (un)lock speech phrase, (un)lock gaze, and (un)lock active zone. These were chosen from the many techniques described in the literature because their characteristics suggest they could be appropriate for this usage context. Whilst prior work has highlighted the importance of clutching to reduce unintended input, the work described in this chapter is the first focused investigation of clutching interactions for touchless medical systems. This work seeks to address this gap in the literature, by providing insight into how clutching affects the user experience of a touchless PACS interface.
- **Detailed qualitative exploration** The work further contributes a detailed qualitative exploration of the challenges and opportunities for deploying touchless interaction beyond PACS, across other interactive systems in medical usage contexts

5.2 Background

Touchless interaction modalities like speech and mid-air gestures are compelling for use in medical settings because they can allow health professionals to provide direct input without the sterility concerns of surface contact. Sterility is crucial to infection control in medical settings, both for protecting patients who are often vulnerable to HAIs and for protecting the well-being of health professionals (most recently highlighted by the COVID-19 pandemic). Touchless modalities allow users to interact directly and immediately without having to break asepsis, issue commands to another person, or utilise the many workarounds that have been developed for interacting with computers in clinical contexts [14, 15, 16, 17]. Many prototypes have shown the potential benefits of touchless input in medical settings [1], e.g., using mid-air gestures [136, 27, 57, 145, 19, 135, 34, 75], foot input [34, 75], voice [146, 27, 75], and proxemics [17]. A common finding in this corpus is that health professionals recognise the potential benefits of touchless input (especially with respect to sterility and direct control over imaging systems). However, good usability is important for such benefits to translate to real working environments and published literature highlight the challenges of deploying touchless clinical systems in a usable and effective way.

Unintentional input is a key usability concern with touchless interaction because touchless sensing systems are ‘always on’ and attempting to recognise intentional input within a large room-scale space [143]. Unintentional input can be especially problematic in medical usage contexts, because teams of health professionals work in close physical proximity and frequently gesture and talk to each other during clinical work [16, 136, 17, 15]. Being able to differentiate between interactive and non-interactive gestures and sentences is therefore key for reducing unintentional input.

Clutching mechanisms may require purposeful actions that act as a ‘mode switch’ or show a person’s intent. Alternatively, systems may infer intent from a set of constraints and contextual cues (e.g., position in the room or eye contact with the display). These actions and cues allow a sensing system to identify input data that corresponds to intentional user input.

Voice clutches such as ’lock’ can be used in conjunction with other interaction modalities for mode switching, since it does not interfere with manual actions, such as mid-air gestures. More complex voice clutches can be integrated into multimodal interactions and used alongside other input methods; for example, Put-That-There [147] used speech and gesture to identify operations and their parameters. Speech commands acted as a clutch, since a gesture without an accompanying utterance, or an utterance without a gesture, would not be treated as input.

Voice clutches were first considered for novel medical imaging systems in the early 1990s. Hinckley et al. [148] investigated the use of tangible props for interacting with neurosurgical visualisations. They used buttons and pedals for clutching, to avoid unintended effects when props were picked up or placed down. A voice clutch was considered and discarded due to the performance limitations of speech recognition technology at the time. They hypothesised that speaking the clutch phrases (*“move [prop]”* and *“stop* *[prop]"*) would introduce a frustrating delay and might distract from other tasks. However, these issues have not prevented the uptake of speech in touchless user interfaces in hospitals [14] and contemporary speech recognition systems would be less affected by such latency. Mid-air gesture systems can likewise use clutch mechanisms to avoid unintentional input. Gestural equivalents of wake words can be used to indicate the beginning of a gesture command sequence. Like wake words, these gestures should be unlikely to occur incidentally. Clutch gestures can be discrete movements or poses that act as a mode switch prior to performing other gestures; e.g., finger snapping [149]. Alternatively, a clutch pose can be held continuously as part of another action; e.g., the hand-on-hip 'teapot' pose while performing gestures with the free hand [150], a finger-to-thumb pinch gesture while moving the hand [151], or an extended thumb while pointing with the index finger [152]. Suitably chosen gestures should reduce the chance of accidental recognition, but confidence can be further increased through the use of a dwell period, where a clutch gesture is held for a brief period.

For clinical contexts, it is important to choose gesture clutch actions that will not occur during other activities or interactions with other health professionals, who often gesticulate to each other [136, 16, 17]. A further constraint on the choice of gesture is introduced by the need for sterility, which restricts where hands can safely move or be placed during the interaction. O'Hara et al. [136] discuss other examples of clutch gestures that were found to be unsuitable in surgical contexts. For example, dwell time was deemed unsuitable because health professionals would often pause for reflection while viewing images.

Finally, the intent to interact can also be inferred by how or where users perform actions. Baudel et al. [133] described the use of an 'active zone', which is an area of space where sensed movements are treated as an intentional input. This can limit the available input space significantly by excluding input in other regions and, if clearly delineated, can help users understand where input will be sensed [153]. Alternatively, information about body posture and gaze can be used to infer an intention to interact [134]. For example, Jacob et al. [135] used information from head and body orientation to determine when a user was intending to interact with a MRI system. While this reduced the false-positive gesture rate in their study, O'Hara et al. [136] suggested those contextual posture cues may be misleading in other usage scenarios (e.g., when looking at images with other HPs).

5.3 User Study

5.3.1 Study Aims

A mixed methods user study was conducted to explore the user experience of touchless control of medical imaging, and to evaluate clutching methods for a touchless PACS interface, with an ecologically valid sample comprising a significant cohort of practising hospital clinicians. HPs with PACS experience were asked to complete a series of tasks using a generic touchless PACS interface, exploring different clutching techniques to lock/unlock the touchless user interface. The aim was to investigate how clutching affects user experience and to better understand the process and challenges of integrating touchless input into PACS workflows. This study sought to address the following research questions:

- **RQ1:** How does the use of clutching affect the user experience of interacting with a touchless PACS interface?
- **RQ2:** How can clutching be effectively integrated into a touchless interface in a clinical setting?
- **RQ3:** What are the operational differences between several commonly used clutching methods for a touchless PACS interface?
- **RQ4:** What are the opportunities, challenges, and limitations associated with touchless PACS interaction?

PACS are often used in challenging operational contexts within hospitals where users frequently have to manage multiple competing demands on their attention. System designers need to better understand how clutching affects the user experience (**RQ1**), so that cognitive demand can be minimized and avoid disrupting the medical tasks that occur in conjunction with PACS usage. By comparing a variety of clutching methods, it will be possible to make informed decisions regarding how to integrate these into a touchless

PACS interface in particular and similar applications for clinical usage contexts in general (**RQ2**).

The clutching methods chosen for this study include explicit and implicit interactions, where there is a potential trade-off between user control (favouring explicit) and interaction demands (favouring implicit); this trade-off is explored, to see which methods are most appropriate for this context (**RQ3**). The resultant findings will provide usable insights for researchers and designers creating future touchless PACS interfaces and touchless interfaces for health professionals (**RQ4**).

5.3.2 Study Design and Procedure

This study used a within-subjects design with four clutch conditions: (1) gesture, (2) speech, (3) gaze, and (4) active zone. These clutch techniques were designed to support both explicit (e.g., gesture and speech where the user performs an affirmative action) and implicit interactions (e.g., inferring intent from body position in the active zone or from eye contact with the PACS interface).

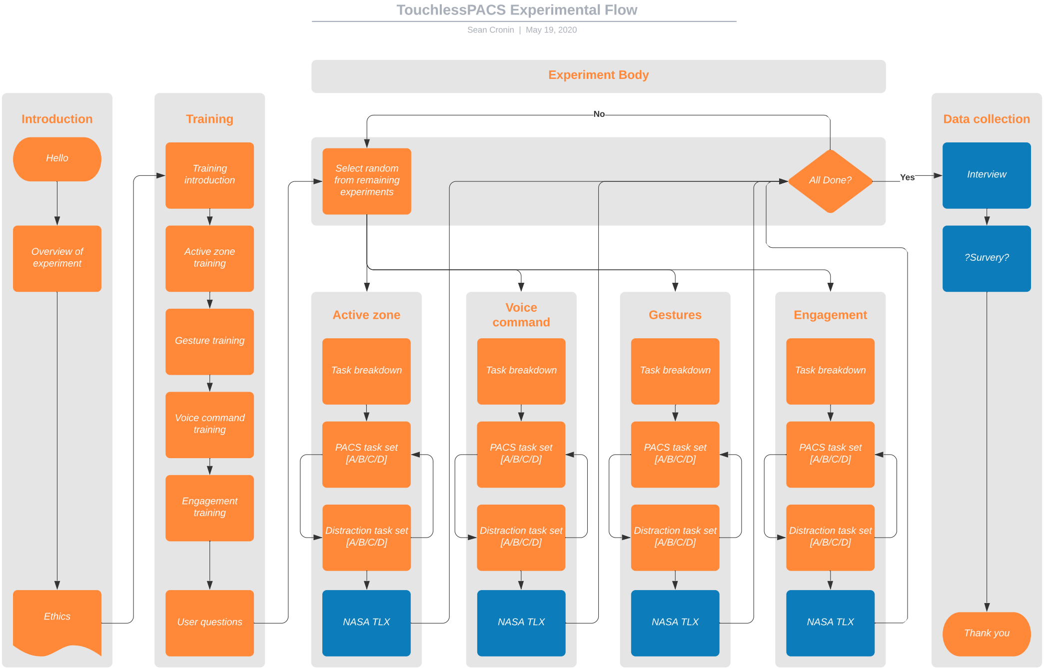


Figure 5.1: Experimental flow design

Participants were required to complete tasks using a custom PACS interface implementation (detailed in Chapter 4, p.75). A custom TMIS was created so that it was possible to integrate all clutch interaction modalities and minimise potential bias as a result of user familiarity with existing PACS software. Frequently used PACS tasks (e.g., opening patient files, image navigation and manipulation, viewing patient reports) that required sustained attention and interaction were selected for the study. The system was placed on a table with a 1 m^2 active zone marked on the floor, 1.5 m in front of the display while conducting the studies with the healthcare participants.

Participants were required to complete a set of off-screen secondary tasks while carrying out the PACS tasks. These tasks were performed in front of the input sensors, necessitating the use of the clutch interactions to avoid false-positive input recognition. Divided attention and multi-tasking scenarios like these are representative of real situations where PACS are used; e.g., discussing images and reports with colleagues, referencing other paperwork and materials, attending to patients, etc. The participants had to step away from the custom PACS to perform the secondary tasks which required them to read a question about an on-screen report, then write down the answer before returning to using the system. Users were allowed to consult the report when answering the off-screen questions, so often had to look at and/or use the system to navigate the interface. This element of the study design was intended to emulate a PACS user switching attention between a PACS display and attending to a patient or discussing images with colleagues.

At the start of each session, participants were trained in using the touchless PACS interface. Video tutorials demonstrated how to perform interactions and participants could practice using the system for as long as they required prior to starting each study. This researcher was present in the room to answer questions and assist during the training phase.

Each condition (i.e., clutch technique) consisted of one block of tasks and condition order was counterbalanced. Each block consisted of fourteen on-screen PACS tasks and off-screen secondary tasks. After each block was completed, participants were asked to complete the NASA-TLX survey to establish a measure of task workload [154]. At the end of the session, a semi-structured interview covering topics related to touchless interaction, the clutching methods, and general PACS usage was completed.

Research ethics approval was obtained from the relevant institutional ethics review committee (Figure B.1, Figure B.2, Appendix G, Appendix H, Appendix I, Appendix J)

5.3.3 Measurements

Each individual session was video-recorded and interaction data was logged by the touchless PACS system. Interaction logs were manually annotated by this researcher using the videos for reference. This ensured that complete logs were compiled with all verifiable user actions, even if the system did not detect them. This enabled a more complete analysis of interaction, e.g., identifying when users performed an incorrect action, 'correct' input action not being recognised, etc.

The total interaction time for each block of tasks was measured. Timing started when the user first attempted to perform the task and ended when the final interaction task was complete. Within this period, the cumulative time for the system in the unlocked state during each block was measured, as this could suggest if a clutch method was at risk of false-positive recognition (e.g., if the system is actively sensing input more often). All clutch interaction events and timestamps were logged, so that the number of transitions between clutch states (i.e., locked/unlocked) could be counted and the total time spent in each state measured. The clutch success rate was calculated as the ratio of successful clutch actions relative to total clutch actions (i.e., including unsuccessful attempts). Participants then completed the NASA-TLX survey after each block of tasks was completed [154].

A semi-structured interview was completed after the final task block, to complement the quantitative data with qualitative feedback. Interviews were structured around findings from previous research that investigated the needs and experiences of clinicians when accessing medical imaging [15]. The interviews aimed to assess user satisfaction when using touchless interaction with medical imaging, assess their thoughts on the four clutching techniques, explore their views on touchless control as an alternative to traditional mouse and keyboard interaction, and discuss attitudes towards touchless interaction with PACS.

Interviews were audio-recorded and transcribed, responses were then thematically analysed. Transcripts were coded in the ATLAS.ti Web tool using the Framework Method described by Gale et al. [89], using the constant comparative method to identify and develop themes in the transcripts. After an initial round of coding, these were structured into the higher-level categories that are discussed in subsection 5.4.3 (p.131). The qualitative findings from the interviews provide additional insight, complementing the quantitative results, and draw on the experiences of a significant sample of practising hospital clinicians and PACS users.

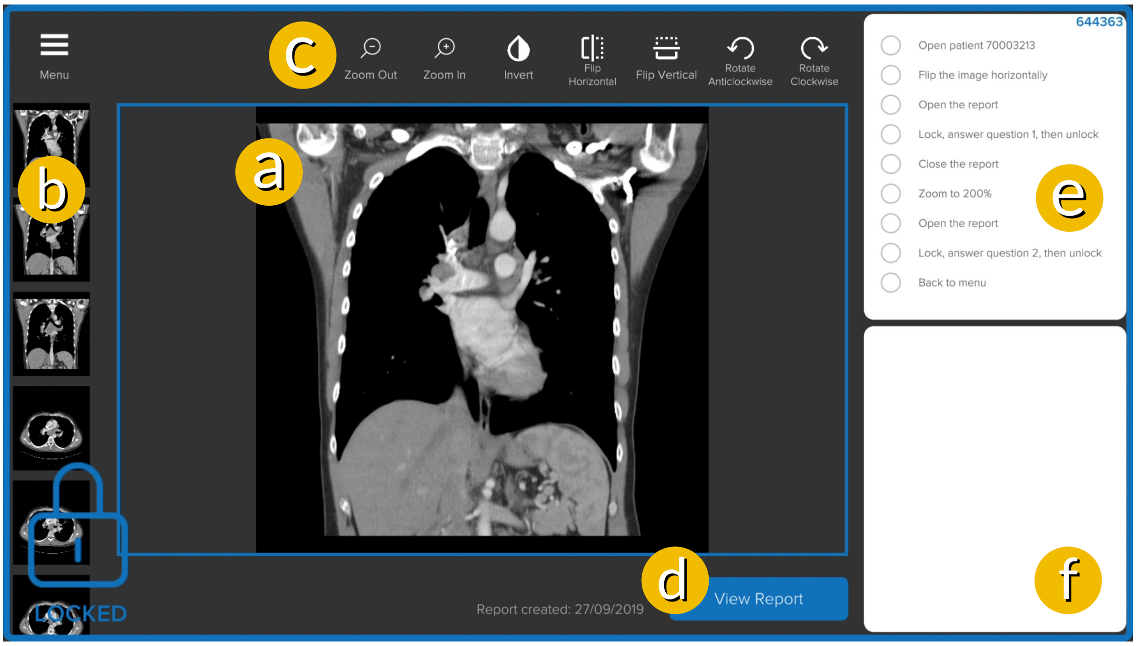
5.3.4 Design and Implementation

PACS interface

A custom generic PACS user interface was developed for this study (detailed in Chapter 4, p.75). This interface employed a simple DICOM loader for displaying authentic medical imaging files, and supported common PACS tasks such as image browsing, image manipulation (e.g., zoom, rotate, invert, flip), and viewing reports. The final set of PACS tasks implemented in the system was based on combining several typical PACS functions presented by Madapana *et al.* with input from clinicians [155]. Participants used speech commands for direct manipulation (e.g., “rotate image clockwise”, “zoom to 200%”), with corresponding mid-air gesture commands.

Figure 5.2 (p.125) shows the custom PACS user interface. This interface has a similar layout to existing PACS software: selected images are shown in the central area (a); a sidebar displays the set of images in the patient file (b); available image manipulation actions are presented in a toolbar (c); and patient reports can be viewed in a new window using the ‘View Report’ button (d). Task instructions and interaction feedback were also displayed in the user interface: a list of PACS task instructions was shown in the top right sidebar (e) and interaction feedback was given in the bottom right sidebar (f).

A dataset built from open-source DICOM images (.dcm files) was used for the experimental tasks. This dataset was acquired from online repositories like DICOM Library [156]. Images were chosen to represent common medical imaging. Patient metadata was modified to standardise patient IDs across all experiments. In order to remove the need for specific medical knowledge (since the participants had different areas of exper-



○ Open patient 70003213
○ Flip the image horizontally
○ Open the report
○ Lock, answer question 1, then unlock
○ Close the report
○ Zoom to 200%
○ Open the report
○ Lock, answer question 2, then unlock
○ Back to menu

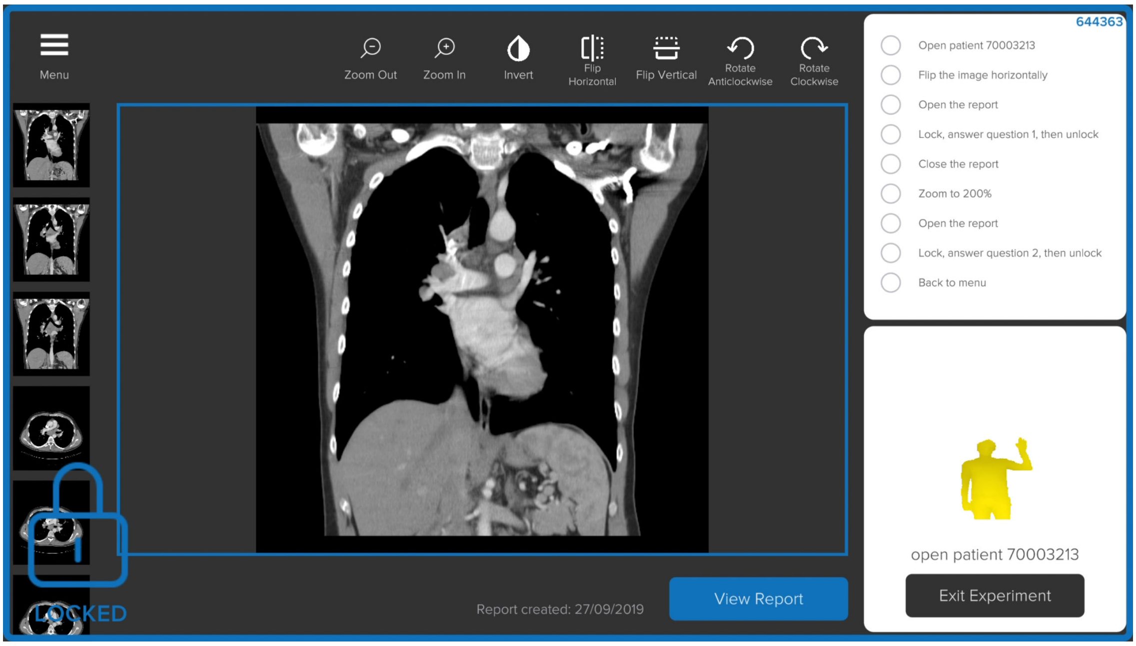
Figure 5.2: The PACS user interface layout (a–d), experiment tasks (e) and touchless interaction feedback panel (f).

tise), a standardised report was used for all experiments, allowing participants to answer secondary task questions directly from the reports without recourse to specialised domain medical knowledge.

Clutch interactions

The four clutch interactions were used to transition between locked and unlocked states in the touchless PACS UI. When the system was unlocked, it would then respond to speech/gesture commands.

The mid-air gesture clutch used a single raised hand, held above shoulder height for one second. This toggled between the locked and unlocked states. This gesture was chosen as a mode switching gesture because it could be robustly detected by a simple depth sensor, even when users are at a distance from the sensor. More importantly, it requires no motion detection and allows the users' hands and arms to be kept close to the body (which is crucial for maintaining asepsis during surgical procedures). It is also an action unlikely to occur naturally in situ. The voice clutch used the words "lock" and "unlock" respectively to switch between input lock states.



○ Open patient 70003213
○ Flip the image horizontally
○ Open the report
○ Lock, answer question 1, then unlock
○ Close the report
○ Zoom to 200%
○ Open the report
○ Lock, answer question 2, then unlock
○ Back to menu

Figure 5.3: The lock gesture has been used to clutch the system.

The gaze clutch determined the estimated gaze orientation to implicitly lock and unlock the touchless interface. If the user was making eye contact with the screen presenting the PACS interface, the system was automatically unlocked and responsive to input actions; conversely, if they were not looking directly at the screen, the system entered the locked state. The study could alternatively have included body posture for a more robust estimation of intention to interact (e.g., as in [134]). However, PACS usage will often require users to divide attention between multiple tasks, so users may glance to and from the screen without necessarily turning their body towards it. For this reason, the decision was made to use gaze alone to infer engagement with the interface.

The active zone clutch used body position and posture input to implicitly lock and unlock the touchless interface. An active zone was defined relative to the position of the screen displaying the PACS interface as a region of 1 m^2, starting from a distance of 1.5 m from the centre of the display. This was marked using tape on the floor. This region provided users with some flexibility regarding where they placed their feet, but was appropriately sized so that, in conjunction with gaze, it was possible to reliably infer an intention to interact with the system. When the user was standing in the active zone and looking at the screen, the system was automatically unlocked and responsive to input actions.

Apparatus

The software used for this study was implemented on the Microsoft Azure platform. The system used the Language Understanding service from Azure Cognitive Services for natural language processing, to map speech commands to user interface operations. The Face service from Azure Cognitive Services was used for facial detection, which determined when the user was engaged and looking at the screen. The Body Tracking SDK from the Azure Kinect DK provided user detection; this was used to track position in the room (active zone) and body posture (gestures).

Input sensing used the Microsoft Azure Kinect DK device, as described in Chapter 4 (p.75). This device has a 1MP depth sensor with room-scale range for user tracking, and an omnidirectional microphone array for robust speech tracking. This device was selected because it could be used to recognise all four of our clutching methods, supports the touchless interactions provided through the user interface, and integrates well with the aforementioned cloud services for accurate and reliable interaction sensing. A 30" monitor positioned on a table surface was used to provide visual output to the participants.

5.3.5 Participants

Thirty-four healthcare professionals (thirty-two practising clinicians and two dedicated PACS staff) were recruited, through existing contacts and a university-affiliated teaching hospital for these studies. All were health professionals with PACS experience, representing varying grades of seniority in the hospital. Table 5.1 (p.128) gives an outline of participants, their speciality area and years of PACS experience. Participants were asked if they had experience with touchless computer interfaces; Twenty-one had used voice interfaces, ten had used gesture or motion controls, and seven had used other devices (e.g., foot pedals). Eighteen participants wore face masks during their study due to COVID-19 restrictions; facemasks and personal protective equipment (PPE) are often worn in situ (especially in surgical contexts). Participants were not compensated for the approximate forty-five –minute study time.

Role / Speciality	**N**	**Experience**
Student	3	0–4 yrs: 3
→ *Radiography*	3	
Senior House Officer (equiv. Resident)	13	0–4 yrs: 11 5–9 yrs: 2
→ *General*	2	
→ *Orthopaedics*	1	
→ *Paediatrics*	1	
→ *Obstetrics & Gynaecology*	3	
→ *Anaesthetics*	1	
→ *Surgical*	5	
Specialist Registrar (equiv. Fellow)	13	0–4 yrs: 2 5–9 yrs: 10 10–14 yrs: 1
→ *General*	4	
→ *Paediatrics*	1	
→ *Obstetrics & Gynaecology*	2	
→ *Orthopaedics*	1	
→ *Surgical*	5	
Consultant (equiv. Attending)	3	10–14 yrs: 2 20+ yrs: 1
→ *Radiology*	2	
→ *Surgical*	1	
Other	2	5–9 yrs: 1 5–9 yrs: 1
→ *PACS Manager*	1	
→ *PACS Clerical Officer*	1	

Table 5.1: Participant role, speciality and years of PACS usage experience.

5.4 Results

An analysis of interaction measurements (subsection 5.4.1, p.129) and NASA-TLX survey results (subsection 5.4.2, p.130) for the cohort follows. This analysis is followed by discussion of the key themes derived from the qualitative analysis of the interviews (subsection 5.4.3, p.131).

5.4.1 Interaction

Figure 5.4 (p.129) shows the number of clutch transitions measured for each condition and the percentage ratio of successful clutch actions. Figure 5.5 (p.130) shows the mean total task time for each condition and the mean cumulative time in the 'unlocked' state for each block.

A mean of 6.31 (SD 4.86) transitions per interaction task was calculated. Friedman's test found a significant effect of clutch method on the number of transitions: $\chi^2 = 27.8, p < .001$. Post hoc Nemenyi tests found that Gaze had more transitions than Active Zone ($p = .007$), Gesture ($p = .02$) and Voice ($p = .001$) respectively.

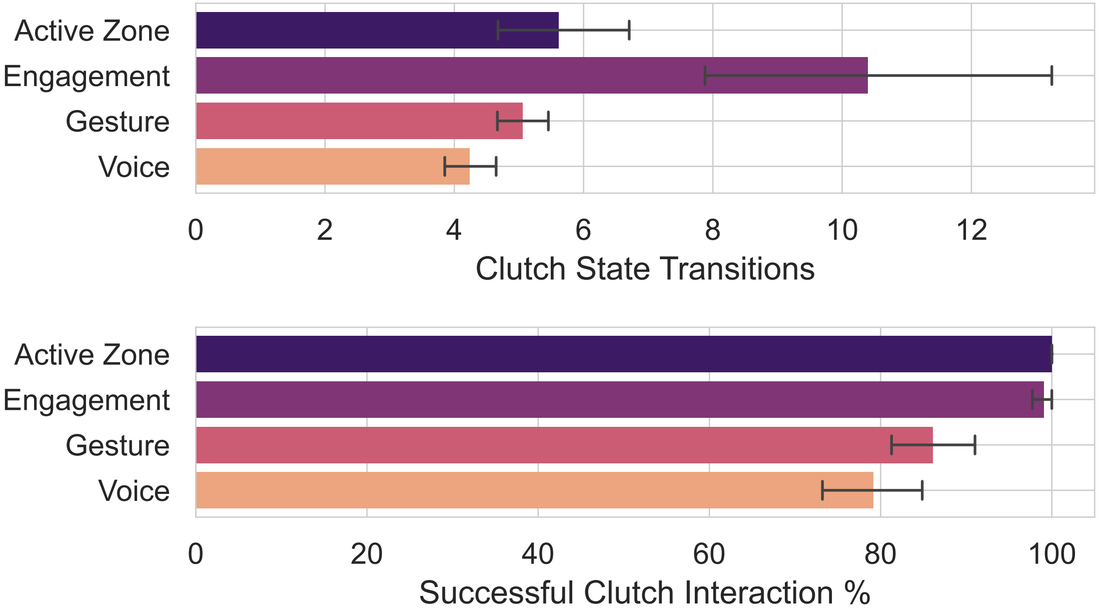


Figure 5.4: Number of transitions and clutch action success rate. Error bars show 95% CIs.

Mean clutch transition success rate was 91%. Friedman's test found a significant effect of method on success rate: $\chi^2 = 44.1, p < .001$. Post hoc Nemenyi tests found higher success rate for Active Zone than Gesture ($p = .008$) and Voice ($p = .001$), and higher success rate for Gaze than Gesture ($p = .02$) and Voice ($p = .001$).

Mean time-on-task for each block was 154.3 seconds (SD 53.5 seconds). A repeated measures ANOVA did not find a significant effect of method on time: F(3, 96) = 0.61, p = .61.

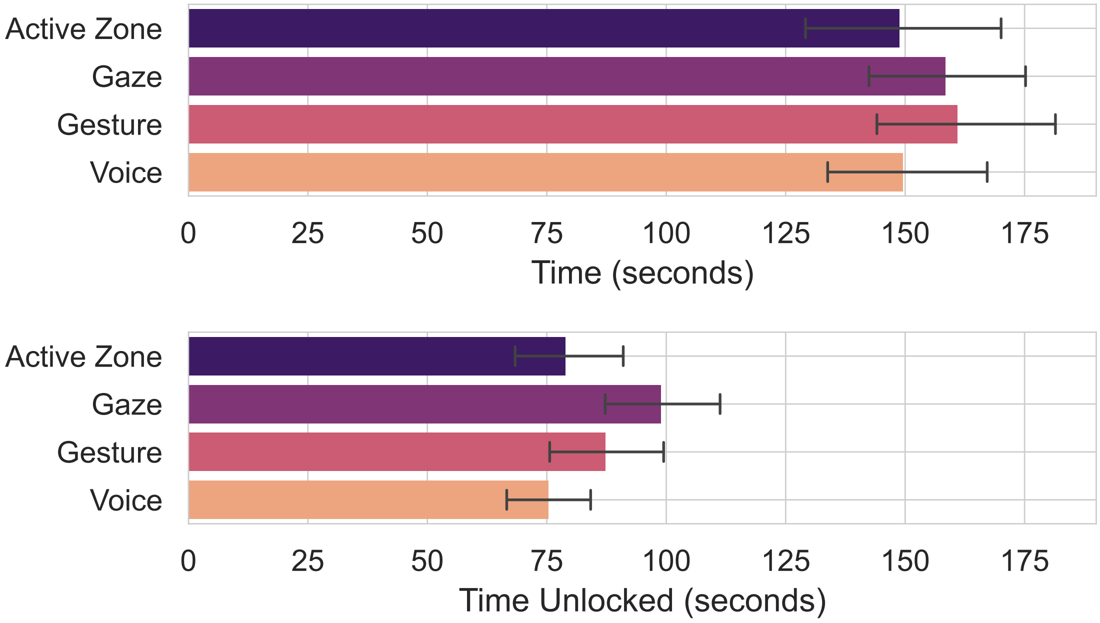


Figure 5.5: Total task time (top) and total time in the unlocked state (bottom). Error bars show 95% CIs.

Mean time unlocked was 85.0 seconds (SD 34.3 seconds), which was 55% of the overall task time. A repeated measures ANOVA found a significant effect of method on total time unlocked: F(3, 96) = 3.42, p = .02. Post hoc t-tests found time unlocked was higher for Gaze than Active Zone (p = .03) and Voice (p = .001).

5.4.2 Task-Load Index

Overall task-load index (TLX) was calculated as the mean of the six components (on a scale of 0–100) [154]. The mean TLX was 23.1 (SD 13.9). Figure 5.6 (p.132) shows mean TLX and CIs determined for each condition, including overall score and six components.

This analysis investigated the effect of condition on overall TLX and on the six TLX components: mental demand, physical demand, temporal demand, perceived performance, effort, and frustration. Note that lower scores are better (i.e., lower mental TLX means less mental demand).

Table 5.2 (p.131) shows Friedman's test results and, where appropriate, the significant results from post hoc Nemenyi comparisons. As can be seen from the Overall TLX results, Active Zone was less demanding than Gesture and Gaze, and Voice was less demanding than Gesture. These differences can be partly explained by key differences in the individual components; i.e., Active Zone was less mentally demanding and less frustrating to operate than Gesture and Gaze, whereas Voice was less mentally demanding and physically demanding than Gesture.

Rating	**Mean**	χ^2	**p-value**	**Sig. Comparisons**
Overall	23.1	21.9	<.001	A < GE (p = .001), A < GZ (p = .006), V < GE (p = .03)
Mental	22.9	26.3	<.001	A < GE (p = .001), A < GZ (p =.02), V < GE (p = .007)
Physical	15.8	12.6	<.001	V < GE (p = .02)
Temporal	19.7	6.08	.11	—
Performance	32.3	5.83	.12	—
Effort	26.3	14.9	.002	A < GE (p = .005)
Frustration	22.5	17.3	<.001	A < GE (p = .008), A < GZ (p = .007)

Table 5.2: Mean task-load index scores (overall and each of the six components), Friedman's test results and, if appropriate, significant Nemenyi test comparisons (A: Active Zone, GE: Gesture, GZ: Gaze, V: Voice).

5.4.3 Interview Findings

This section considers the key themes derived from analysis of the interview transcripts. Participants are anonymised by assigning a unique number identifier and their professional role, codified in Table 5.3 (p.133).

The following sections discuss thematic findings regarding the use of the clutching interactions (in Table 5.4.3, p.133), followed by a discussion of themes related to the

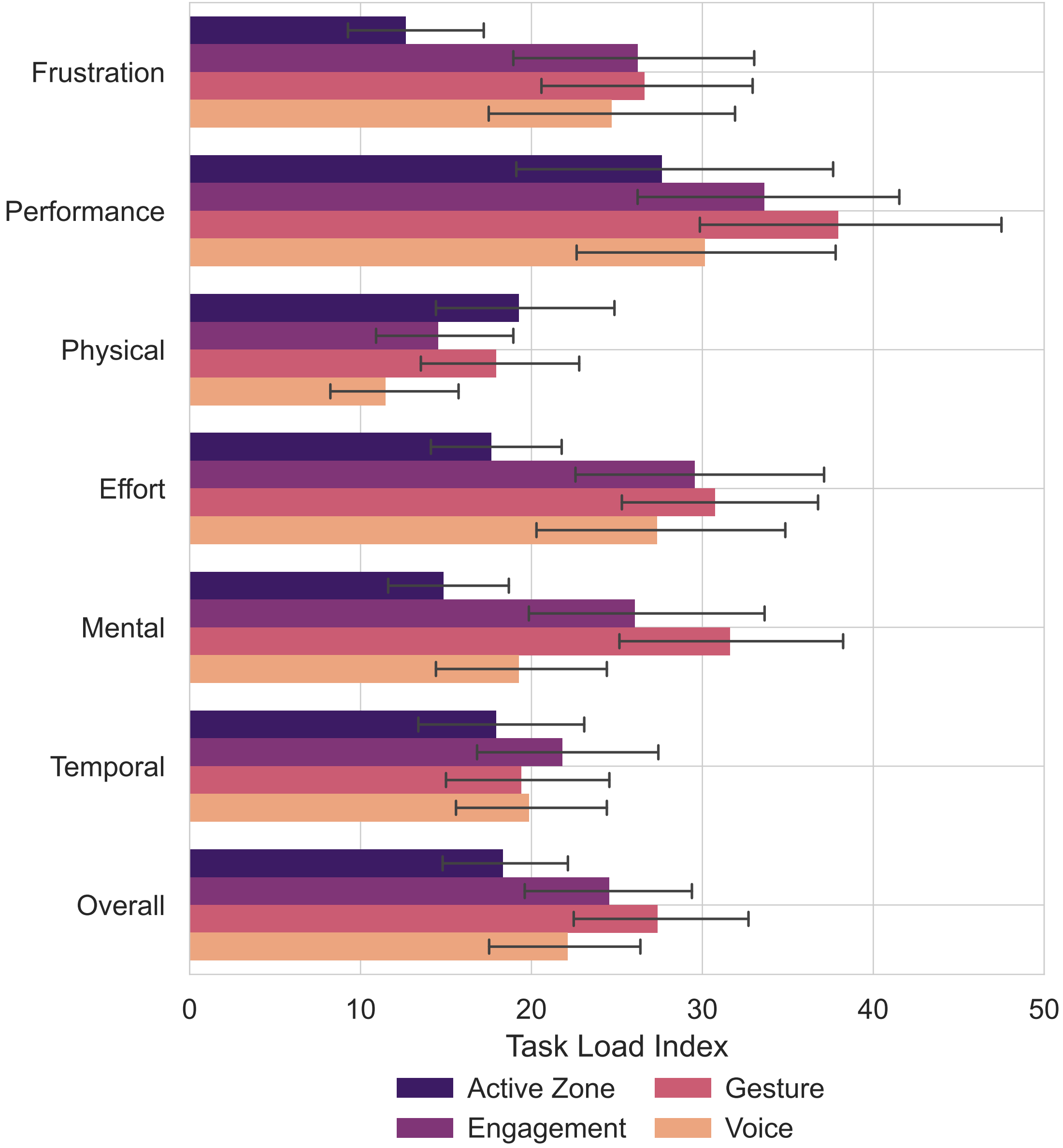


Figure 5.6: Mean TLX scores. Error bars show 95% CIs.

Code	Professional role
PACS	Admin staff (e.g., PACS manager)
STDT	Student
SHO	Senior House Officer
SREG	Specialist Registrar
CS	Consultant Surgeon
CR	Consultant Radiologist

Table 5.3: Professional role codes

general use of touchless interaction (in Table 5.4.3, p.136).

Clutching interactions

Voice For many of the health professionals, voice was the *"most straightforward"* [SHO 13] and *"easiest, most intuitive"* [SREG 3] clutch method to use. Most (21) had prior experience using voice control (e.g., digital assistants), so familiarity was perceived as an advantage. Another advantage of voice in this context was that the unlock and lock commands could be issued from anywhere: *"it removes the element of where you're standing"* [SHO 13] and did not disrupt tasks elsewhere in the room, unlike the other methods that required standing within line of sight of the Azure Kinect DK sensor.

Voice input was not without issues, however. Some users reported it was *"frustrating when the voice wasn't listening to you to unlock and lock"* [SHO 10]. This was often the result of misrecognising speech commands: *"not recognising certain phrases from myself"* [SREG 8]. Misrecognition often occurred due to factors out of the user's control. An important observation was the potential impact of PPE: *"voice recognition, when you're wearing an FFP3 mask might be a problem"* [SREG 12] because it muffles speech.

Many users pointed out the negative effect of ambient noise in a busy environment with other people, e.g., SREG 1 and SREG 6 both said the operating theatre can be very noisy and this could cause issues when detecting the clutch commands, whilst SREG 6 and PACS 1 highlighted the potential for false-positive input from others talking nearby who were not part of the interaction. Being around others could also cause other difficulties when using speech to unlock the system, as the user may need to interrupt ongoing conversations: *"it would be great if you didn't have to stop talking, issue a command, and start talking again"* [SREG 10, SHO 10].

The study used "lock" and "unlock" as key phrases for (un)locking the system, although some participants noted that these may not be appropriate for all usage contexts within the hospital. For example, *"locking and unlocking are a common term in orthopaedics for actually fixing things, so it might cause confusion"* [SREG 5]. In practice, the key phrases for clutching will need to be chosen carefully, to avoid false-positive input from words that have domain meaning and may occur often in that context.

Gesture Gesture input was a novel interaction modality for most users (only 10 had prior experience, mostly from motion controls in games, e.g., Nintendo Wii or Xbox Kinect). Whilst only one clutch gesture was used in the study (which acted as a mode switch), some users noted the mental demand associated with remembering this as part of the wider gesture set: *"it [was] a little tiring to remember the gestures"* [SREG 10, CR 1]; this was also reflected in the TLX scores. Some participants noted that this was largely due to the novelty of these gestures and that they would become *"second nature"* [CR 1] with experience; indeed, health professionals *"are invariably tech-savvy and adaptable... it's not rocket science"* [CR 1] and gestures would become easier to learn and use with experience.

Similar to the voice clutch modality, participants raised concerns about the negative effects of PPE on the clutch gestures, both in terms of recognition difficulties and their own inability to perform certain gestures: *"[they] probably wouldn't be captured, just because you're wearing a gown in theatre, if you were scrubbed"* [SREG 1], *"could never use that one if I was wearing any type of scrubs or gown"* [SHO 5]. This poses an interaction challenge, as those with the greatest need to interact with PACS in theatre are likely to be the health professionals who are scrubbed.

Gaze For a small number of health professionals, the gaze clutch method was *"definitely the best"* [SREG 10], *"my preferred method"* [SHO 12], as it streamlined interaction and required no additional effort to begin issuing speech or gesture commands for the PACS interface: *"you do not have to raise your hand or say anything, that was good"* [STDT 1]. Participants also commented positively on the lack of clutch errors, an observation supported by the data (99.1% recognition). However, it was the least preferred and the majority found it challenging—it was *"very unnatural feeling"* [SHO 13].

It seems most participants experienced a conflict between their understanding of the gaze clutch and its actual behaviour. Gaze was an implicit clutch that automatically unlocked the touchless PACS interface whilst a user was looking at the system. For many users, this was undesirable—they wanted to be able to look at the system without it clutching for input (e.g., to refer to the patient files for the secondary task): *"I was looking at the screen to get the report, I automatically kept engaging it"* [SREG 3]. This loss of control over the clutch state—the inability to prevent the system unlocking when glancing at the screen—was frustrating to them. As a result, some participants described strategies to avoid unlocking while looking at the screen, e.g., sideways glances so the system did not think they were looking.

Despite frustrations with the gaze unlocking, some participants found it to be a particularly compelling modality for initiating interaction; they described existing frustrations with the need to repeatedly unlock computers each time they wish to use the system, due to data protection and security policies: *"locking full-stop is what bothers me the most about PACS... if there was an option to never lock throughout the whole operation that would be useful"* [SREG 3]. It was suggested that an improved version of gaze, used alongside facial recognition for authentication, could address such concerns, i.e., for unlocking the system *and* from an input recognition perspective.

Active zone The active zone clutch was notable for the lack of frustration expressed by participants during the interviews. As might be expected from the 100% recognition rate, *"I don't think it missed a beat once, that was good"* [STDT 1]. Unlike gaze, which was also an implicit clutch method, participants seemed positive about the automatic (un)locking with the active zone: e.g., *"I think that's good for safety, it means you don't have to unlock it ... [then] you just walk away and it automatically locks"* [SHO 1]. Participants also suggested this could simplify data protection practices: e.g., *"I can see the active zone* *being useful, like when you walk out of your office, it will lock it automatically"* [STDT 1].

A key advantage of the active zone method was that it was a good fit for existing PACS usage and could integrate well with existing workflows and practices. Participants explained that in real PACS, they will often move to a computing terminal, issue commands via mouse and keyboard, then step away again. Active zone required similar behaviours, e.g., *"stepping away is something you'd be doing anyway"* [SREG 7].

General views on clutching Participants generally appreciated the value of clutching as a tool for avoiding accidental input—*"the most important part of this"* [SHO 10]—especially in situations where having the correct information on the screen is vital. As observed with gaze, security and data protection also arose in discussion about clutching. The need to authenticate before PACS usage adds workflow friction and, whilst out of the scope of this research, many saw the potential for touchless interaction to aid authentication. Common HP suggestions included; using facial or speech recognition to identify the active user of the system, such that clutching and authentication could happen concurrently.

Touchless interaction

After discussing the clutch methods, participants were asked to discuss their thoughts on touchless PACS interaction in general. Five main themes emerged during analysis: (1) Workflow—how touchless interaction could impact and integrate with existing clinical practice; (2) Sterility—benefits of touchless input for sterility, especially in light of the COVID-19 pandemic; (3) Environment—challenging aspects of the usage context; (4) Adoption—integrating touchless technology into the health profession; and (5) Applications—promising use cases for touchless input.

Workflow Many of the health professionals saw the potential for touchless interaction to improve their workflow and were excited about this: e.g., *"it's the future of what we're doing in PACS... you just look at the screen, it will come on, you just say "open that patient", it will just come up... it's gonna replace keyboard and mouse I think, you know, it's just a matter of time"* [CR 2]. For most, the key benefits were the potential for more efficient interaction: e.g., *"it's much easier, it's quicker, it's much less frustrating"* [CS 1] and *"it's innovative and potentially the future, and allows one to potentially be more efficient, so, yes, I did like it."* [CR 1]

Also important was the ability to interact with PACS directly, which is not always possible in situ. To maintain sterility and avoid cross-contamination, health professionals cannot always directly use computing devices. Instead, they need to give commands to another person who controls the system on their behalf, although this is not without operational friction: e.g., *"what we usually do was ask one of the scrubbed nurses to open up, but they don't necessarily [know] what image sequence you wanted to open up"* [SHO 2].

Touchless interaction avoids this indirect input and enables the health professional to interact directly: e.g., *"you have to ask for someone to stay with you for several hours to look at the scan, [doing it yourself] is really very helpful"* [SREG 13], *I think it would be really useful in theatre if we wanted to scroll through images while we're scrubbed and couldn't use a mouse and keyboard* [SREG 5], and *"you're trying to avoid having to ask other people in the room, such as the circulating nurse who may not be familiar with the technology... certainly, if you can talk to it, or gesture to it to pull up the exact image that you need, I think it would be very useful"* [CS 1].

Whilst these benefits are compelling, it was clear that reliability is a key factor in willingness to adopt a touchless PACS interface: e.g., *"nothing's 100%, if you can rely on it 95%, people will engage... [but if not] people have something that works, even if it's slower, they know it works"* [CR 2]. Some noted it would take time for touchless interaction to be integrated into their workflow, due to familiarity with existing input modalities: e.g., *"I'd prefer a touchless system in the long run, but all systems require you to kind of engage with it, you know, and get used to it... at this moment in time, just more used to mouse and keyboard.* [CR 2], *"I'm used to using the mouse and keyboard... it just takes time, because you're thinking about what you're doing before you do it... with a bit of practice, it probably would be as easy"* [SREG 6], and *"particularly when you're doing a repetitive task, such as we do in radiology, it would become second nature"* [CR 1].

Participants thought different interaction modalities would be better suited to different tasks within the PACS workflow. Speech and gestures would be ideal as shortcuts for simple and repetitive actions like zooming, panning, moving through image sequences, but less suited to more complicated tasks: e.g., *"once it is kept simple, it is better than mouse and keyboard... but if there was a command for zooming in, raising your hand or something, yeah... for anything more complicated, I would prefer mouse and keyboard"* [STDT 1].

Sterility One of the main perceived benefits of touchless interaction was that it would allow health professionals to interact with PACS under sterile conditions and help reduce cross-contamination between surfaces and people: *"it decreases kinds of cross-contamination in a hospital setting"* [SREG 11]. Touchless input can overcome this challenge: e.g., *"we shouldn't be touching things, I don't think we should be touching things... I should be able to sit at a computer and be able to sift through, use a hand gesture, to sift through a scan"* [SHO 10], *"I think in theatre, I can imagine it being used a lot, say if like somebody is scrubbed and wanted to see a scan and there was a big monitor on the wall, I think it'd be really useful... not even in theatre but in other scenarios where people are scrubbed or in sterile environments"* [SREG 6], and *"I envision touchless interaction being particularly useful in high-risk areas where there is a high risk of contact being made, such as a busy ED ward in a COVID-19 setting or theatre, in what is meant to be a sterile area"* [SREG 8].

The significance of sterility was amplified during this study by the COVID-19 pandemic, during which this research took place. Social distancing and increased focus on sterility had a disruptive effect on workflow and interaction both with patients and computing systems, and is likely to have a lasting effect on shared hardware in the health profession. This increased the perceived benefit of touchless input for PACS and other computing tasks: *"because of this pandemic and COVID-19, everything is being moved towards minimal touch or touchless, so yes, why not [PACS displays]"* [SHO 11], *"I was very pleased with it, especially in times of COVID-19 where we're trying to minimise contact with things, so I think it's a great idea... it would be a perfect time to introduce this into the wards"* [SHO 9] , and *"I think it's exceptionally useful, particularly at the moment in the middle of a COVID-19 pandemic when we're all sharing the same keyboard, and* *maybe aren't paying attention to sanitising between them, so certainly from that point of view I'd feel a whole lot happier with [touchless]"* [CS 1].

Environment Whilst this study was conducted in several hospital settings, the rooms in which the study sessions occurred were generally quiet and free from disruption. In reality, the actual usage context is busy and noisy, and many participants expressed concern that this could limit effectiveness, reliability and usability. As noted earlier, PPE could cause interaction difficulties, e.g., facemasks muffling speech and protective gowns causing gesture detection issues.

Health professionals often work with others and the presence of bystanders was seen as a potential issue for gesture recognition: *"it may not pick up your [gestures] or may not see you exactly doing them"* [SHO 12]. Having others nearby could also cause issues of occlusion and ambiguity over who the active user is, especially for the active zone clutching method: e.g., *"one barrier I envisage is space, particularly for the active zone feature"* [SREG 8], since that might create an area of the room that other health professionals need to avoid standing in.

Background noise was seen as a key problem for voice recognition: *"hospitals are noisy, noisy places"* [SREG 6], including conversations between others that might affect the PACS interface: e.g., *"if you were talking and then I was talking, the system wouldn't understand"* [PACS 1] and *"[it would be good if] it isn't distracted by other voices speaking around it... it might pick up their voice"* [SREG 3]. Recognition failure was one concern with noise, but false-positive actions could also cause disruption, e.g., *"more importantly, if [I am] spending an hour on a report and someone walks in and says a word like 'cancel' for whatever reason, and it deletes everything, no one's going to be happy"* [SREG 12].

Some noted the importance of room layout, so that visual displays would be clearly viewable and input sensors would be able to capture input: e.g., *"sometimes you operate from the top of the patient or the bottom of the patient, so looking back at the screen might be an issue"* [SREG 12] and *"I'm short-sighted, so I have to stand particularly close to be leaning in to see, I guess you could just always reposition the camera if that was the case"* [SHO 12].

Adoption There were many observations about factors that may affect the adoption of touchless interaction in hospitals. Two users suggested people may feel uncomfortable using touchless input actions around other people: *"you may think you look a bit foolish doing the arm movements, but overall, if you have sort of a quiet area, it's definitely better"* [SHO 12].

Many of the health professionals identified the need for training and discussed the learning effort they experienced during the study: *"it did take me a little bit of time to realise how to use it properly"* [CS 1] and *"you might need just to train them well"* [SREG 13]. There was a positive outlook on the impact of training by some: *"there has to be good training and adaptation by the user group...I wouldn't envisage that as a problem"* [CR 1].

One participant suggested that training could also encourage adoption, by helping novice users see the time-saving benefits of touchless input: *"find the 10 or 20 things that they actually use a computer for and say this is how you do it, and it's actually going to be much quicker for you"* [CR 2]. Demonstrating the benefits could help encourage those who are reluctant to change working practice, which is *"a pretty systemic issue"* [SREG 2] and so *"showing people that it is actually going to make their life easier is probably the most important thing"* [CR 2].

Applications The research focus in this qualitative study was on interaction with a generic touchless PACS interface. Many of the health professionals discussed other areas in the medical profession where touchless interaction could be useful in the interviews. The operating theatre was a commonly suggested environment, due to the need for complex imaging but difficulty to directly interact with it: e.g., *"in the operating theatre, if we're doing a complex case that relies on complex imaging such as cross-section CT scanning angiography where you're trying to match the image to the operative site... you're trying to avoid having to ask other people in the room, such as the circulating nurse who may not be familiar with the technology... if you can talk to it, or gesture to it to pull up the exact image that you need, I think it would be very useful"* [CS 1].

Radiology was also proposed as a compelling use case, where voice control is desirable as it does not require body position changes or the use of hands: *"I would see it as being useful in both diagnostic and interventional radiology, whereby you don't have to be employing other parts of your body to do stuff that you're actually trying to concentrate on... voice control would be very handy in intervention, to take an image to mag up, rather than having to instruct a radiographer or to do it yourself physically* [CR 1].

Finally, the catheterization lab was another environment identified where touchless interaction could provide workflow improvements. Like the other suggestions, touchless input would remove the need to operate via a technician to control the system: *"in the cath lab, if you want to look at an old image there is a technician who sits outside who will load the images, like a radiographer, for the person who is performing the procedure to have a look at... usually, you'd look at it before the procedure, but if you needed to remind yourself, maybe it would be useful to be able to ask for it to come up—if you could say 'load it up on screen two'... if you had control of loading and scrolling through the image at a certain speed... you're not trying to communicate probably quite a sophisticated interpretation through someone else* [SREG 3].

5.5 Discussion

Touchless user interfaces are compelling for use in hospital environments because they can help reduce the spread of pathogens. This is more important than ever after the COVID-19 pandemic heightened awareness of the infection control risks associated with shared input devices—especially in hospitals. In addition, touchless input also has the potential to improve PACS usage through faster interaction and by facilitating direct input without the need for a proxy user. For touchless technology to be deployed successfully in this context, it needs to be reliable, enhance rather than hinder clinical practice, and be easy to use in demanding situations without contributing to additional cognitive load. This research study looked at methods for mitigating false-positive input, a crucial criterion for reliable touchless interaction. A combination of quantitative and qualitative findings provide insight into the ways touchless interaction can be integrated into clinical contexts, and inform the design of clutching methods for more reliable input in this context.

5.5.1 Touchless Interaction in Clinical Contexts

This study evaluated a generic touchless PACS interface with a representative, varied sample of health professionals in hospital environments. It was conducted with experienced PACS users, thereby gaining valuable insight into their needs and requirements for a touchless user interface. Key themes relevant to how touchless interaction could benefit PACS usage are identified; an important finding being participants displayed a desire for adding touchless capabilities to other hospital systems.

Benefits could be realised in any aseptic environment where imagery access is vital, e.g., the operating room or during intraventional radiology [**RQ4**]. The *Workflow* and *Sterility* themes illustrate that existing workflows are built around maintaining asepsis: i.e., avoiding contact with input devices and shared surfaces, issuing commands to a nurse or intermediate user, and having to anticipate imagery needs in advance of procedures. Other work notes that users may ignore imagery entirely when control is not possible due to sterility constraints [15], which is not ideal for patient outcomes.

Participants recognised the potential that touchless technologies have in addressing these barriers to interaction: it allows them direct control and facilitates interaction whilst maintaining asepsis [**RQ1**]. Participants expressed a strong desire for this level of interactive control, citing how much it could improve their workflow. Touchless input can also remove a significant source of frustration and inefficiency: e.g., the challenges of explaining precise imagery needs to a non-expert. In the worst case, having to break asepsis for interaction and then scrubbing back in, can add considerable delay. Another potential benefit of touchless input was the faster interaction it affords, e.g., by allowing users to issue a spoken or gesture command from anywhere in the operational space, even without taking their hands away from a patient. These findings are consistent with prior work, which has also highlighted the benefits of touchless input in this context (as discussed in section 5.2, p.117).

A number of challenges and possible limitations to touchless interaction in the clinical usage context, were explored in the *Environment* theme [**RQ4**]. Gesture recognition may sometimes be impeded by protective gowns or visors, and participants noted that busy environments could confuse input sensing. Similarly, speech recognition could be affected by face masks, visors, and cross-talk. These ecological challenges have implications for interaction design: e.g., which gestures can be robustly sensed through PPE, which gestures require the least physical space, which commands are less affected by muffled speech? Similarly, there are implications for deployment: e.g., what is the most suitable sensing technology and the optimal sensor placement? Such questions provide an agenda for future research, with relevance beyond clinical application areas (e.g., industrial settings where protective equipment, ambient noise, and crowded environments are also common).

The study found that background noise and ambient sounds could also be challenging for voice control in certain hospital environments; e.g., the operating theatre can be particularly noisy. A voice-controlled touchless system would need to work reliably in such noisy spaces, also participants suggested that having to repeat commands would negate any efficiency benefits. Another auditory challenge comes from other people; clinicians often work alongside many people. Overlapping conversations could confound speech recognition and there was a sense of concern that a touchless PACS user would need to interrupt other ongoing conversation so that speech commands could be issued in relative silence. More robust speech recognition, e.g., detecting commands in mid-sentence, could help with addressing this.

A Microsoft Azure Kinect DK was used in the prototype system, a commodity device with powerful SDK support that simplified the development of the touchless PACS interface. Low-cost devices and emerging sensing frameworks (e.g., Microsoft Azure, Google MediaPipe, NVIDIA DeepStream) will make it easier and less costlycheaper for robust touchless user interfaces to be developed on the scale that regionalised hospitals or a national health service will require.

Impact of COVID-19 Though the user studies presented (C. 3 and C. 5) in this thesis were not intended to examine the impact of the COVID-19 pandemic on HP opinions, comparing HP responses from before the start of the COVID-19 pandemic with HP responses following reveals two significant insights.

First, prior to the COVID-19 pandemic, many HPs reported not being concerned about sterility when interacting with medical imaging, with HPs reporting not considering sterility when interacting with medical imaging. There was a sense that it was the user's own "grim" on the surface the user was making contact with and was therefore less of a concern, e.g., CR2:"No...it's my own grime...I'm not worried about that." In contrast, following the advent of the COVID-19, HPs reported a greater need to avoid contact with surfaces, e.g.,SHO10: "We shouldn't be touching things.". COVID-19 was cited as a motivating factor in HP desire to reduce surface contact, both to protect patients and the HP themselves, with HPs expressed concern regarding shared use of devices such as keyboards. Future work regarding the persistence of this increased level of concern regarding reducing surface contact and enhancing sterility would be of interest.

Second, following the advent of the COVID-19 pandemic, HP attitudes to adoption of touchless technology became increasingly positive. HPs had been receptive to the idea of touchless control before COVID-19, e.g., SREG3: "I think if I was there we would definitely use it...it would just be handier to use if you were busy doing something in theater.", though some HPs were unsure that touchless interaction would deliver value to their workflow, e.g., SHO9: "It doesn't seem this immediately intuitive to my day to day work." Following the advent of COVID-19, HPs seemed to find it easier to visualise touchless control being a part of their workflow especially in high-risk areas, e.g., SREG8: "I envision touchless interaction being particularly useful in high-risk areas where there is a high risk of contact being made, such as a busy ED ward in a COVID setting or theatre in what's meant to be a sterile area." HPs felt that touchless interaction could result in an improved workflow, and the introduction of a real-world TMIS may now be more easily achieved due to increased acceptance amongst HPs, especially in areas such as the OR and IR (SHO5: "I could definitely see it going into theater or interventional radiology.").

5.5.2 Clutching Methods for Touchless Input

The principal aim of this study was to investigate clutching methods for a touchless PACS interface. Clutching can improve gesture and speech recognition by providing users with a way to *address* a touchless user interface, while mitigating false-positive input and inadvertent operation. A variety of clutching methods are available, and this work brings clarity to which are suitable for clinical usage contexts.

Strengths and limitations were identified for each of the clutch techniques (Table 5.4, 145) that suggest that, depending on operational context, they may be more or less suitable for touchless interaction in clinical settings[**RQ3**]. Voice was the most familiar interaction modality and other work has suggested a preference for speech commands in hospitals (e.g., [15]). However, environmental constraints and recognition issues can impede accurate speech recognition. The study highlights the need to choose the speech lexicon carefully, e.g., the "lock" and "unlock" words are frequently used during orthopaedic procedures. Table 5.4 (p.145) presents the relative strengths and limitations of the touchless modalities investigated in this work.

Modality	**Strengths**	**Limitations**
Gesture	Combines dependable performance with the option of silent operation.	Users found this the most difficult to learn. Not suitable for hands-busy contexts. PPE, such as surgical gowns, has the potential to degrade performance.
Voice	Found to be intuitive, with many users having prior experience with voice control.	Failed recognition was frustrating to users. PPE, such as facemasks, has the potential to degrade performance.
Gaze	Potential for easy integration with other touchless modalities, especially with its high-performance accuracy.	Users had low confidence in performance, required checking for success multiple times. Deemed unnatural feeling to users.
Active Zone	Users found this the least demanding, with high accuracy enhancing their experience. Would work well with existing PACS usage habits. Strong potential for real-world application.	May be unsuitable for environments with limited space within a shared area.

Table 5.4: Strengths and limitations of touchless clutch modalities

It was hypothesized that gaze would be easy for users to understand and a suitable implicit clutching method: i.e., simply look at the system to unlock it. This method caused more user frustration than anticipated, due to users often wanting to look at the system to refer to PACS images and reports, without it unlocking and becoming responsive to input. Some users tried to work around this behaviour, e.g., glancing to avoid making direct eye contact. In this case, the clutch method was not congruent with the manner PACS images and reports were used during secondary tasks, which often required glancing at the screen.

Both quantitative and qualitative findings appear to support the use of the active zone [133] as a good choice for this context. This is possibly due to the simple sensing requirements of this method making it easier for humans and computers alike to operationalise: viz. users understood how it worked and were able to assert control over the system by moving in/out of the active zone. The system was able to robustly detect when the user was intending to interact as body position could be reliably estimated regardless of posture and the confounding effects of the presence of PPE, etc. Perhaps most importantly, the active zone was a good fit for existing usage habits and clinical workflow, so came naturally to many users. Marking the active zone on the floor with tape, so that users knew where their input would be sensed, was also an effective analogue solution to a key usability challenge when using sensor-based interfaces.

Active zone appears compelling for use in clinical contexts, although its use raises interesting questions for future work. Where should the active zone be located in the clinical field, and how large should it be?; could it impede other health professionals in the operating theatre, for example? Should there be multiple active zones to facilitate multiple users or to facilitate the need for a single user to move around during surgery, for example? Should *any* user in the active zone be given control of the system, or just a primary user, and how would user identification be facilitated?

5.5.3 Clutching + Authenticating

Hospitals understandably enforce strict information security policies to protect sensitive patient data, often requiring terminals to be unlocked each time they are used and automatically locking users out after very short periods of inactivity. This is a key source of frustration with existing PACS usage: authentication slows down interaction, e.g., having to re-authenticate a number of times when moving through a sequence of images. Many of the health professionals expressed a desire for clutching to work alongside authentication; indeed, this would be necessary in practice, since a touchless user interface would not be responsive if a terminal was locked or logged out.

Many participants seemed to appreciate that touchless user interface sensors could potentially identify them; using biometric approaches like facial recognition from optical gesture trackers, retina scans from gaze trackers, and voice recognition [**RQ2**]. This is a compelling topic for future work, as there are significant productivity gains to be had by making it easier for users to 'unlock' a touchless PACS system, both from an authentication and touchless input perspective. This work envisages a multimodal approach that streamlines access to medical imagery and reports in a touchless PACS interface, through combining active zone clutching with a biometric identification component. There are interesting challenges associated with the use of PPE, although persistent user tracking could allow authenticated users to be tracked before they scrub into their protective equipment and would keep the system unlocked so long as those known users remained in view.

During this study, Ireland's Health Service was affected by a devastating ransomware attack that brought NIMIS, PACS and many other hospital systems offline, severely reducing clinical work rates. Participants reflected on poor extant security practices during the interviews. They reported that sharing passwords was common to get around what they saw as restrictive security policies. Workarounds that lead to poor security practices and the increasing risk of cyber-attacks add further motivation for new authentication modalities to be researched in this context.

5.6 Conclusions

In this chapter, HP experiences with a prototype TMIS (section 4.4, p.102) were explored. HP views on touchless control, and, specifically, their attitudes towards touchless interaction with PACS are discussed. The results of HP interviews are analysed in this chapter to explore user satisfaction when using touchless interaction with medical imaging. HP thoughts on four clutching techniques are assessed, with Table 5.4 (p.145) providing an overview of the pros and cons of the clutching techniques investigated. It is clear that each of the clutching techniques investigated exhibited strengths and weaknesses; from HP feedback, appropriate selection of clutching mechanism is dependent on the context of use.

This study, together with the preceding chapters, has served to illustrate, from a variety of perspectives, the complexity of the design process for TMIS. In the next chapter, we put forward a characterisation of the development process to help provide some structure to future developers and researchers in approaching and managing the wide range of issues which will impact on the success of the systems they develop.

Chapter 6

A Framework for the Design and Development Process of Touchless Medical Imaging Systems

Within this thesis, multiple perspectives on the design of TMIS have been explored, highlighting many important design decisions and trade-offs faced by the designers of these systems. However, the number and variety of issues to be addressed can make it difficult to approach the design of these systems in a systematic fashion. Gathering together the findings of the thesis, this chapter provides a framework for the design and development process of TMIS, to guide future development. The aim is to provide a means for understanding the design space, and for articulating and contextualising design rationale for TMIS.

According to Moran and Carroll, design problems should be decomposed into manageable sub-problems [139]. As touchless control becomes increasingly relevant, there is a need to provide reliable resources for future developers and researchers to streamline their development process. As such, this chapter presents a framework for the design and development process of TMIS. The terms "framework" refers to a broad range of theoretical and practical concepts [128]. Generally, a framework is taken to provide a structure to guide design, programming, and research. Several examples of frameworks are discussed in the literature; O'Hara *et al* present framework of properties of touchlessness [112], Karam develops a theoretical framework for the research and design of gesture systems [128], Wigdor *et al.* and Nguyen present frameworks concerning the development of touchless NUIs [157, 129], and Placitelli and Gallo propose a framework to enable rapid prototyping of touchless user interfaces [158].

This thesis seeks to provide a conceptual design framework that allows the complexity of developing such systems to be better understood and mitigated, while delivering systems that anticipate the requirements of end users, their workflows, and their operational context. As such, the basic structure of the design framework in this chapter is presented in several stages; an overview of the design process, qualitative research as a key step, hardware selection decisions, software design decisions, user experience (UX) design decisions, and touchless interaction design decisions. These stages were adopted to represent high-level stages of the design and development of TMIS.

Several experiments and studies were conducted that contributed the qualitative and quantitative results that led to the development of the framework. Chapter 2 (p.11), which investigates existing approaches and findings, provides a basis for developing the framework. The qualitative investigation presented in Chapter 3 (p.45) provides insight into the needs of HPs in the context of PACS. The qualitative study informed multiple design decisions, such as the inclusion of voice commands. Comparing the process of designing touchless systems with and without qualitative research (as in Chapter 4, p.75), emphasised the impact of such research on the design process and the need to include qualitative research in the framework presented here. The process of developing the prototype systems, (Chapter 4 (p.75)), provides an opportunity to gain an understanding of complex issues related to the design process for touchless systems. The differences in hardware capabilities between the developed prototype systems in terms of the interactions they enabled, inform the discussion of hardware selection in this chapter. Chapter 5 (p.115) involved the exploration of HP user experiences with a TMIS, as well as comparing several clutching mechanisms, using a significant cohort of real-world HP users. This testing of various design decisions present in the experimental system, inform the framework in this chapter.

6.1 Overview of the Design Process

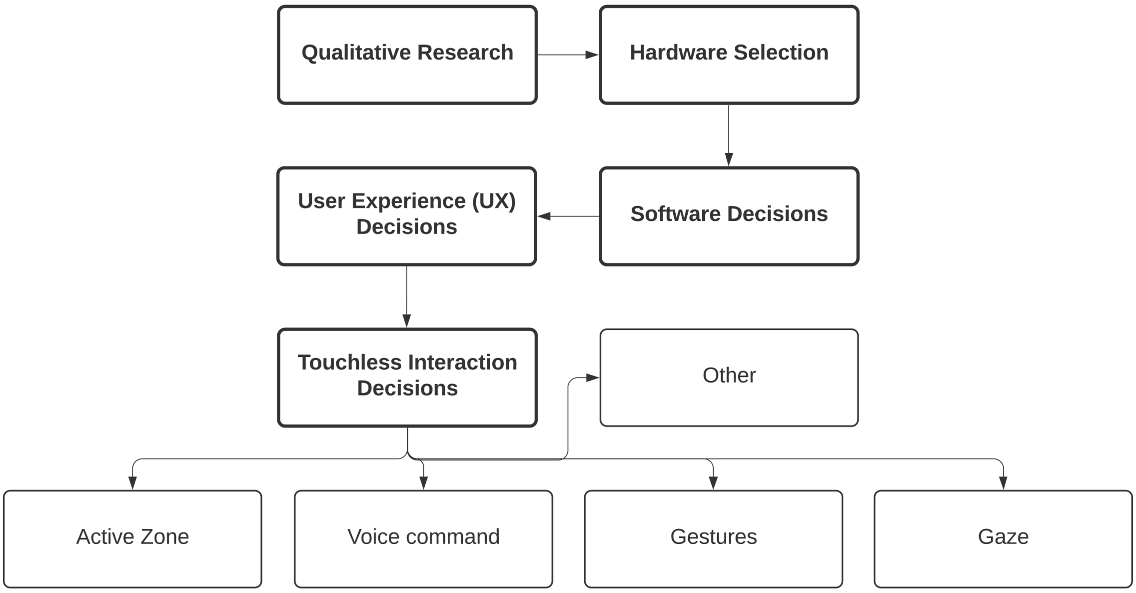


Figure 6.1: Design Decisions when Approaching a Touchless Interface for Medical Imaging

6.2 Understanding the Context: A Key First Step

A useful approach to developing a rich understanding of end-users, their roles, requirements, experiences, and attitudes regarding a technology is through qualitative research. A variety of analysis approaches such as grounded theory or thematic analysis provide appropriate methods for conducting such qualitative research.

For example, grounded research is a strategy for performing qualitative research [159] concerned with generating a theory or model which is "grounded" in data, which is systematically gathered and analysed. Grounded research is appropriate when little is known about a phenomenon [160]. A grounded theory approach may be appropriate when generating a deeper understanding of users and their needs, seeking to exhaust the data to generate a new model to guide the design of a TMIS.

The qualitative study presented in Chapter 3 (p.45) of this thesis is an example of a thematic analysis approach using the framework method, applied to understanding HP's when accessing medical imaging. An overview of the proposed research element of the design process is presented in Figure 6.2, p.152.

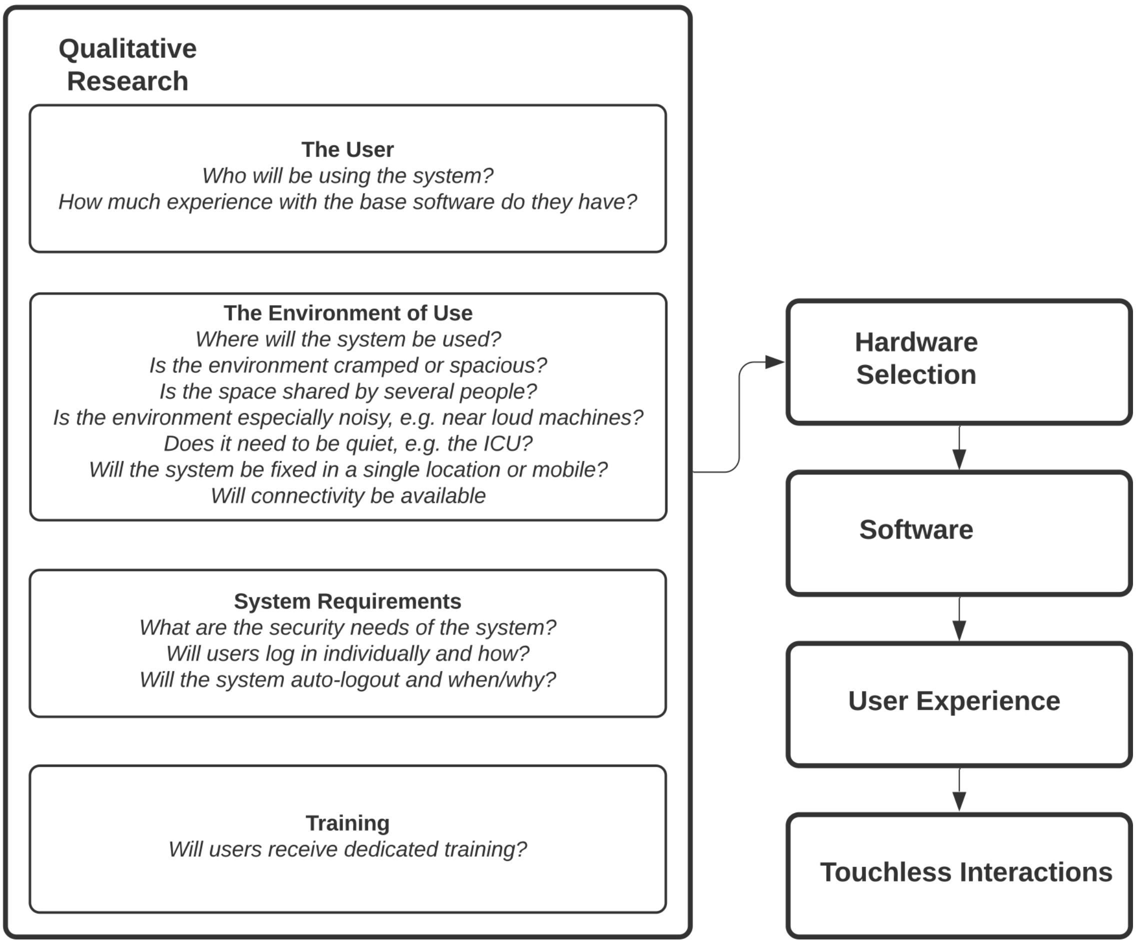


Figure 6.2: Qualitative research is a key part of the design process

6.2.1 The User

Who will be using the system(s)?

In order to better understand the user, their preferences, and requirements, it is essential to understand their work practices and how a touchless system integrates into their workflows; such insights can be gained through interviewing users and performing contextual enquiries [2]. HPs are in general a busy user group, and if a novel system does not fit well with their workflow, or their perception is that it hinders their ability to get their work done, they will be very reluctant to use that technology. Different HPs use very different workflows, for example, a surgeon's operating workflow in the OR is very different from that of a radiologist reviewing images in their office. By understanding these workflows, their similarities and differences, a touchless system can be designed to best suit the target users; for example, radiologists already use voice dictation and commands, so introducing a touchless system that prioritizes voice commands may reduce the learning effort for this user type and improve acceptance.

The role of the HP within the hospital can also impact the range of software functions they require (subsection 3.4.3, p.55). Though they may be using the same PACS, a clinician on the ward is unlikely to need to perform the sophisticated image reconstructions, and analysis that a radiologist may be performing. If a primary user group requires only a specific set of core functionality most of the time, that should prompt the designer to use a proven robust set of touchless interactions for those commands, with fallback to traditional input, such as mouse and keyboard, for deeper interaction. An overly complex touchless lexicon may result in more time and effort for users to master, risking that those users who may not have the time, or motivation required, will not commit to learn more than a subset of available interactions. A system may need to support multiple categories of user with significantly different needs. Using techniques such as providing support for user profiles to customise user roles with a touchless experience that best suits their needs.

How much experience with the existing software does the user have?

When adapting existing software to support touchless interactions, it is important to understand the level of experience that users already have with the existing software system. If users have prior experience with existing touchless technologies, such as foot pedal control for colonoscopies, more sophisticated touchless interactions may be a suitable design choice. Existing user experience may enable users to adapt to additional functionality when using a more functionally complete touchless system. However, in the case of users naive to using touchless modalities, such as mid-air gestures, it may be appropriate to design a touchless system that better aligns with more common touchless modalities, such as voice control. Even if users do not have significant experience with such input mechanisms in a medical context, the pervasive use of touchless or voice-driven consumer devices may help users feel more comfortable with touchless modalities in the more nuanced clinical setting.

6.2.2 The Operational Environment

Where will the system be used?

A clear understanding of the operational environment in which the system is used will guide a number of design decisions and can help the designer avoid unexpected issues. A detailed description of the relevant hospital environment should be considered early in the design process. For example, if a system is to be deployed in the OR, understanding the different zones within the OR and their layout (Figure 6.3, p.155) can inform the design process.

It may not be possible to fully determine all the environmental parameters and constraints. A system that will be mobile and move throughout the hospital is likely to be exposed to a greater range of environmental conditions than a static system that while normally fixed, can be relocated to other hospital locations. It is, however, essential that the best effort is made to understand the space and its limitations; contextual enquiries can be used to gain an understanding of operating context [2]. This is especially so when designing touchless systems, as the impact of the environment on user interaction and

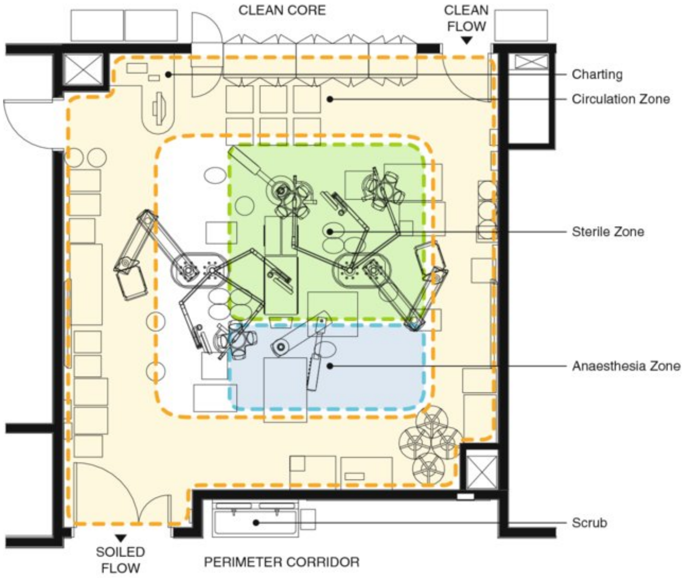


Figure 6.3: An example of the various surgical zones in an OR setting. There is a firm division between sterile and non-sterile zones [161]. ©2013 Stantec

system performance may be greater, often in unanticipated ways, than experienced with traditional systems.

Is the environment spatially restricted?

It is important to understand what spatial constraints will apply in the area(s) in which the system will be operated. Such constraints may be due either to limited physical space or to the number of people operating within the same area. Alternatively, if the system is located in a large area, this may exceed the operating range of the system sensors and lead to recognition errors. Depending on the system design, a larger area may be required [19]. Different touchless modalities may be more appropriate depending on how restricted or otherwise the space is. For example, active zone might not be a suitable modality in an area that is small, or in a congregated area, as may occur in a shared office space. Active zone requires one or more dedicated physical regions for the user to step in and out of. If such dedicated regions cannot be guaranteed, then gestures or voice commands may be more effective modes. In less spatially constrained environments, such as those used for interventional radiology, active zone would be an appropriate modality, especially when used as a clutching mechanism.

Is the space shared?

The impact of a user operating the system on the workflow of others should be considered. For example, voice commands may distract other people from performing their jobs; a crowded operative space may be a source of issues during system use [23]. However, if a space is shared by several people all working as a team, such as in the OR, this would be less of an issue. In a quieter working environment, such workflow interference will be reduced, for example, a radiologist within their own office can issue voice commands without having to worry about distracting anyone else.

What are the noise levels in the environment?

Another environmental factor to consider is how noisy the area around the system is likely to be. High levels of ambient noise and surrounding conversation can have a significant adverse impact on the performance of speech recognition [20], especially if the user is not directly adjacent to the system's microphone array. In a noisy environment the use of voice commands may be deprecated in the design, or modified to operate with microphones worn by the user for example. This could be achieved through individual headsets; however, this may be undesirable as these devices present maintenance and sterility overheads. Such devices may also not fit well from a user comfort or workflow perspective.

Does the system need to operate quietly?

It should also be considered if there are limitations to the sound levels associated with operating the system. For example, in an intensive care unit (ICU), there may be a requirement for the system to operate silently, especially during the night [162]. This could limit using voice commands. Feedback to the user by the system would also be best provided silently, perhaps through on-screen prompts.

Will the system be fixed in a single location or be mobile?

Mobility is another important consideration when designing a touchless system [48], i.e., whether the system will be located in a defined area, such as in the OR, or will it be mobile, such as a mobile ward computer. This will affect decisions regarding what hardware can be used. For example, a mobile system will be unable to utilize a room-scale fixed-camera setup, but will instead need to rely on the sensor(s) that move with the system.

Will network connectivity be available?

It is important to understand if a system will have access to the hospital networks and the Internet. A key question that will guide the design of a system is whether public cloud processing of touchless interactions will be permitted by the hospital. There are multiple powerful platforms that cloud technologies provide; these have the potential to bring very high levels of sophistication to the user experience, as well as extensive system capabilities. Cloud-based natural language processing tools provide a very powerful alternative that can be developed more rapidly than using local-only tools and libraries. Because processing occurs in the cloud platform, any improvements to vocabulary and accent processing capability, for example, is immediately available to all deployed systems. Another example of a cloud-enabled technology is facial recognition. Apart from being extremely performant, a cloud-based solution allows for user authentication and data processing to be performed across multiple devices as well as across multiple hospital sites. The facility and performance of cloud platforms will need to be balanced with hospitals remaining keen to store/process as much data local-only in order to maintain compliance with data privacy regulations such as GDPR (EU) and HIPAA (US) and to enhance security. Understanding what a particular hospital will permit is vital to the design decision to incorporate connected imaging systems with public cloud platforms.

6.2.3 System Requirements

What are the security needs of the system?

When designing any medical imaging system it is important to understand the sensitivity level of the data that will be accessed using the system. Such systems are generally used to access private patient data that is subject to high levels of security and privacy regulations. Existing systems, such as PACS, employ information security measures such as unique user profiles and strict auto-logout policies to help protect this data. Breaches of health information security can result in high financial costs and endanger patients [163]. By understanding the nature of the identity data and its sensitivity, appropriate levels of touchless security can be considered in the design, such as the use of facial recognition to replace traditional username and password authentication.

Will the system require auto-logout + when/why?

Hospitals usually have auto-logout requirements for computer systems, especially those used to access patient information. Such automatic logouts aim to minimise health information security risks [163]. It is important to understand what the security policy and auto-logout requirements a hospital expect in terms of time-to-logout, and what qualifies as session activity to prevent premature logout, e.g., can presence detection be used to keep a system active?

6.2.4 Training Requirements

Will users receive dedicated training?

It is useful in the qualified research phase to establish what attitudes users have towards training, and whether users will receive dedicated training with a novel touchless system or if they might be expected to 'learn it on the job', perhaps supported by input from a system champion, e.g., the lead surgeon [27]. Though some users may have experience with existing touchless technologies, such as foot pedals, it may be an unfamiliar, novel means of using medical software for many users. Time for extensive training sessions is often not available in a busy hospital and not thought desirable. Hospital administrators may prefer that new users learn how to operate the system through use, rather than having to allocate additional training hours. Learning techniques such as micro-lessons may be appropriate [164]. This insight is important to interface design. A system that must be 'self-evident' enough for users to learn to operate through use will require a simplified interface with user prompts to guide them.

6.3 Hardware Design Decisions: Selecting Appropriate Hardware

The first set of technical design decisions is related to hardware. If a touchless interface is being retrofitted into an existing system, the hardware choices available may be constrained at the design stage, e.g., an OR will likely already have monitors installed, which may predetermine factors such as screen size. It is important to investigate the technology available, their abilities, limitations, costs, and in what contexts they may be optimally used (Figure 6.4, p.159).

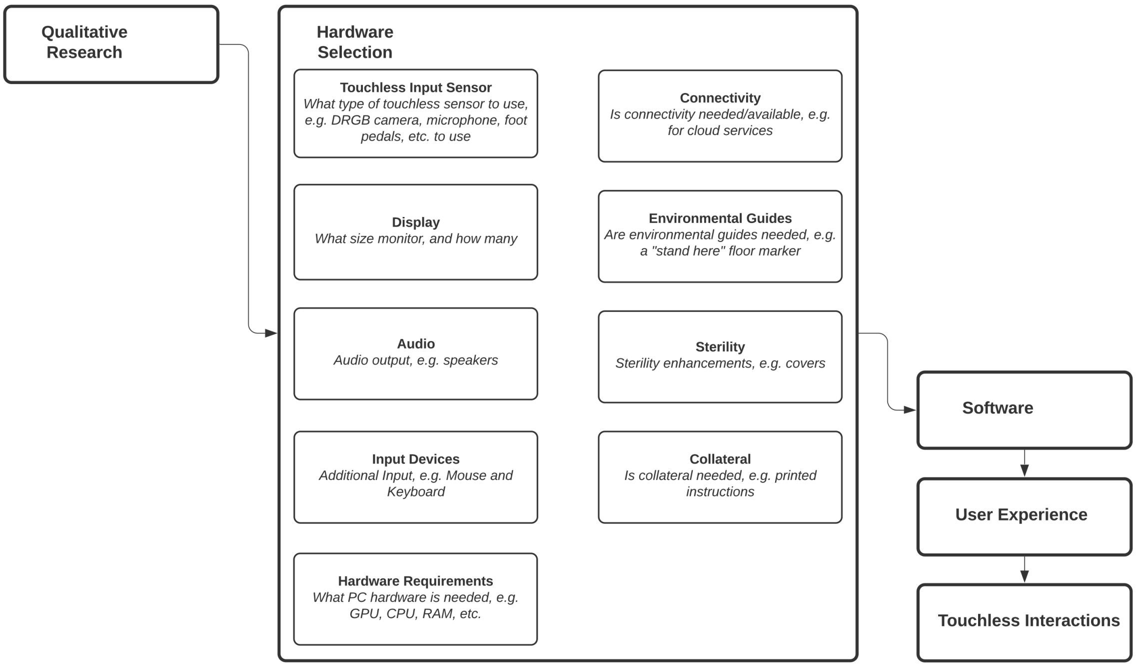


Figure 6.4: Hardware design path

6.3.1 Touchless Input Sensor

What class of touchless sensor to use?

The choice of touchless sensor is impacted by system requirements as well as hardware availability and cost. The gesture input set of the system will impact the choice of sensor based on the input modalities they enable. Mewes *et al.*'s chart (Figure 6.5, p.160) of touchless interaction methods provides a useful guide in this respect.

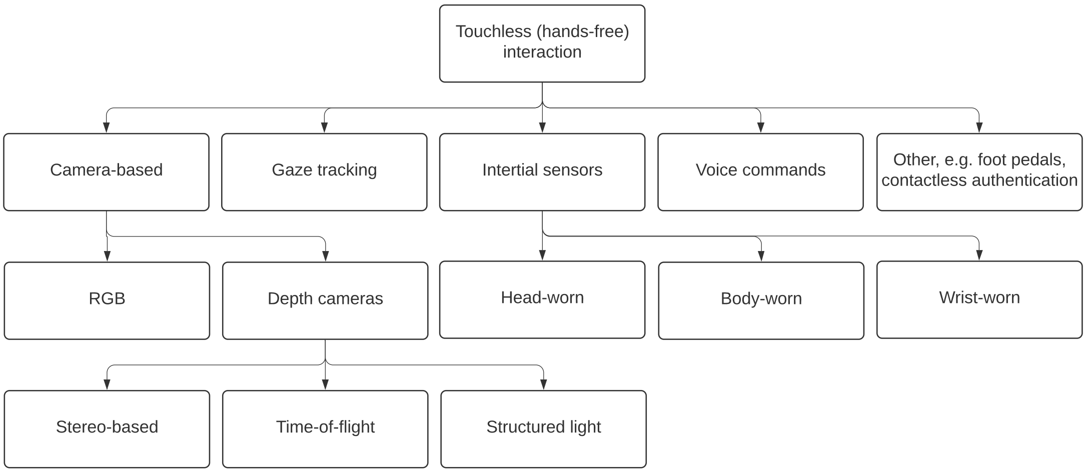


Figure 6.5: Mewes presents a chart showing various touchless interaction methods. [1]

For example, if audio input is required, a sensor with good quality microphone(s) should be used. When considering design related to audio input as a touchless interface there are several factors that must be addressed. These include context, e.g., available space, noise levels, audio input, user proximity to the sensor, cost, component availability, support life cycle, and the required level of computer hardware resources (e.g., high-powered PC, sound card). Commercially available sensors have been found to be suitable for use in hospital environments [28].

6.3.2 Display

What size display monitor, and how many?

The question of display monitor size and quantity should be considered as this will be an essential component for usability. This choice is impacted by a number of factors, firstly, how close to the system display the user is likely to be. Users may be farther from a touchless system than usual, as they will not be limited by access to a mouse and keyboard. To account for this additional distance, a larger display may support visibility from farther away, more like a kiosk design. However, this may not always be the case. If a touchless system is designed to be used close to the user, with touchless technology replacing a mouse and keyboard (the Leap Motion is a good example of such a device), an overly large monitor may prove overwhelming for users as they may not be able to see the entire screen at once.

Another design approach is through providing multiple displays [68]. One display could be used to provide touchless interaction feedback and information to the image viewer, while another primary display could show the user interface and images they are interacting with.

If an existing system is being adapted for touchless input, the monitor size may be pre-determined by existing equipment. If this is the case, any interface elements being designed should be sized accordingly to allow for good visibility while not occupying all of the available display space.

6.3.3 Audio

Audio output devices

A combination of factors, including ambient noise, sound level restrictions in quiet areas, the content to be carried, and the nature of user feedback, may require that audio output devices such as speakers are part of a system design. Audio feedback has been found to help novice users organize their thoughts [35]. It is important to understand if users will receive audio feedback from the system, e.g., a chime to confirm a gesture has been recognized. If the system will be used in areas with volume limitations, for example, at night in areas where patients are sleeping, it may be necessary to design silent feedback to the user, or provide a non-verbal touchless means of muting the system.

6.3.4 Input Devices

Additional input devices

When adding touchless interaction to existing systems, it is likely that interaction using mouse and keyboard will already be available and may continue to be required. For entirely new systems, provision of a conventional mouse and keyboard should be carefully considered as this may compromise the ability to sterilise the system contact surfaces. Sanitation of such surfaces can take up to twenty minutes each time [25]. This decision will depend on whether or not touchless interaction supports the complete operation of the system, furthermore, it may be necessary to provide traditional input, such as mouse and keyboard, in the event that touchless recognisers are unavailable. This hybrid approach would provide for touchless interaction for the most common input tasks, with fallback input for the less commonly used or cumbersome to perform functionality. This approach could be beneficial in terms of limiting touchless lexicon complexity, thereby decreasing the learning effort for users. A risk that must be considered is the possibility that providing a mouse and keyboard may have the effect of some users avoiding using touchless interactions at all. If users revert to using traditional interaction, this will defeat the improvement in aseptic use through touchless interaction. This may be an issue if improved asepsis is a design goal.

6.3.5 Computer Hardware Requirements

What computer hardware configuration is required?

It is important to understand the factors that impact this part of the design process.

Firstly, the recommended requirements of the touchless sensor should be considered. For example, does the sensor require GPU functionality via a dedicated graphics card? Will the system be performing intensive CPU-based calculations on the local hardware and require an advanced CPU? Different input devices will have significantly different requirements and capabilities and it is important to understand these during the design phase. The Azure Kinect DK used in this research provides a relevant example. At launch, this sensor required a dedicated GTX 1070 graphics card (or better); without such a dedicated graphics card, a standard desktop computer simply could not support the camera. Such hardware is not commonly found in more business-oriented devices, being more commonly found in gaming and media editing/production setups. Further, systems that come with fast processors and such dedicated graphics cards are generally more expensive per unit; hospitals are generally cost sensitive when it comes to devices such as ward-based computers and this additional cost may be undesirable. Increasingly, modern RGBD cameras are capable of performing much of the processing within the camera processors, reducing the load on and hardware cost of the attached computer.

Secondly, the operational environment should be considered, e.g., an overly loud computer may not be appropriate for an environment such as the ICU [162]. In this research, both a tower desktop PC with a separate display and an all-in-one PC with an integrated display were tested. Though nearly identical in capability, one key difference observed between these two devices during testing was air flow and fan size. The restricted air flow and smaller cooling fans within the all-in-one PC chassis meant, that during intensive use, e.g., when the Kinect 3D scene data was being processed, the fan noise level was considerably higher for the all-in-one when compared to the tower desktop computer. If a device is to be used in an environment where noise levels need to be managed, an all-in-one device (or any device with restricted thermals and smaller fans) may not be appropriate. Instead, the designer may choose to use a system with larger, quieter fans and greater cooling airflow. In an environment where space is more limited, an all-in-one PC may well be preferable as this compact unit can be wall-mounted, attached to a movable stand, or can be placed on a desk with minimal footprint.

Thirdly, it is important to consider resource constraints. It is unrealistic to expect the hospital to provide high-end computing resources, costing thousands of euro. The reduction of costs as a benefit been discussed in research such as [44] and [45]. If designing a device that is intended to be deployed across multiple locations in a hospital, design decisions will need to balance performance (and user experience) against cost. For this reason, the ability to off-load much of the gesture, voice, and facial recognition processing to secure cloud platforms may be essential to reducing the requirement for sophisticated and costly local computer systems.

6.3.6 Connectivity

Is connectivity needed/available?

The design options for a touchless system with and without connectivity can be quite different. As a designer the question must be asked 'is connectivity available and is it needed?'. If connectivity is not available (either due to lack of infrastructure or due to hospital policy), then a fully offline design should be adopted. However, if connectivity is available, the designer should consider if it is needed to deliver a reliable system or to improve the UX of a touchless system.

Cloud-based processing for touchless interactions has the potential to dramatically improve system performance and UX, especially on systems using lower-end hardware. Provided a robust and performant Internet connection is available, cloud technologies can enable sophisticated natural language processing, along with advanced computer vision processing (for applications such as facial recognition). Ideally, it is best to use wired connectivity for any stationary system. This is the optimal choice for reliable connectivity, improved data rates, and reduced latency, which is key as users are often intolerant of slow system response times. In cases where a wired connection is not an option, either due to a lack of infrastructure or because a system needs to be mobile; available wireless options, such as GSM (4G/5G), WiFi, or Bluetooth, will need to be considered.

6.3.7 Environmental Guides

Are physical environmental guides needed?

A system design may include elements such as in-world guide markers for user position, to indicate active zone, for example. These elements can be useful in certain situations, by removing ambiguity for the user regarding where to position themselves. However, these elements may increase the complexity of installing a touchless system by introducing additional elements, such as floor marking tape suitable for heavily used OR environments.

6.3.8 Sterility

Asepsis of non-touchless elements

One perceived benefit of touchless systems is the reduction in cross contamination. Design of a touchless system that is to be shared by multiple users must consider how to enhance the sterility of the non-touchless elements of the system. If a mouse and keyboard are part of a touchless system, either due to being part of existing hardware or as a fallback, it may be worth considering the inclusion of asepsis enhancements, such as a plastic wipeable cover (Figure 6.6, p.165) for the keyboard, as computers and their peripherals are difficult to sterilise [22].

Figure 6.6: A removable, wipeable keyboard cover

6.3.9 Collateral

Is illustrative collateral required?

While on-screen instructions are valuable, there may also be a need to provide additional printed collateral to inform users regarding a system. Such collateral may come in the form of wall-mounted informational posters, user guides providing comprehensive instruction, or 'cheat sheet' print-outs listing the most common touchless commands for a system. Where possible, a system should be designed not to require such additional resources, however, a hospital may require the inclusion of this material.

6.4 Software Design Decisions: Selecting Appropriate Software Tools and Development Environments

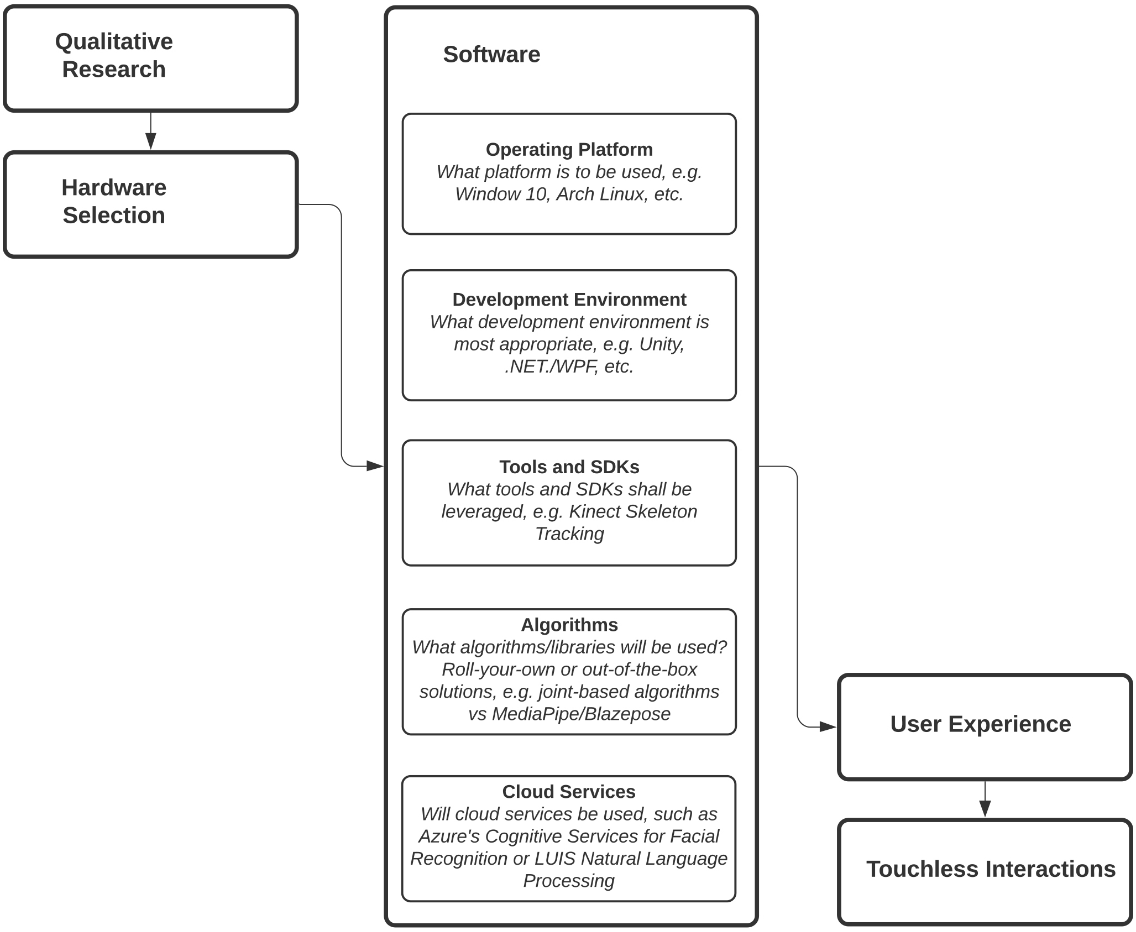


Figure 6.7: Software design path

6.4.1 Operating System Platform

What operating system platform will be used?

Certain sensors may only work to their full capability (or at all) in the context of a limited set of operating system platforms. A good example of this is the Microsoft Kinect. The primary target operating system platform for this device is their own Microsoft Windows operating system. As a result, tools such as SDKs, skeleton tracking applications, and various libraries are mostly available only for the Windows operating system. This may limit the design decision to the use of a Windows operating system, rather than alternative Unix-based operating systems such Mac OS or Linux.

It is also important to consider the security capabilities of different platforms and development tools. Hospitals process a significant amount of sensitive patient information and data security is important to consider—[163], especially if any touchless features are leveraging cloud-based processing. Understanding what security features are required during design and development is essential, as this is likely to be challenged by a hospital IT department. System design should include mechanisms for managing software updates, feature updates, security patches, and firmware updates, as well as complying with applicable security policies by design.

6.4.2 Development Environment

What development environment is most appropriate?

Bespoke or proprietary technologies may require specific development workbenches, e.g., if they use embedded operating systems based on Linux. Generally, there is little need to change from a development team's usual development environment as common programming languages, revision management, and collaboration tools can be used.

6.4.3 Tools, SDKs, and Algorithms

What tools and SDKs will be leveraged? Create-your-own or out-of-the-box solutions?

During the design stage, it is valuable to consider what tools are available for the processing of touchless interactions and their capabilities. This is required to determine if it is possible to use existing software solutions or if a custom solution needs to be developed. Services such as Google's MediaPipe or NVIDIA DeepStream offer significant performance benefits in terms of performing human tracking using lightweight hardware. However, it is important to research these solutions and understand their strengths and limitations, as well as their licensing requirements.

It should be considered whether an existing or a custom tool will allow the provision of a better user experience. Some touchless modalities are already supported by highly robust solutions. For example, solutions to the problem space of natural language processing have been developed and supported for a considerable time with tools such as Dragon Speech Recognition or LUIS available today. However, even though the available solutions for gesture recognition continue to expand and improve, a hybrid approach using existing and custom software may be needed for gesture-based input analysis.

If an existing tool can accelerate the development process, it may prove a compelling approach, especially if a touchless system needs to be adapted to various environments and touchless modalities. By using modular components it is possible to develop a touchless interface much more efficiently than by coding custom solutions. However, this improvement to code release speed should be balanced against the life-cycle support of external tools. External tools do risk being made obsolete, or abandoned, by their developers so it is important, therefore, to choose tools that will be supported for as long as possible through the life cycle of the touchless system. This is particularly important for hospital systems, which usually have duty lifetimes of a decade or more.

6.4.4 Cloud Services

Will cloud services be used, such as Azure's Cognitive Services for Facial Recognition or LUIS Natural Language Processing?

Before committing to including cloud technologies in the design of a touchless system it is vital to understand the hospital administration position on such technologies. In order to protect the data of their staff and patients, hospitals may either limit or outright reject what processing can be performed in the public cloud. For example, a hospital may permit natural language processing in the cloud, but insist that facial recognition for authentication processing be performed locally due to the sensitivity of staff biometric data. Several factors affect the adoption of cloud-based processing in hospitals, including; data security, enabling technology, human factors, organisational factors, and environment [165]

When considering cloud technologies for a touchless system, the designer should consider if processing can be executed locally with equivalent performance. If this is the case, local processing should generally be considered the preferable design choice as it will be more reliable and have lower lag times, resulting in a better user experience. Always consider whether a processing task can be performed offline. If cloud processing platforms can provide levels of functionality that cannot be replicated locally, that afford HPs superior use of the touchless system as part of their workflow then they are worth including in the design if possible.

Systems dependent on cloud services risk that a loss in Internet connectivity may prevent users from operating the system. This could have serious consequences if it were to happen at a critical point in a medical procedure. It is essential to design a system to be resilient to this possibility. This can be achieved by including appropriate messaging to the user when connectivity is poor, as well as supporting limited offline interaction processing, or failover operation using touch-based input devices.

6.5 User Experience Design Decisions

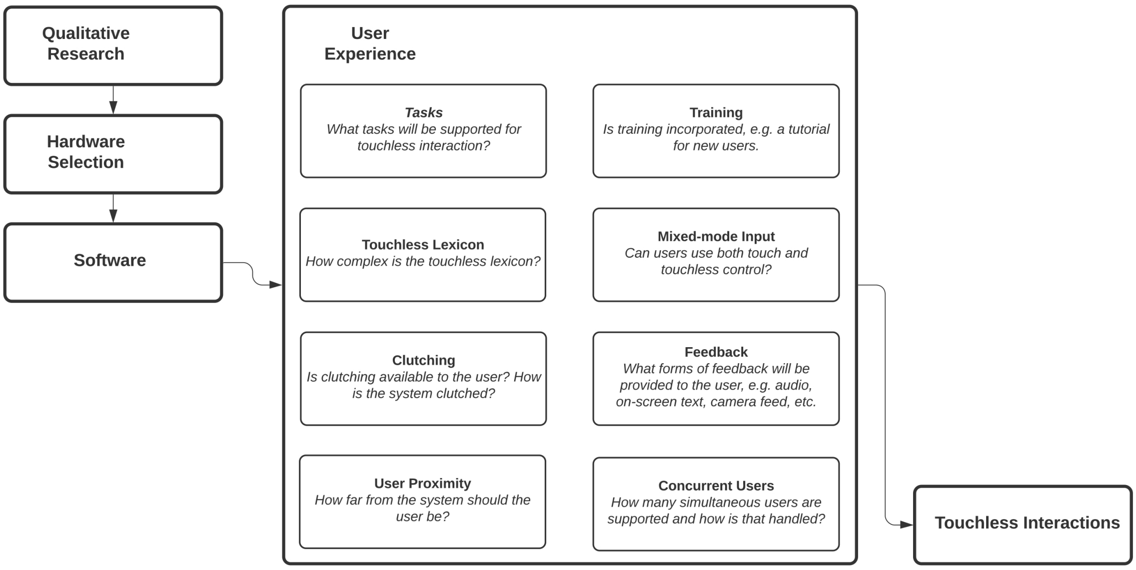


Figure 6.8: User experience design path

6.5.1 Tasks and Touchless Lexicon

What tasks will be supported through touchless interaction?

It is important for the designer to gain a comprehensive understanding of the functionality of the base PACS software (see subsection 2.2.2, p.17). This information may then be combined with an understanding of the most common functions used by the target users, to develop a core set of operations that need to be supported using touchless interactions. Experiences with the prototype systems employed in this dissertation suggest that it will be possible to deliver the majority of the required PACS viewer functionality through touchless input, by careful selection and consideration of which functions will deliver the greatest benefit to HPs.

There are a number of reasons for limiting the complexity of the touchless lexicon that a designer should be aware of. First, not all touchless interactions are equally robust; an overly large lexicon may impact a system's ability to discern between gestures [48]. Some gestures will be more reliably detected by a sensor than others. It is important to test various inputs to determine a recognition confidence index for each gesture to identify which are reliable and which are prone to failure. This will guide the design process by giving the designer a measure of what touchless input types can reliably be supported with a given sensor setup.

Second, limiting the size of the touchless lexicon can help to reduce the learning effort for users. This may be achieved by focusing on providing core functionality required by the users, thereby avoiding overwhelming the user with a large set of novel interactions that users may find challenging to remember. This is more likely to apply to gesture-based systems where users can take longer to become familiar with a system. The corollary to a limited lexicon set, is that it may be limiting for advanced users who require deeper levels of functionality.

6.5.2 Clutching

Is clutching required? How is the system to be clutched?

Clutching should be included as a necessary function in touchless medical systems [37, 75]. Though clutching will be used more extensively in some workflows than others, it brings multiple benefits in terms of user experience, as well as minimizing errors, protecting data and privacy, e.g., by preventing accidental input from opening another patient's file or leaving a record open on-screen while attending to other tasks.

There are two primary input categories to select from, automatic and discrete. Examples of automatic inputs include gaze detection and active zone, whereas inputs from gestures, voice commands, and foot pedals are examples of discrete input. Where clutching is to be used, the automatic method of active zone clutching may often be the most effective choice, integrating easily with most user workflows. However, if active zone is not possible as an input, perhaps due to space constraints or sensor choice, discrete clutching remains a good option from a user experience perspective.

As a core usability criterion, it is important that clutching and unclutching a system using touchless input be as intuitive to the user, and reliable as possible.

6.5.3 User Proximity

How far from the system is the user likely to be?

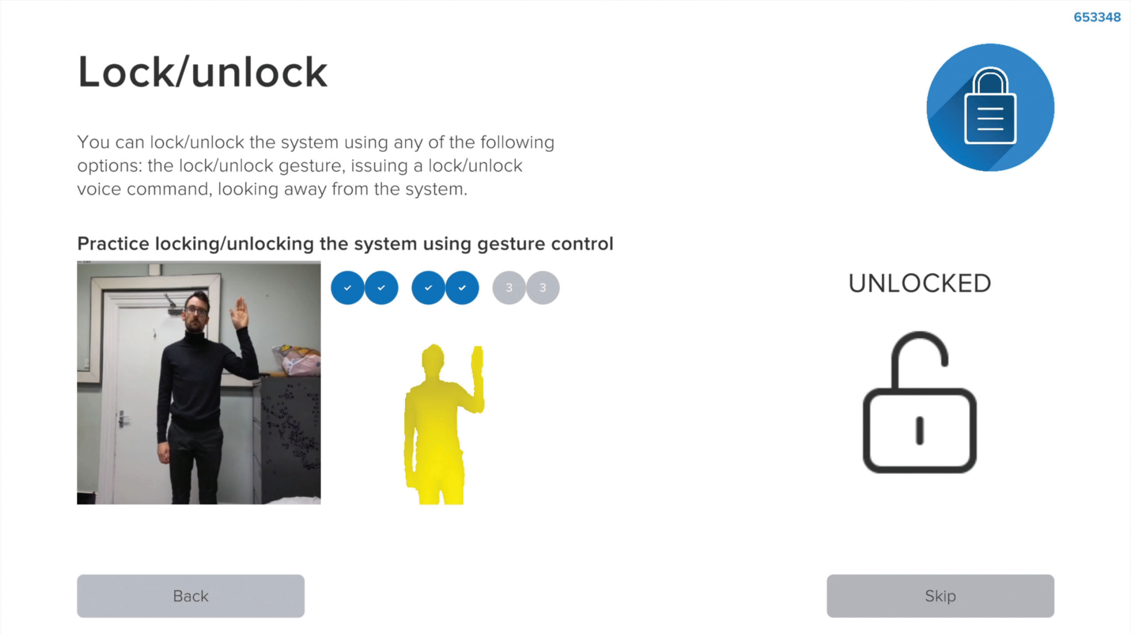
Depending on the choice of sensor, a user may need to be within a specified location to interact with a touchless system. Devices such as the Leap Motion require the user's hands to be directly above the sensor to operate. In comparison, devices such as the Azure Kinect DK have wide fields of view and can allow the user to be positioned in a range of locations. Furthermore, gesture recognition performance can be impacted by distance from the sensor [17]. The user's workflow will likely influence the choice of sensor depending on the degrees of freedom position and movement they require.

6.5.4 User Training

Is user training incorporated?

It is important to consider how user training will be conducted when it comes to touchless systems. Providing appropriate training/onboarding will help ensure a positive initial user experience, helping build user confidence in a novel technology [36].

There are various approaches available to delivering this training. Two suitable options to consider are; conducting a comprehensive single training session (as was implemented in this work due to the constraints of the experimental context, see Figure 6.9, p.172), or use of micro-learning training that is delivered contextually, as discussed in Lacey *et al.* [164]. Though delivering full system training in a single session is easier to envision and implement, it will generally result in a poorer user outcome (usually attributed to information overload). Delivering too much new information at once, risks users becoming overwhelmed and only retaining a portion of the information. This may result in users avoiding certain functions and operations that will be perceived as unclear, or difficult to perform.



Lock/unlock

You can lock/unlock the system using any of the following options: the lock/unlock gesture, issuing a lock/unlock voice command, looking away from the system.

Practice locking/unlocking the system using gesture control

Figure 6.9: Training was integrated into the experimental prototype system. Here the user is shown the gesture/task combination; there is a demo video for them to imitate, they are shown a real-time representation of themselves to better understand what they system 'sees', and they can observe the outcome of their action with strong visual feedback.

In contrast, delivering training through micro-learning may allow users to incrementally build their competence with a touchless system. Rather than attempting to memorize all available functionality at once, they instead will be trained on functions that are appropriate to their workflow. The question of how best to design this training is beyond the scope of this section, but drawing on learning theory and micro-learning best practices, is highly recommended.

6.5.5 Mixed-mode Input

Can a hybrid system providing both touch and touchless control be used?

Depending on the level of sterility required, the question of whether users can operate the system using both touchless interactions and traditional touch interactions should be considered. For highly sterile environments, such as the OR, a designer may adopt for a strictly touchless approach. However, this may not fit with users' desired workflows. Surgeons often dictate instructions to other clinical staff to be input via mouse and keyboard on their behalf. Some users may desire to retain this ability even with a touchless system.

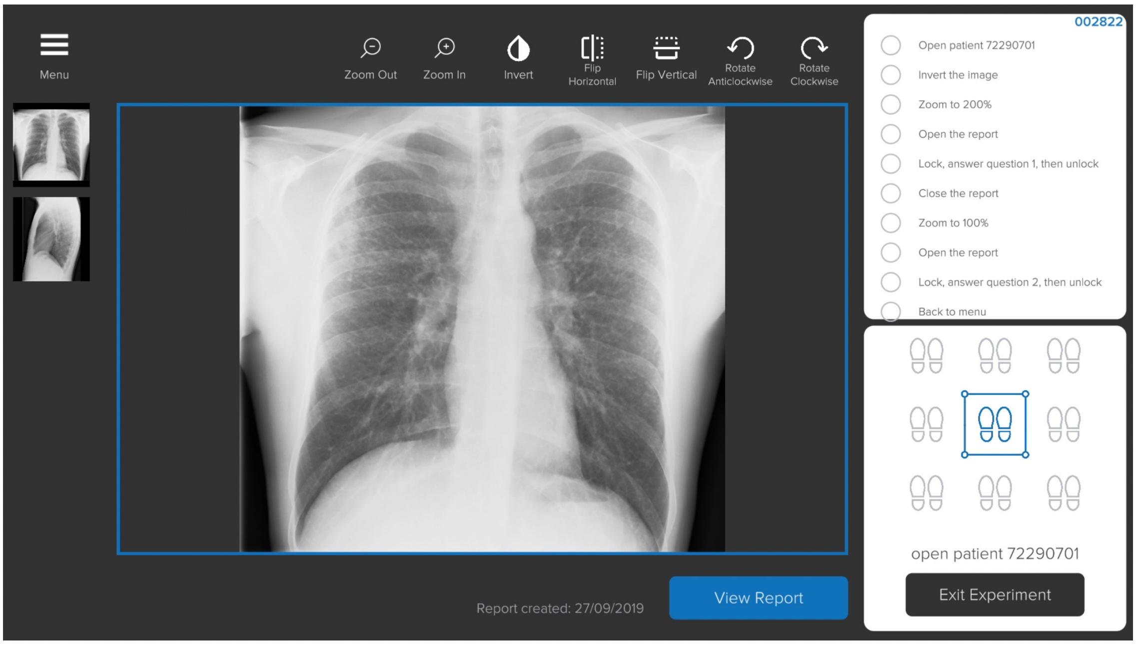
Other users may prefer a hybrid approach where they can use a combination of mouse and keyboard and touchless input simultaneously. For example, radiologists may want to use voice commands to trigger macros, in conjunction with a mouse and keyboard for tasks such as zooming, etc. In this situation, touchless interaction can help improve a user's workflow without concern for sterility.

6.5.6 User Feedback

What forms of feedback will be provided to the user?

Providing appropriate and timely feedback to the user is a key part of ensuring a positive user experience with a touchless system, helping to improve user acceptance [1, 74]. Users are more forgiving of errors, and feel less lost when clear feedback is provided. Properly designed feedback improves user confidence with the touchless elements of a system. Feedback can also be used to confirm significant actions, such as issuing a 'delete' command, or when opening a patient file to confirm that the correct patient record has been selected.

There are several available mechanisms for providing feedback to the user. The most immediate is the feedback on the display, either on the primary monitor or on an adjacent secondary monitor (as in [60, 65]). Feedback is generally most successful when the user is presented with a real-time view of what the system is perceiving, as well as prompts regarding successful and unsuccessful attempts at touchless control. By providing this real-time system view of the operator, a user can feel confident they are in the correct location, are speaking loudly enough, etc. Feedback related to unsuccessful interactions allows the user to attempt those interactions again without undue delay. If possible, feedback regarding how the operator can correct interaction failures would be desirable, though this feedback can be difficult to generate in practice.



002822

○ Open patient 72290701
○ Invert the image
○ Zoom to 200%
○ Open the report
○ Lock, answer question 1, then unlock
○ Close the report
○ Zoom to 100%
○ Open the report
○ Lock, answer question 2, then unlock
○ Back to menu

open patient 72290701

Report created: 27/09/2019

Figure 6.10: A mini-map showing the user's position relative to the active zone provided clear feedback to the user, allowing them easily position themselves to operate the prototype system.

Audio feedback is the most natural alternative to on-screen feedback as users are already familiar with many sources of audio feedback. However, hospital environments may be either too loud for audio feedback, or require devices to operate silently.

Though common in handheld devices, haptic feedback is less straightforward to implement in the context of a touchless medical imaging system. Technical solutions do exist that could be used to provide haptic feedback, such as haptic gloves, or floor-mounted rumble motors, but neither are common in hospitals. It is also preferable when designing a touchless system to minimize the amount of hardware the user must wear; such equipment may encumber the user and is at risk of failing or being misplaced. Technology that enables touchless ultrasonic feedback has become possible in recent years. Rakkolainen *et al.* present a compelling survey of the applications of mid-air ultrasound haptics, including feedback solutions such as the ability to evoke the sensation of touching something in mid-air by focusing ultrasound from a phased array of transducers [137, 138]. Such technologies offer a compelling option for providing feedback to the user, however, they are not yet mature enough for real-world application.

Can multiple simultaneous touchless modalities be supported?

Limiting the number of input modalities can reduce system complexity, both in terms of hardware and software. From this research, it was observed that in general, users settle on using a single input modality for most interactions, even when more than one modality is available. It should be noted that though users tended to use a single modality, not all users gravitated to the same touchless input. Other work has investigated multi-modal touchless input, combining voice commands and gesture control [68].

It may be necessary to consider a combined approach, e.g., using voice command to select a mode, such as zooming, and gestures to use that mode. This approach would be beneficial in terms of reducing the complexity of gestures, which could reduce learning effort for users.

6.5.7 Concurrent Users

How many concurrent users are supported and how is that handled?

Analysis of the design objective workflow will help determine how many concurrent users within a session are likely to be supported. Most systems will only need to support a single user input in a session, similar to conventional input where there is a single mouse and keyboard for a device. However, there are instances where it may be desirable to provide support for input from more than one user. The OR is a good example where it may be desirable to allow several surgeons to participate and control the system. If the primary surgeon is unable to interact with the system, due to both hands being busy, another surgeon may interact with the system on their behalf. This would also allow an expert user to control the system without the need to descrub.

6.6 Touchless Design Decisions: Selecting and Designing Appropriate Touchless Modalities and Interactions

It is of benefit to understand whether the touchless system will use a single touchless modality or be multi-modal. This will be heavily dependent on sensor capability and performance, with different sensors supporting different numbers of touchless modalities.

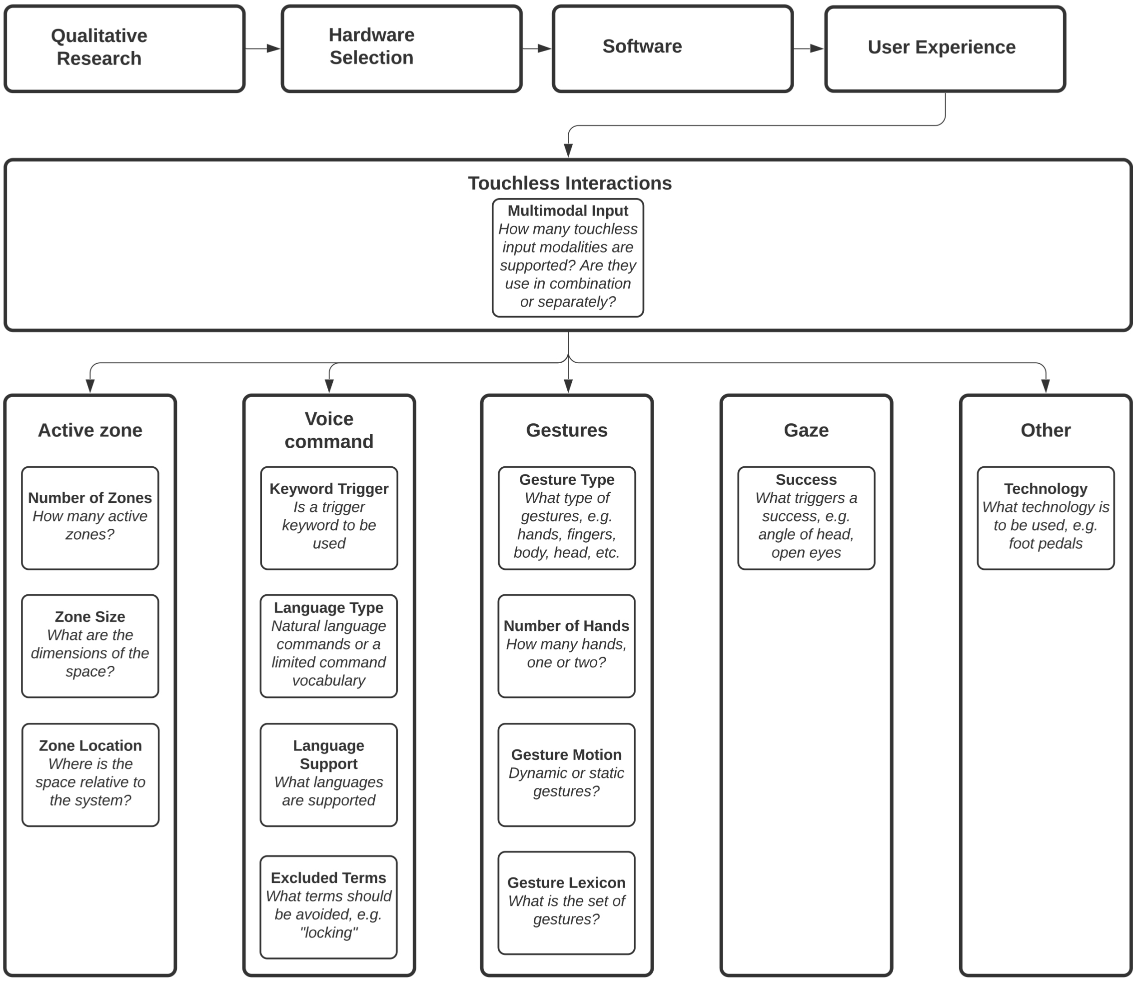


Figure 6.11: Designing touchless interaction modalities

6.6.1 Multi-modal Input

How many touchless input modalities are supported? Are they used in combination or separately?

It may be necessary to adopt a multi-modal approach to touchless interaction. This can be approached by differentiating touchless modalities for different tasks (as in Nishihori *et al.*), by using a combination of touchless modalities simultaneously (as in Park *et al.*), or by allowing the user to choose between touchless modalities.

Nishihori *et al.* demonstrated such multi-modal interaction by combining voice and gesture commands. In their solution, the nature of the task affected the choice of touchless interaction. "The right hand was assigned for analog operations, such as image enlargement and rotation, and the operation mode was set to the left hand. Voice input was used for operations such saving and reading the current coordinates and for switching the viewpoint direction" [68]. This approach acknowledges the relative strengths of different touchless interaction types, with gestures being better suited to continuous (or analogue) input, such as zooming, and voice commands being better suited to discrete commands, such as performing a save record operation.

Park *et al.* present a system that allows users to combine two touchless modalities; foot pedal with hand gestures [61]. This allowed users to perform interactions that, without the foot pedal, would require both hands, which is not always possible in a hands-busy setting.

6.6.2 Active Zone

How many active zones are required?

Depending on the operational context, a touchless system incorporating active zones may be designed to use one or more active zones. A single active zone is likely to be the most common scenario, with most systems being used by a single user in a single location. However, there are situations where multiple active zones may be beneficial, either to allow a single user to operate the system from several key locations, or within the space to facilitate interaction from multiple simultaneous users. These scenarios require different design considerations.

If a system allows only one user in a session, but has multiple active zones, there may be problems if someone other than the user (aka a non-user) has erroneously entered an active zone not occupied by the user. Communicating the current state of the system clearly to the operator can help to mitigate for this. The level of system complexity is increased when multiple simultaneous users can interact within a session. A designer needs to consider how many users are to be supported, as well as whether input from all users is to be treated with equal priority. Designing a multi-user session touchless system should be performed with a complete understanding of the real-world workflow that it will support.

What are the dimensions of the active zone region?

Depending on the environment and the use case, different sizes of active zone regions may be considered. In this research, it was found that a region of 1m^2 was an acceptable working area for users. This region gave participants sufficient space so as to not feel restricted, while allowing them to easily leave the active zone in any given direction using a single step. Restricting an active zone to less than 1m^2 present in ergonomic challenges to the user. Larger regions may be considered, such as an enlarged active zone covering key surgical space in the OR. When using such larger areas, it is important to consider the risk of accidental activation by non-users entering the region. This is especially true where overall space is limited relative to the active zone and users/non-users are unable to avoid triggering a system unintentionally.

Where is the operating space relative to the system?

The proximity of the operating space relative to the system may impact the design of the GUI. If the operating position is sufficiently far from the system interface/sensor, traditional GUI design may result in an interface that is difficult for the user to read. It may also be valuable to understand if the active user is likely to be in a fixed location relative to, but not adjacent to, the system, e.g., in the OR, as this may impact the choice of supported touchless modalities. For example, if the user is likely to be in a fixed location, active zone may be an appropriate clutching mechanism to support, whereas, if the user is highly mobile, an alternative clutching mechanism may be more suitable.

6.6.3 Voice Command

Is a trigger keyword required?

When designing a voice command vocabulary, it is worth considering the inclusion of a trigger command [47]. Often HPs will switch between communicating with colleagues and controlling the system. Depending on the frequency that users switch between these activities, it may be that system-wide clutching may not be sufficient. Per command clutching, i.e., trigger commands can provide users with more granular control of the system.

There are situations where having to issue a trigger command every time may be frustrating for the user. An example of this could be a radiologist using the system in their office. Here the user may wish to issue a series of commands in succession. They are also unlikely to be talking to another person during these operations. In this instance, being able to simply unlock the system, issue the command sequence, and then lock the system again may be the most suitable design choice.

Natural language commands or a limited command vocabulary?

When considering voice control, there is a choice to be made between providing full natural language control or a limited vocabulary of commands. It was observed in this research study that even when natural language interaction was available, many users gravitated to using primary command terms, such as 'lock' or 'rotate'. Implementing a system with a limited set of commands may be preferable as it is more amenable to offline processing. Two-word commands may be more meaningful to the user than single-word input [40].

An area where natural language processing may be valuable, is in completing more complex tasks such as opening patient records, or construction of compound commands, such as "open study X and study Y".

What spoken languages are supported?

Though a single language is usually sufficient for operating a touchless system, support for more than one language is required when users cannot use the system through the primary interface language with competence. Providing speech recognition for multiple languages with similar levels of recognition performance may be a significant implementation challenge, requiring access to cloud speech processing functionality.

What commands should be avoided?

When considering suitable voice commands for clutching a system, careful consideration must be given to avoiding domain-specific words, such as 'lock' and 'unlock', which may seem obvious choices, but may be used by medical users to have other meanings. For example, in an orthopaedic environment, existing workflows may contain references to 'locking screws'. The similarity between these terms and the command 'lock' may cause problems both for the user and the system. A system can be designed to avoid commands that overlap with existing operational vocabulary if this requirement is established during the qualitative research phase of the design. Care should also be taken to ensure that commands are sufficiently distinct from each other phonetically to avoid overlap and misrecognition [46].

6.6.4 Gestures

What type of gestures are supported?

Gestures that use fingers, hands or arms are the most common. They are more straightforward for users to perform than full-body gestures. It is important to consider the context in which a user is to use a gesture. For example, during surgical procedures, gestures should be uncomplicated and practical, such as opening the palm of the hand, and not require the user to look away from the surgical display [66, 68].

Although finger-based gestures can be easy for the user to perform and useful to convey information such as numerical values, many sensors and their associated recognisers are not designed for finger detection. Choosing to use finger-based gestures will limit the range of sensors available when designing a TMIS, as finger gestures are potentially difficult to detect reliably at different angles and distances [65]. Modern tool kits such as BlazePose are designed to perform robust full body detection but do not currently provide sufficient finger level resolution to be useful in real-world applications.

Other types of gestures can be considered, such as head or full body-based gestures, however, they are more niche. Head gestures may be useful in situations where both of a user's hands are busy. Full body gestures should generally be avoided as they have too great an impact on a users ability to perform other tasks. Such options should be considered when there is a compelling reason not to use hand/arm gestures, while also requiring a gesture-based interface.

How many hands for the gesture, one or two?

When designing the gesture set for a system that uses hand/arm based gestures, the designer may choose to use one-handed gestures, two-handed gestures, or both. The advantage to one-handed gestures is that they allow the user to perform the gesture while still using their other hand. This may be important in hands-busy contexts such as the OR. Further, in Hassan *et al.*, users suggested a preference for single-handed gestures [55]. In contrast, using two-handed gestures can allow for a greater range of inputs that may be appropriate for particular inputs. According to Grandhi *et al.*, two-handed gestures should be used for the manipulation of objects, with the non-dominant hand providing a frame of reference and the dominant hand performing the transitive gesture [141]. By augmenting one-handed gestures with a second hand, it is possible to increase the range of inputs dramatically. Furthermore, in some instances, two-handed gestures may be easier to detect as they provide a larger target for the sensor to detect. A possible design choice may be to use one-handed gestures for most commonly used system functions, reserving two-handed gestures for less commonly used functions; this would allow a more comprehensive gesture lexicon while allowing the majority of users to use a simpler set of one-handed gestures.

Dynamic or static gestures?

Depending on the nature of the task, it may be more appropriate to choose either a dynamic or a static gesture. Dynamic gestures are best suited to continuous adjustment, e.g., zooming or panning. This is also true where a gesture is manipulating an object on the display, such as a 3D scan; in this case, the gesture should be a dynamic iconic representation of the motion required for the manipulation [141]. Static gestures are more appropriate for toggle or discrete tasks, e.g., inverting an image or opening a report. Dynamic gestures demand a greater level of user engagement than static gestures.

What is the gesture set?

When designing a gesture set, looking to the existing literature is a recommended starting point (Table 2.6, p.23). Hassan *et al.* note that any gesture set should be intuitive, with most users able to learn and accept a smaller set of gestures [55]. Many sets of gestures have been tested and documented in existing research and may provide a useful resource when compiling a design gesture set. It is also worth investigating what gestures are supported by any of the tools being used, such as BlazePose. Often these 'built-in' gestures will have been tested and optimized and can provide reliable performance and reduced system development times.

It is essential to test any candidate gestures thoroughly in order to understand how well they perform with a given sensor. An overly large gesture set can lead to gesture overlap and reduce a system's ability to discern between different gestures [55]. Generally, it is preferable to prioritize gestures that perform reliably, allocating such gestures to key functionality. Users are generally happy to learn a set of gestures provided they feel these gestures feel natural, work well, and can be relied upon.

6.6.5 Gaze

When using gaze-based interaction, it is important to consider how and when to keep a system properly calibrated. As successful eye-based gaze control is dependent on a high degree of sensor accuracy, Chatelain *et al.* propose re-calibration once every 24 days, with the goal of maintaining a mean accuracy of 3 degrees with a 95% confidence [72]. Fortunately, modern eye-tracking cameras are easy to install, calibrate, and use [49]. The requirement for such calibration is not necessary for solutions that use head position as a proxy for gaze.

What triggers a successful input?

Depending on factors such as sensor performance, and user distance from the sensor, the designer may choose either to use the operator's eyes or head to determine gaze direction.

If using head angle as a metric for gaze, it is important to consider sensor placement. Should the sensor be positioned adjacent to the screen, in which case the angle of the face relative to the sensor is directly equivalent to gaze at the screen, or if the sensor is offset from the display, in which case appropriate modifications are needed to calculate head angle relative to the display. It may be preferable to keep the sensor close to the display to simplify implementation.

6.6.6 Other Touchless Modalities

What technology is to be used?

When considering alternative touchless technologies which are not directly discussed herein, there are three key steps to approaching the design process. First, it is important to understand the technology, its capabilities, strengths, and weaknesses. Researchers have investigated a broad range of touchless control technologies (subsection 2.2.3, p.21).

Second, it is key to seek the experiences of existing users of the technology, if possible. Talking to experienced users will give deeper insight and help to prevent avoidable errors at the design stage. If it is possible to talk to users who use similar workflows to the design requirement, this would also be a valuable approach.

Third, it is useful to ask 'why use this technology?'. Sometimes, the touchless technology may already be available in a hospital, such as foot pedals or 3D input gloves. A hospital that has invested in such technologies may require a product that works with their existing infrastructure or is it likely to be familiar to their users. Alternatively, these technologies may simply be more appropriate for a given task or context. For example, foot pedals make a lot of sense when both a user's hands are occupied and are commonly used in catheterisation labs to operate motorised X-ray detectors.

6.7 Conclusion

In this chapter, the design process has been decomposed into a number of identifiable and manageable stages [139]; qualitative research, hardware selection, software decisions, user experience decisions, and touchless interactions. The design process for successful touchless interfaces for hospitals is more complex than may be immediately apparent. The discussions presented in this chapter aim to reduce this complexity for future researchers and developers by allowing them to anticipate decisions, trade-offs, and challenges. Each of these decisions and trade-offs needs to be considered within the context of the user and the environment. It is important to develop an appropriate understanding of the context in order to fully understand the user cohort needs, and how to best approach the various decisions required for such a system.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

Touchless input is compelling for clinical user interfaces because touchless input modalities such as gesture and speech can allow health professionals to interact directly while maintaining asepsis. The need for sterility has vital implications for interaction and the role of computing systems in clinical practice, e.g., as seen in the qualitative analysis presented in this dissertation and in prior published work on this topic. Touchless interaction has the potential to improve user experience and, more importantly, clinical workflow through sterile interaction and the technologies that enable this are becoming increasingly capable and deployment-ready.

The literature review indicates a gradual move from concern regarding technical implementation difficulties towards more fundamental aspects such as the design of gesture languages and the potential impact of touchless systems on medical practice, particularly in the OR. It is clear that whilst progress has been made in the field, the literature does not support any particular instance of the technology being mature enough to gain widespread acceptance or adoption. There have been performance improvements, and commoditisation of touchless technologies. These capabilities have not yet been fully exploited in the medical context. The literature supports the technical feasibility of using these types of systems, and provides a valuable corpus of studies of imaging-related tasks in the OR.

The literature is limited by the lack of more extensive studies and ecologically valid evaluations which represents a significant barrier to going beyond proof of principle, to adoption. Providing benchmark tasks for particular contexts would support comparative studies, particularly in the context of future studies examining surface contamination as an outcome. While there is an understandable focus on application in the OR and interventional radiology environments as the most frequently examined use cases, the use of touchless systems in other areas within the hospital context should also be explored.

The development of two prototype medical image systems described in this thesis using different generations of sensor provides additional context and perspective on the process of designing such systems. There was a significant difference in capability between the two generations of Kinect; from the perspective of a system designer, this allowed for more rapid development of various features, as well as the inclusion of functionality and performance that was not previously possible. It is now technically feasible for a single developer to develop a functional TMIS for research purposes; however, there remain a large number of design decisions that need to be addressed. As technical capabilities continue to improve, the range of functions that can be supported also increases. Though this allows for more advanced TMIS, supporting a greater range of user and environmental needs, it also increases the complexity of designing these systems. For example, as it becomes easier to provide multi-modal, multi-zone, or multi-user capability, the design complexity is increased.

This work also explored the user experience of clinicians interacting with a TMIS, and took a critical first look at clutching interactions for initiating and directing interaction within a touchless PACS interface paradigm. Clutching helps to avoid false-positive gesture and speech recognition, a key aspect of reliable touchless input. Four clutching techniques were evaluated within a touchless PACS interface. The sample of thirty-four health professionals provided a wealth of qualitative data. The resultant findings provide valuable insight into the potential benefits of touchless interaction in clinical settings, and explore the challenges of reliably using touchless input in this context. The study analyses the user experience of different clutching techniques, and how well they integrate with PACS workflows. Though active zone [133] was determined to be the most promising in the context of this work, being the least demanding and working well with existing

PACS usage practices, it should be noted that each of the touchless clutching techniques displayed both strengths and weaknesses (see Table 5.4, p.145). When considering the design of a TMIS, it is important to be aware that optimal clutch technique may depend on the operational context.

The benefits to touchless PACS interaction are further supported by the work in this thesis; HPs further express their desire for touchless input capabilities for hospital systems beyond PACS, from imaging scanners to regular desktop computers. Whilst interesting challenges for future work have emerged during this analysis, it also provides a framework for understanding the design process for the development of touchless PACS technology for deployment in clinical settings. Touchless interaction has the potential to improve clinician workflow, productivity, and possibly even patient outcomes. Clinicians will be able to perform their jobs more effectively, treating patients while improving sterility and reducing the spread of pathogens from shared surfaces and input devices.

7.2 Future Work

This work has raised several interesting questions for research in future work.

The qualitative investigation presented in Chapter 3 (p.45) revealed the impact of operational context on HP use of PACS. The prototype systems presented in Chapter 4 (p.75) implemented levels of multi-modal control, however, the impact of multi-modal input on the user experience was not the focus of the investigation presented in Chapter 5 (p.115). Additional investigation of these topics would bring greater insight to the design of TMIS for clinical contexts. A particular issue of interest is the mapping between appropriate interaction techniques and clutching mechanisms, and specific clinical contexts. HPs expressed a desire for PACS access from mobile devices, such as smartphones and tablets; this could provide a powerful new tool for the clinician, which would be of particular interest in the context of non-radiologist users, whose needs may focus more on accessing radiologist reports with associated images to provide additional context.

Data governance remains a concern among some HPs when using novel technologies, such as TMIS. The use of touchless authentication, e.g., biometric facial recognition, as well as appropriate clutching mechanics may provide tools to enhance the protection of patient data in clinical contexts. Investigation into the role of authentication and clutching in this context would be a valuable step in bringing touchless systems into real-world hospitals.

To enable robust future studies, there is a need for the development of standardised task sets and associated touchless lexicons to be used for experimental benchmarks. To date, most experimental research in this area has been performed with limited cohorts; there is a need for studies employing larger cohorts to ensure the ecological validity of findings. Furthermore, there is a need for contamination studies, in order to better understand the role of touchless technologies in reducing cross-contamination via contact with surfaces.

There is a clear need for improvements to PACS training for clinicians, and more convenient user support could also be helpful. Future work investigating learning delivery methods would be of benefit. Research should focus on determining which learning tools are most effective and best suit the clinician's workflow. In order to best target training, there would be a benefit to research into which PACS tools are most used by various user groups. This would allow training to be targeted and less overwhelming, enabling the clinician to take greater advantage of appropriate subsets of PACS features.

Novel hardware configurations provide a compelling research topic. One such configuration that could bring significant value in the OR context is an array of Azure Kinect DK sensors positioned to enable room-scale tracking. Such a setup could allow tracking of a primary user from authentication in a non-sterile area to engaging in surgery in a sterile area. Robust, persistent user tracking would allow improvements to both security and TMIS performance, both of which are current concerns for touchless systems.

Finally, future work validating the design decisions presented in this work would be of significant value.

Bibliography

[1] André Mewes, Bennet Hensen, Frank Wacker, and Christian Hansen. Touchless interaction with software in interventional radiology and surgery: a systematic literature review. *International Journal of Computer Assisted Radiology and Surgery*, 12, 2 2017.

[2] Wang Shi Hui, Weihong Huang, Jianzhong Hu, Kun Tao, and Yonghong Peng. A new precise contactless medical image multimodal interaction system for surgical practice. *IEEE Access*, 8, 2020.

[3] B Allegranzi. Report on the burden of endemic healthcare-associated infections worldwide, 2011.

[4] Carl Suetens, Katrien Latour, Tommi Kärki, Enrico Ricchizzi, Pete Kinross, Maria Luisa Moro, Béatrice Jans, Susan Hopkins, Sonja Hansen, Outi Lyytikäinen, Jacqui Reilly, Aleksander Deptula, Walter Zingg, Diamantis Plachouras, and Dominique L Monnet. Prevalence of healthcare-associated infections, estimated incidence and composite antimicrobial resistance index in acute care hospitals and long-term care facilities: results from two european point prevalence surveys, 2016 to 2017. *Eurosurveillance*, 23, 11 2018.

[5] Alessandro Cassini, Diamantis Plachouras, Tim Eckmanns, Muna Abu Sin, Hans-Peter Blank, Tanja Ducomble, Sebastian Haller, Thomas Harder, Anja Klingeberg, Madlen Sixtensson, Edward Velasco, Bettina Weiß, Piotr Kramarz, Dominique L. Monnet, Mirjam E. Kretzschmar, and Carl Suetens. Burden of six healthcare-associated infections on european population health: Estimating incidence-based

disability-adjusted life years through a population prevalence-based modelling study. *PLOS Medicine*, 13, 10 2016.

[6] ECDC. Annual epidemiological report on communicable diseases in europe 2008, 2008.

[7] Maureen Schultz, Janet Gill, Sabiha Zubairi, Ruth Huber, and Fred Gordin. Bacterial contamination of computer keyboards in a teaching hospital. *Infection Control & Hospital Epidemiology*, 24, 4 2003.

[8] Vincenzo Russotto, Andrea Cortegiani, Santi Maurizio Raineri, and Antonino Giarratano. Bacterial contamination of inanimate surfaces and equipment in the intensive care unit. *Journal of Intensive Care*, 3, 12 2015.

[9] B.S. Cooper, G.F. Medley, and G.M. Scott. Preliminary analysis of the transmission dynamics of nosocomial infections: stochastic and management effects. *Journal of Hospital Infection*, 43, 10 1999.

[10] Google LLC. Mediapipe. `https://google.github.io/mediapipe/`.

[11] Manomotion. Manomotion. `https://www.manomotion.com/`.

[12] NVIDIA. Deepstream SDK. `https://developer.nvidia.com/deepstream-sdk`.

[13] Gerald Morrison. Touchless interaction in a post-covid-19 world. `https://www.forbes.com/sites/forbestechcouncil/2020/08/26/touchless-interaction-in-a-post-covid-19-world/?sh=69762a842891`, 8 2020.

[14] Seán Cronin and Gavin Doherty. Touchless computer interfaces in hospitals: A review. *Health Informatics Journal*, 25, 12 2019.

[15] Séan Cronin, Bridget Kane, and Gavin Doherty. A qualitative analysis of the needs and experiences of hospital-based clinicians when accessing medical imaging. *Journal of Digital Imaging*, 4 2021.

[16] Rose Johnson, Kenton O'Hara, Abigail Sellen, Claire Cousins, and Antonio Criminisi. Exploring the potential for touchless interaction in image-guided interventional radiology. ACM Press, 2011.

[17] Helena M. Mentis, Kenton O'Hara, Abigail Sellen, and Rikin Trivedi. Interaction proxemics and image use in neurosurgery. ACM Press, 2012.

[18] WHO. Hcai effect on europe, 2014.

[19] Justin H. Tan, Cherng Chao, Mazen Zawaideh, Anne C. Roberts, and Thomas B. Kinney. Informatics in radiology: Developing a touchless user interface for intraoperative image control during interventional radiology procedures. *RadioGraphics*, 33, 3 2013.

[20] J. P. Wachs, H. I. Stern, Y. Edan, M. Gillam, J. Handler, C. Feied, and M. Smith. A gesture-based tool for sterile browsing of radiology images. *Journal of the American Medical Informatics Association*, 15, 2 2008.

[21] Craig Feied, Michael Gillam, Juan Wachs, Jonathan Handler, Helman Stern, and Mark Smith. A real-time gesture interface for hands-free control of electronic medical records. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2006:920, 2006.

[22] Mithun George Jacob, Juan Pablo Wachs, and Rebecca A Packer. Hand-gesture-based sterile interface for the operating room using contextual cues for the navigation of radiological images. *Journal of the American Medical Informatics Association*, 20, 6 2013.

[23] Matt Strickland, Jamie Tremaine, Greg Brigley, and Calvin Law. Using a depth-sensing infrared camera system to access and manipulate medical imaging from within the sterile operating field. *Canadian Journal of Surgery*, 56, 6 2013.

[24] Jonathan E. dela Cruz, John C. Shabosky, Matthew Albrecht, Ted R. Clark, Joseph C. Milbrandt, Steven J. Markwell, and Jason A. Kegg. Typed versus

voice recognition for data entry in an electronic health record: Emergency department physician time utilization and interruptions. *Western Journal of Emergency Medicine*, 15, 7 2014.

[25] Luigi Gallo. A study on the degrees of freedom in touchless interaction. ACM Press, 2013.

[26] Kenton O'Hara, Gerardo Gonzalez, Abigail Sellen, Graeme Penney, Andreas Varnavas, Helena Mentis, Antonio Criminisi, Robert Corish, Mark Rouncefield, Neville Dastur, and Tom Carrell. Touchless interaction in surgery. *Communications of the ACM*, 57, 1 2014.

[27] Kenton O'Hara, Gerardo Gonzalez, Graeme Penney, Abigail Sellen, Robert Corish, Helena Mentis, Andreas Varnavas, Antonio Criminisi, Mark Rouncefield, Neville Dastur, and Tom Carrell. Interactional order and constructed ways of seeing with touchless imaging systems in surgery. *Computer Supported Cooperative Work (CSCW)*, 23, 6 2014.

[28] Yonjae Kim, Simon Leonard, Azad Shademan, Axel Krieger, and Peter C. W. Kim. Kinect technology for hand tracking control of surgical robots: technical and surgical skill comparison to current robotic masters. *Surgical Endoscopy*, 28, 6 2014.

[29] Cherng Chao, Justin Tan, Edward M. Castillo, Mazen Zawaideh, Anne C. Roberts, and Thomas B. Kinney. Comparative efficacy of new interfaces for intra-procedural imaging review: the microsoft kinect, hillcrest labs loop pointer, and the apple ipad. *Journal of Digital Imaging*, 27, 8 2014.

[30] Luigi Gallo, Alessio Pierluigi Placitelli, and Mario Ciampi. Controller-free exploration of medical image data: Experiencing the kinect. IEEE, 6 2011.

[31] Guilherme Cesar Soares Ruppert, Leonardo Oliveira Reis, Paulo Henrique Junqueira Amorim, Thiago Franco de Moraes, and Jorge Vicente Lopes da Silva. Touchless gesture user interface for interactive image visualization in urological surgery. *World Journal of Urology*, 30, 10 2012.

[32] Lars C. Ebert, Gary Hatch, Garyfalia Ampanozi, Michael J. Thali, and Steffen Ross. You can't touch this. *Surgical Innovation*, 19, 9 2012.

[33] Ali Bigdelou, Loren Schwarz, and Nassir Navab. An adaptive solution for intra-operative gesture-based human-machine interaction. ACM Press, 2012.

[34] Shahram Jalaliniya, Jeremiah Smith, Miguel Sousa, Lars Büthe, and Thomas Pederson. Touch-less interaction with medical images using hand & foot gestures. ACM, 9 2013.

[35] David Liu, Mark Zucherman, and William B. Tulloss. Six characteristics of effective structured reporting and the inevitable integration with speech recognition. *Journal of Digital Imaging*, 19, 1 2006.

[36] Guillermo M. Rosa and MarÍa L. Elizondo. Use of a gesture user interface as a touchless image navigation system in dental surgery: Case series report. *Imaging Science in Dentistry*, 44, 2014.

[37] Stanislas Mauser and Oliver Burgert. Touch-free, gesture-based control of medical devices and software based on the leap motion controller. *Studies in health technology and informatics*, 196:265–270, 2014.

[38] Nathaniel Rossol, Irene Cheng, Rui Shen, and Anup Basu. Touchfree medical interfaces. IEEE, 8 2014.

[39] Stephan Schroder, Nina Loftfield, Benjamin Langmann, Klaus Frank, and Eduard Reithmeier. Contactless operating table control based on 3d image processing. IEEE, 8 2014.

[40] Julian Arcos Argoty and Pablo Figueroa. Design and development of a prototype of an interactive hospital room with kinect. ACM Press, 2014.

[41] Ana Luisa Trejos, Karen Siroen, Christopher D. W. Ward, Shahan Hossain, Michael D. Naish, Rajni V. Patel, and Christopher M. Schlachta. Randomized control trial for evaluation of a hands-free pointer for surgical instruction during laparoscopic cholecystectomy. *Surgical Endoscopy*, 29, 12 2015.

[42] Terutaka Marukami, Shoko Tani, Atsuko Matsuda, Keiko Takemoto, Akiko Shindo, and Hiroshi Inada. A basic study on application of voice recognition input to an electronic nursing record system -evaluation of the function as an input interface-. *Journal of Medical Systems*, 36, 6 2012.

[43] Hyunseok Peter Kang, S. Joseph Sirintrapun, Rick J. Nestler, and Anil V. Parwani. Experience with voice recognition in surgical pathology at a large academic multi-institutional center. *American Journal of Clinical Pathology*, 133, 1 2010.

[44] John A. Pezzullo, Glenn A. Tung, Jeffrey M. Rogg, Lawrence M. Davis, Jeffrey M. Brody, and William W. Mayo-Smith. Voice recognition dictation: Radiologist as transcriptionist. *Journal of Digital Imaging*, 21, 12 2008.

[45] Rodney H Strahan and Michal E Schneider-Kolsky. Voice recognition versus transcriptionist: Error rates and productivity in mri reporting. *Journal of Medical Imaging and Radiation Oncology*, 54, 10 2010.

[46] Alexandre Alapetite. Impact of noise and other factors on speech recognition in anaesthesia. *International Journal of Medical Informatics*, 77, 1 2008.

[47] Alexandre Alapetite. Speech recognition for the anaesthesia record during crisis scenarios. *International Journal of Medical Informatics*, 77, 7 2008.

[48] Juan Pablo Wachs, Mathias Kölsch, Helman Stern, and Yael Edan. Vision-based hand-gesture applications. *Communications of the ACM*, 54, 2 2011.

[49] A. Faro, D. Giordano, C. Spampinato, D. De Tommaso, and S. Ullo. An interactive interface for remote administration of clinical tests based on eye tracking. ACM Press, 2010.

[50] Aristotelis Perrakis, Werner Hohenberger, and Thomas Horbach. Integrated operation systems and voice recognition in minimally invasive surgery: comparison of two systems. *Surgical Endoscopy*, 27, 2 2013.

[51] Stefan Soutschek, Jochen Penne, Joachim Hornegger, and Johannes Kornhuber. 3-d gesture-based scene navigation in medical imaging applications using time-of-flight cameras. IEEE, 6 2008.

[52] Miroslav Nagy, Petr Hanzlicek, Jana Zvarova, Tatjana Dostalova, Michaela Seydlova, Radim Hippman, Lubos Smidl, Jan Trmal, and Josef Psutka. Voice-controlled data entry in dental electronic health record. *Studies in health technology and informatics*, 136:529–534, 2008.

[53] Robert Hoyt and Ann Yoshihashi. Lessons learned from implementation of voice recognition for documentation in the military electronic health record system. *Perspectives in health information management*, 7:1e–1e, 1 2010.

[54] Marianne T. Luetmer, Christopher H. Hunt, Robert J. McDonald, Brian J. Bartholmai, and David F. Kallmes. Laterality errors in radiology reports generated with and without voice recognition software: Frequency and clinical significance. *Journal of the American College of Radiology*, 10, 7 2013.

[55] Wan Nurazreena Wan Hassan, Noor Lide Abu Kassim, Abhishek Jhawar, Norsyafiqah Mohd Shurkri, Nur Azreen Kamarul Baharin, and Chee Seng Chan. User acceptance of a touchless sterile system to control virtual orthodontic study models. *American Journal of Orthodontics and Dentofacial Orthopedics*, 149, 4 2016.

[56] Ambreen Zaman, Lars Reisig, Anke Verena Reinschluessel, Huseyin Bektas, Dirk Weyhe, Marc Herrlich, Tanja Döring, and Rainer Malaka. An interactive shoe for surgeons: Hand-free interaction with medical 2d data. ACM, 4 2018.

[57] Julian Hettig, Patrick Saalfeld, Maria Luz, Mathias Becker, Martin Skalej, and Christian Hansen. Comparison of gesture and conventional interaction techniques for interventional neuroradiology. *International Journal of Computer Assisted Radiology and Surgery*, 12, 9 2017.

[58] Behrang Homayoon, John Chung, Ripal T. Gandhi, and David M. Liu. Early clinical experience with a touchless image navigation interface for the endovascular suite. *Minimally Invasive Therapy & Allied Technologies*, 29, 6 2020.

[59] Adam A. Dmytriw, Kevin Phan, Jamie Tremaine, Matt Strickland, Ivan Radovanovic, Timo Krings, and Vitor M. Pereira. Evaluation of a touchless angiography suite computer controller for neuroendovascular procedures. *Clinical Neuroradiology*, 31, 3 2021.

[60] Benjamin Hatscher, Maria Luz, Lennart E. Nacke, Norbert Elkmann, Veit Müller, and Christian Hansen. Gazetap: towards hands-free interaction in the operating room. ACM, 11 2017.

[61] Ben Joonyeon Park, Taekjin Jang, Jong Woo Choi, and Namkug Kim. Gesture-controlled interface for contactless control of various computer programs with a hooking-based keyboard and mouse-mapping technique in the operating room. *Computational and Mathematical Methods in Medicine*, 2016, 2016.

[62] Yuanyuan Feng, Uchenna A. Uchidiuno, Hamid R. Zahiri, Ivan George, Adrian E. Park, and Helena Mentis. Comparison of kinect and leap motion for intraoperative image interaction. *Surgical Innovation*, 28, 2 2021.

[63] Andrés Felipe Botero-Ospina, Sara Isabel Duque-Vallejo, John Fredy Ochoa-Gómez, and Alher Mauricio Hernández-Valdivieso. Touchless control module for diagnostic images at the surgery room using the leap motion system and 3d slicer software. *Revista Facultad de Ingeniería Universidad de Antioquia*, 3 2017.

[64] Vinícius Paes de Camargo, Renato Balancieri, Heloise Manica Paris Teixeira, and Guilherme Corredato Guerino. Touchless modalities of human-computer interaction in hospitals. ACM, 10 2021.

[65] Markus Bockhacker, Hannah Syrek, Max Elstermann von Elster, Sebastian Schmitt, and Henning Roehl. Evaluating usability of a touchless image viewer in the operating room. *Applied Clinical Informatics*, 11, 1 2020.

[66] Yukio Oshiro, Kenoki Ohuchida, Toshiyuki Okada, Makoto Hashizume, and Nobuhiro Ohkohchi. Novel imaging using a touchless display for computer-assisted hepato-biliary surgery. *Surgery Today*, 47, 12 2017.

[67] Glebys Gonzalez, Naveen Madapana, Rahul Taneja, Lingsong Zhang, Richard Rodgers, and Juan P. Wachs. Looking beyond the gesture: Vocabulary acceptability criteria for gesture elicitation studies. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62, 9 2018.

[68] Masahiro Nishihori, Takashi Izumi, Yoshitaka Nagano, Masaki Sato, Tetsuya Tsukada, Asuka Elisabeth Kropp, and Toshihiko Wakabayashi. Development and clinical evaluation of a contactless operating interface for three-dimensional image-guided navigation for endovascular neurosurgery. *International Journal of Computer Assisted Radiology and Surgery*, 16, 4 2021.

[69] John Chung and David M. Liu. Experimental assessment of a novel touchless interface for intraprocedural imaging review. *CardioVascular and Interventional Radiology*, 42, 8 2019.

[70] Cherie-Ann Nathan, Vinaya Chakradeo, Kavita Malhotra, Horacio D'Agostino, and Ravish Patwardhan. The voice-controlled robotic assist scope holder aesop for the endoscopic approach to the sella. *Skull Base*, 16, 5 2006.

[71] Andreas M. Hötker, Michael B. Pitton, Peter Mildenberger, and Christoph Düber. Speech and motion control for interventional radiology: requirements and feasibility. *International Journal of Computer Assisted Radiology and Surgery*, 8, 11 2013.

[72] Pierre Chatelain, Harshita Sharma, Lior Drukker, Aris T. Papageorghiou, and J. Alison Noble. Evaluation of gaze tracking calibration for longitudinal biomedical imaging studies. *IEEE Transactions on Cybernetics*, 50, 1 2020.

[73] Vittoria Frau, Carlo Cuccu, and Lucio Davide Spano. MR^2 : a mixed reality interface for navigating medical records. ACM, 6 2019.

[74] Pei-Ying Chiang, Chun-Chi Chen, and Chih-Hsien Hsia. A touchless interaction interface for observing medical imaging. *Journal of Visual Communication and Image Representation*, 58, 1 2019.

[75] Benjamin Hatscher and Christian Hansen. Hand, foot or voice: Alternative input modalities for touchless interaction in the medical domain. ACM, 10 2018.

[76] Matthew B. Morgan, Barton F. Branstetter, Jonathan Mates, and Paul J. Chang. Flying blind: Using a digital dashboard to navigate a complex pacs environment. *Journal of Digital Imaging*, 19, 1 2006.

[77] Reeva Lederman, Rogier van de Wetering, and Lucy Firth. Pacs contribution to hospital strategy via improved workflow, 2008.

[78] Stirling Bryan, Gwyn Weatherburn, Martin Buxton, Jess Watkins, Justin Keen, and Nicole Muris. Evaluation of a hospital picture archiving and communication system. *Journal of Health Services Research & Policy*, 4, 10 1999.

[79] John Mongan and David Avrin. Impact of PACS-EMR integration on radiologist usage of the EMR. *Journal of Digital Imaging*, 31, 10 2018.

[80] Brucel I. Reiner, Eliot L. Siegel, Frank Hooper, Stephen M. Pomerantz, Zenon Protopapas, Elliott Pickar, and Lois Killewich. Picture archiving and communication systems and vascular surgery: Clinical impressions and suggestions for improvement. *Journal of Digital Imaging*, 9, 11 1996.

[81] David B. Hayt, Steven Alexander, James Drakakis, and Nia Berdebes. Filmless in 60 days: The impact of picture archiving and communications systems within a large urban hospital. *Journal of Digital Imaging*, 14, 6 2001.

[82] Eliot L. Siegel and Bruce Reiner. Work flow redesign: The key to success when using pacs. *Journal of Digital Imaging*, 16, 3 2003.

[83] Malathi Srinivasan, Eric Liederman, Noralyn Baluyot, and Robert Jacoby. Saving time, improving satisfaction: The impact of a digital radiology system on physician

workflow and system efficiency. *Journal of healthcare information management : JHIM*, 20:123–131, 2006.

[84] C Trumm, D Morhard, B Ertl-Wagner, C Glaser, and M Reiser. Impact of ris/pacs integrated speech recognition on report availability. *Radiology management*, 30 6:16–23; quiz 24–6, 2008.

[85] David B. Hayt and Steven Alexander. The pros and cons of implementing pacs and speech recognition systems. *Journal of Digital Imaging*, 14, 9 2001.

[86] James Chen, John Bradshaw, and Paul Nagy. Has the picture archiving and communication system (pacs) become a commodity? *Journal of Digital Imaging*, 24, 2 2011.

[87] Rogier van de Wetering and Ronald Batenburg. Towards a theory of pacs deployment: An integrative pacs maturity framework. *Journal of Digital Imaging*, 1 2014.

[88] Isla M Hains, Andrew Georgiou, and Johanna I Westbrook. The impact of pacs on clinician work practices in the intensive care unit: a systematic review of the literature. *Journal of the American Medical Informatics Association*, 19, 7 2012.

[89] Nicola K Gale, Gemma Heath, Elaine Cameron, Sabina Rashid, and Sabi Redwood. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology*, 13, 12 2013.

[90] Virginia Braun and Victoria Clarke. Thematic analysis. 2012.

[91] Virginia Braun and Victoria Clarke. *Successful qualitative research: A practical guide for beginners.* sage, 2013.

[92] Victoria Clarke and Virginia Braun. Thematic analysis. In *Encyclopedia of critical psychology*, pages 1947–1952. Springer, 2014.

[93] Virginia Braun and Victoria Clarke. One size fits all? what counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18:328–352, 7 2021.

[94] Daniel R. Gale, M. Elon Gale, Rebecca K. Schwartz, Victorine V. Muse, and Richard E. Walker. An automated pacs workstation interface. *American Journal of Roentgenology*, 174, 1 2000.

[95] Ahmed Hosny, Chintan Parmar, John Quackenbush, Lawrence H. Schwartz, and Hugo J. W. L. Aerts. Artificial intelligence in radiology. *Nature Reviews Cancer*, 18, 8 2018.

[96] Bridget Kane and Saturnino Luz. Achieving diagnosis by consensus. *Computer Supported Cooperative Work (CSCW)*, 18, 8 2009.

[97] Who surgical safety checklist. `https://www.who.int/patientsafety/safesurgery/checklist`.

[98] Rogier van de Wetering, Ronald Batenburg, Johan Versendaal, Reeva Lederman, and Lucy Firth. A balanced evaluation perspective: Picture archiving and communication system impacts on hospital workflow. *Journal of Digital Imaging*, 19, 12 2006.

[99] Kent Fridell, Lars Edgren, Lars Lindsköld, Peter Aspelin, and Nina Lundberg. The impact of pacs on radiologists' work practice. *Journal of Digital Imaging*, 20, 12 2007.

[100] Mehmet Top. Physicians' views and assessments on picture archiving and communication systems (pacs) in two turkish public hospitals. *Journal of Medical Systems*, 36, 12 2012.

[101] Molly E. Roseland, Ella A. Kazerooni, Janet E. Bailey, Gary D. Luker, Richard H. Cohan, and Matthew S. Davenport. Needs assessment using a structured prioritization schema: An open letter to pacs vendors. *Journal of the American College of Radiology*, 16, 2 2019.

[102] HSE. The hse and ehealth ireland welcome the commencement order for the individual health identifier. `https://www.hse.ie/eng/services/news/media/pressrel/`

`the-hse-and-ehealth-ireland-welcome-the-commencement-order-for-the-individual-health-`
`html`, 2017.

[103] Jonas F. Ludvigsson, Petra Otterblad-Olausson, Birgitta U. Pettersson, and Anders Ekbom. The swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *European Journal of Epidemiology*, 24, 11 2009.

[104] Simona Cohen, Flora Gilboa, and Uri Shani. Pacs and electronic health records. 5 2002.

[105] Kamran Sartipi, Krupa A Kuriakose, and Weina Ma. An infrastructure for secure sharing of medical images between pacs and ehr systems. pages 245–259. IBM Corp., 2013.

[106] Benita Cox and Nina Dawe. Evaluation of the impact of a pacs system on an intensive care unit. *Journal of Management in Medicine*, 16, 4 2002.

[107] Kent Fridell, Peter Aspelin, Lars Edgren, Lars Lindsköld, and Nina Lundberg. Pacs influence the radiographer's work. *Radiography*, 15, 5 2009.

[108] Steve G. Langer. Impact of speech recognition on radiologist productivity. *Journal of Digital Imaging*, 15, 12 2002.

[109] Luigi Lepanto, Guy Paré, and Alain Gauvin. Impact of PACS deployment strategy on dictation turnaround time of chest radiographs. *Academic Radiology*, 13, 4 2006.

[110] Guy Paré and Marie-Claude Trudel. Knowledge barriers to pacs adoption and implementation in hospitals. *International Journal of Medical Informatics*, 76, 1 2007.

[111] Maria Karam and m.c Schraefel. A taxonomy of gestures in human computer interactions. *Electronics and Computer Science*, 01 2005.

[112] Kenton O'hara, Richard Harper, Helena Mentis, Abigail Sellen, and Alex Taylor. On the naturalness of touchless. *ACM Transactions on Computer-Human Interaction*, 20, 3 2013.

[113] Dan Saffer. *Designing Gestural Interfaces: Touchscreens and Interactive Devices.* O'Reilly Media, Inc., 2008.

[114] Robert J.K. Jacob, Audrey Girouard, Leanne M. Hirshfield, Michael S. Horn, Orit Shaer, Erin Treacy Solovey, and Jamie Zigelbaum. Reality-based interaction: A framework for post-wimp interfaces. ACM Press, 2008.

[115] Donald A. Norman. Natural user interfaces are not natural. *Interactions*, 17(3):6–10, May 2010.

[116] Julie Rico and Stephen Brewster. Usable gestures for mobile interfaces. ACM Press, 2010.

[117] Erving Goffman. *The Presentation of Self in Everyday Life.* Penguin Books, London, 1990.

[118] Ludwig Wittgenstein. *The Visible and the Invisible: The Intertwining—The Chiasm.* Oxford, 1969.

[119] Maurice Merleau-Ponty. *Phenomenology of Perception.* Routledge, 1962.

[120] Maurice Merleau-Ponty. *The Visible and the Invisible: The Intertwining—The Chiasm.* Northwestern University Press, 1968.

[121] Astrid Twenebowa Larssen, Toni Robertson, and Jenny Edwards. The feel dimension of technology interaction. ACM Press, 2007.

[122] Muhammad Zahid Iqbal and Abraham Campbell. The emerging need for touchless interaction technologies. *Interactions*, 27, 7 2020.

[123] Sarah Perez. Amazon's alexa app hits the top of the app store for the first time. `https://techcrunch.com/2017/12/27/amazons-alexa-app-hits-the-top-of-the-app-store-for-the-first-time`, 2017.

[124] Ben Rubin. Amazon sees alexa devices more than double in just one year. `https://www.cnet.com/home/smart-home/amazon-sees-alexa-devices-more-than-double-in-just-one-year`, 1 2020.

[125] Brendan O'Flynn, Javier Sanchez-Torres, Salvatore Tedesco, and Michael Walsh. Challenges in the development of wearable human machine interface systems. pages 10.4.1–10.4.4. IEEE, 12 2019.

[126] Don Norman. *The Design of Everyday Things*. Revised edition, 2013.

[127] Jenny Preece, Helen Sharp, and Yvonne Rogers. *Interaction Design: Beyond Human-Computer Interaction*. 2015.

[128] Maria Karam. A framework for research and design of gesture-based human computer interactions, 2006.

[129] V. T. Nguyen. Enhancing touchless interaction with the leap motion using a haptic glove. 2014.

[130] Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassell. A Survey of Design Issues in Spatial Input. In *Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology - UIST '94*, pages 213–222. ACM Press, 1994.

[131] Victoria Bellotti, Maribeth Back, W. Keith Edwards, Rebecca E. Grinter, Austin Henderson, and Cristina Lopes. Making Sense of Sensing Systems: Five Questions for Designers and Researchers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '02*, pages 415–422. ACM Press, 2002.

[132] Pawel Woźniak, Morten Fjeld, and Shengdong Zhao. Limiting trial and error. ACM Press, 2014.

[133] Thomas Baudel and Michel Beaudouin-Lafon. Charade: Remote Control of Objects Using Free-Hand Gestures. *Communications of the ACM*, 36(7):28–35, 1993.

[134] Julia Schwarz, Charles Claudius Marais, Tommer Leyvand, Scott E. Hudson, and Jennifer Mankoff. Combining Body Pose, Gaze, and Gesture to Determine Intention

to Interact in Vision-Based Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '14*, pages 3443–3452. ACM Press, 2014.

[135] Mithun Jacob, Christopher Cange, Rebecca Packer, and Juan P Wachs. Intention, context and gesture recognition for sterile mri navigation in the operating room. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications - CIARP '12*, pages 220–227. Springer, 2012.

[136] Kenton O'Hara, Gerardo Gonzalez, Abigail Sellen, Graeme Penney, Andreas Varnavas, Helena Mentis, Antonio Criminisi, Robert Corish, Mark Rouncefield, Neville Dastur, and Tom Carrell. Touchless Interaction in Surgery. *CACM*, 57(1):70–77, 2014.

[137] Ismo Rakkolainen, Antti Sand, and Roope Raisamo. A survey of mid-air ultrasonic tactile feedback. pages 94–944, 12 2019.

[138] Ismo Rakkolainen, Euan Freeman, Antti Sand, Roope Raisamo, and Stephen Brewster. A Survey of Mid-Air Ultrasound Haptics and Its Applications. *IEEE Transactions on Haptics*, 14:2–19, 2020.

[139] Thomas P. Moran and John M. Carroll. *Design rationale : concepts, techniques and use.* Mahwah, N J: Lawrence Erlbaum, 1996.

[140] Vito Gentile, Alessio Malizia, Salvatore Sorce, and Antonio Gentile. Designing touchless gestural interactions for public displays in-the-wild, 2015.

[141] Sukeshini A. Grandhi, Gina Joue, and Irene Mittelberg. Understanding naturalness and intuitiveness in gesture production. ACM, 5 2011.

[142] Rick Kjeldsen and Jacob Hartman. Design issues for vision-based computer interaction systems. In *Proceedings of the 2001 workshop on Perceptive User Interfaces - PUI '01*, pages 1–8. ACM Press, 2001.

[143] Euan Freeman, Stephen Brewster, and Vuokko Lantz. Do That, There: An Interaction Technique for Addressing In-Air Gesture Systems. In *Proceedings of the*

2016 CHI Conference on Human Factors in Computing Systems - CHI '16, pages 2319–2331. ACM Press, 2016.

[144] Euan Freeman, Dong-Bach Vo, and Stephen Brewster. Haptiglow: Helping users position their hands for better mid-air gestures and ultrasound haptic feedback. In *Proceedings of the IEEE World Haptics Conference*, WHC '19, pages 289–294. IEEE, 2019.

[145] Stefanie TL Pöhlmann, Elaine F Harkness, Christopher J Taylor, and Susan M Astley. Evaluation of kinect 3d sensor for healthcare imaging. *Journal of medical and biological engineering*, 36(6):857–870, 2016.

[146] Helena M. Mentis, Kenton O'Hara, Gerardo Gonzalez, Abigail Sellen, Robert Corish, Antonio Criminisi, Rikin Trivedi, and Pierre Theodore. Voice or gesture in the operating room. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, page 773–780, New York, NY, USA, 2015. Association for Computing Machinery.

[147] Richard A. Bolt. "Put-That-There": Voice and Gesture at the Graphics Interface. *ACM SIGGRAPH Computer Graphics*, 14(3):262–270, 1980.

[148] Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassell. Passive real-world interface props for neurosurgical visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '94*, pages 452–458. ACM Press, 1994.

[149] David Fleer and Christian Leichsenring. MISO: A Context-Sensitive Multimodal Interface for Smart Objects Based on Hand Gestures and Finger Snaps. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology - UIST '12*, pages 93–94. ACM Press, 2012.

[150] Robert Walter, Gilles Bailly, and Jörg Müller. StrikeAPose: Revealing Mid-Air Gestures on Public Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, pages 841–850. ACM Press, 2013.

[151] Andrew D. Wilson. Robust computer vision-based detection of pinching for one and two-handed gesture input. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology - UIST '06*, pages 255–258. ACM Press, 2008.

[152] Euan Freeman, Julie Williamson, Sriram Subramanian, and Stephen Brewster. Point-and-Shake: Selecting from Levitating Object Displays. In *Proceedings of the 36th Annual ACM Conference on Human Factors in Computing Systems - CHI '18*, page Paper 18. ACM Press, 2018.

[153] Hannah Limerick. Call to interact: Communicating interactivity and affordances for contactless gesture controlled public displays. In *Proceedings of the 9TH ACM International Symposium on Pervasive Displays*, PerDis '20, page 63–70, New York, NY, USA, 2020. Association for Computing Machinery.

[154] Sandra G. Hart and Lowell E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In Peter A. Hancock and Najmedin Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139–183. North-Holland, 1988.

[155] Naveen Madapana, Glebys Gonzalez, Rahul Taneja, Richard Rodgers, Lingsong Zhang, and Juan Wachs. Preference elicitation: Obtaining gestural guidelines for pacs in neurosurgery. *International Journal of Medical Informatics*, 130, 10 2019.

[156] Dicom library. `https://dicomlibrary.com/`.

[157] Daniel Wigdor and Dennis Wixon. *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1st edition, 2011.

[158] Alessio Pierluigi Placitelli and Luigi Gallo. Toward a framework for rapid prototyping of touchless user interfaces. pages 539–543. IEEE, 7 2012.

[159] Barney Glaser and Anselm Strauss. *Discovery of grounded theory: Strategies for qualitative research.* Aldine, 1967.

[160] Ylona Chun Tie, Melanie Birks, and Karen Francis. Grounded theory research: A design framework for novice researchers. *SAGE Open Medicine*, 7:205031211882292, 1 2019.

[161] Paul R. Barach and Bill Rostenberg. Design of cardiac surgery operating rooms and the impact of the built environment, 2015.

[162] Anand D Padmakumar, Victoria Bhasin, Timothy N Wenham, and Andrew R Bodenham. Evaluation of noise levels in intensive care units in two large teaching hospitals – a prospective observational study. *Journal of the Intensive Care Society*, 14:205–210, 7 2013.

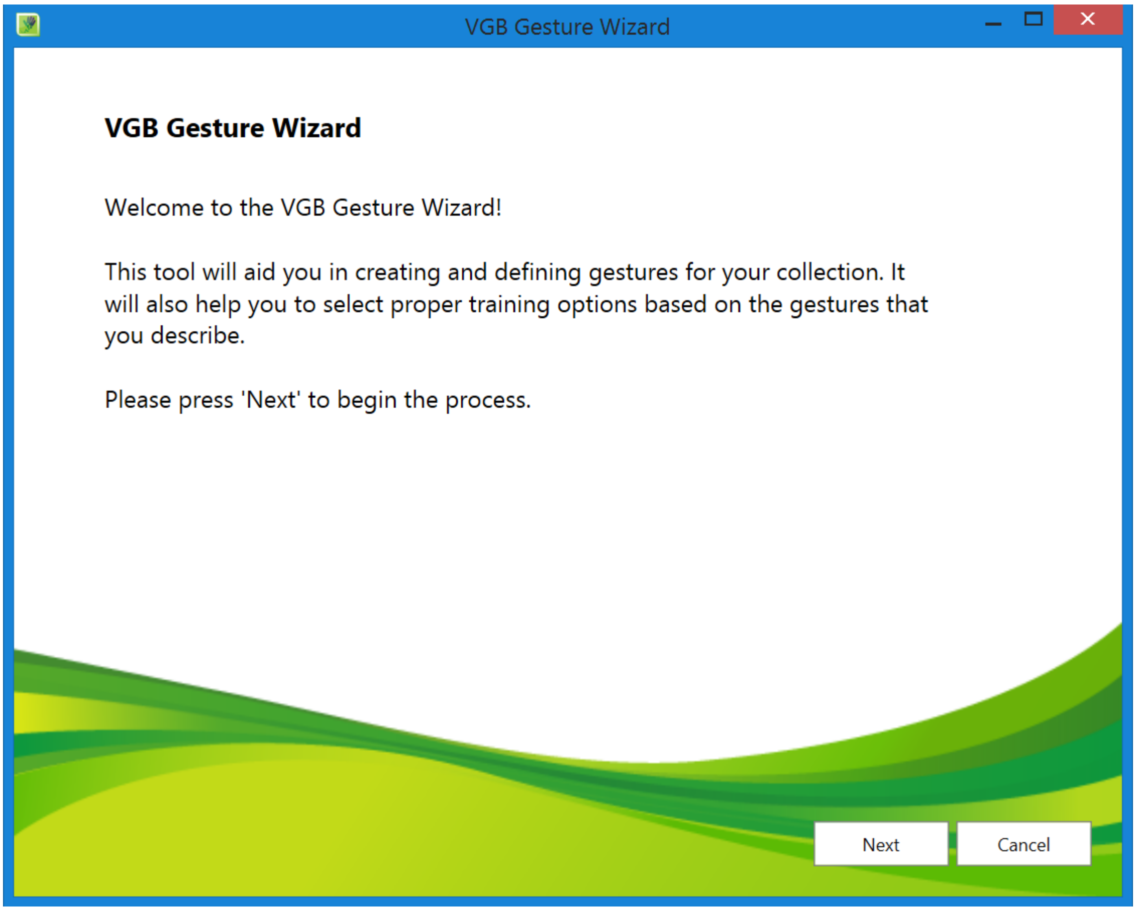
[163] Julie K. Taitsman, Christi Macrina Grimm, and Shantanu Agrawal. Protecting patient privacy and data security. *New England Journal of Medicine*, 368:977–979, 3 2013.

[164] Gerard Lacey, Lucyna Gozdzielewska, Kareena McAloney-Kocaman, Jonathan Ruttle, Sean Cronin, and Lesley Price. Psychomotor learning theory informing the design and evaluation of an interactive augmented reality hand hygiene training app for healthcare workers. *Education and Information Technologies*, 10 2021.

[165] Jiunn-Woei Lian, David C. Yen, and Yen-Ting Wang. An exploratory study to understand the critical factors affecting the decision to adopt cloud computing in taiwan hospital. *International Journal of Information Management*, 34:28–36, 2 2014.

Appendix A

Initial Prototype Software Tools



VGB Gesture Wizard

Welcome to the VGB Gesture Wizard!

This tool will aid you in creating and defining gestures for your collection. It will also help you to select proper training options based on the gestures that you describe.

Please press 'Next' to begin the process.

Figure A.1: When creating a gesture project the VGB Gesture Wizard provided a convenient interface for setting basic parameters such as whether the gesture was one or two-handed, and if the gesture relied on hand states.

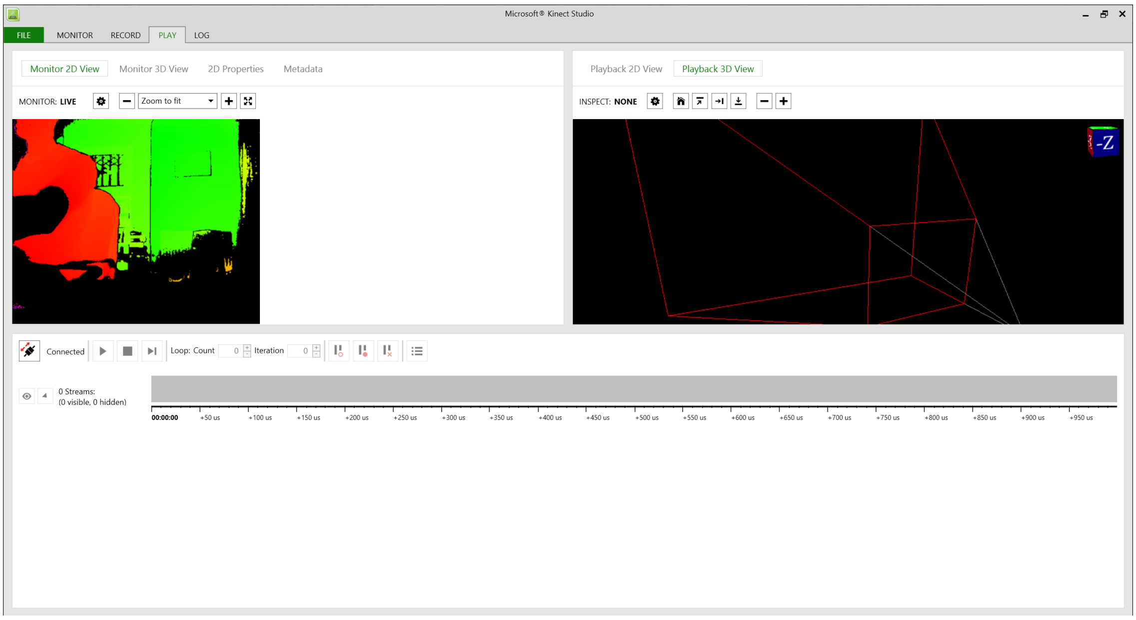


Figure A.2: The playback view of Kinect Studio allowed opening and viewing of recorded eXtended Event Files (.xef) files.

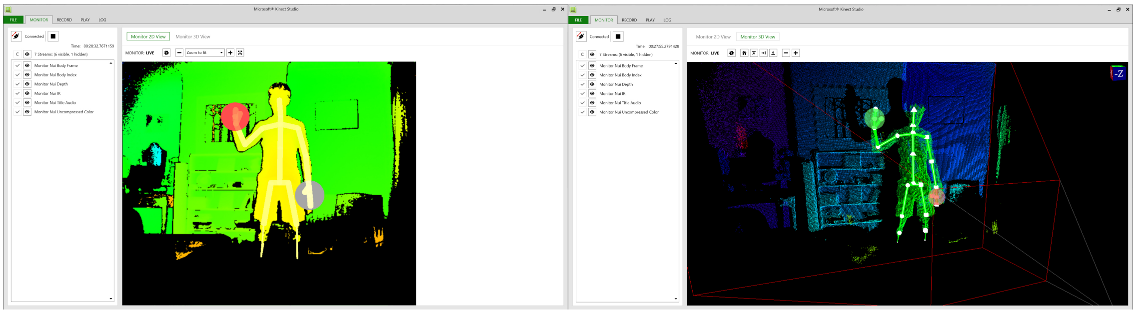
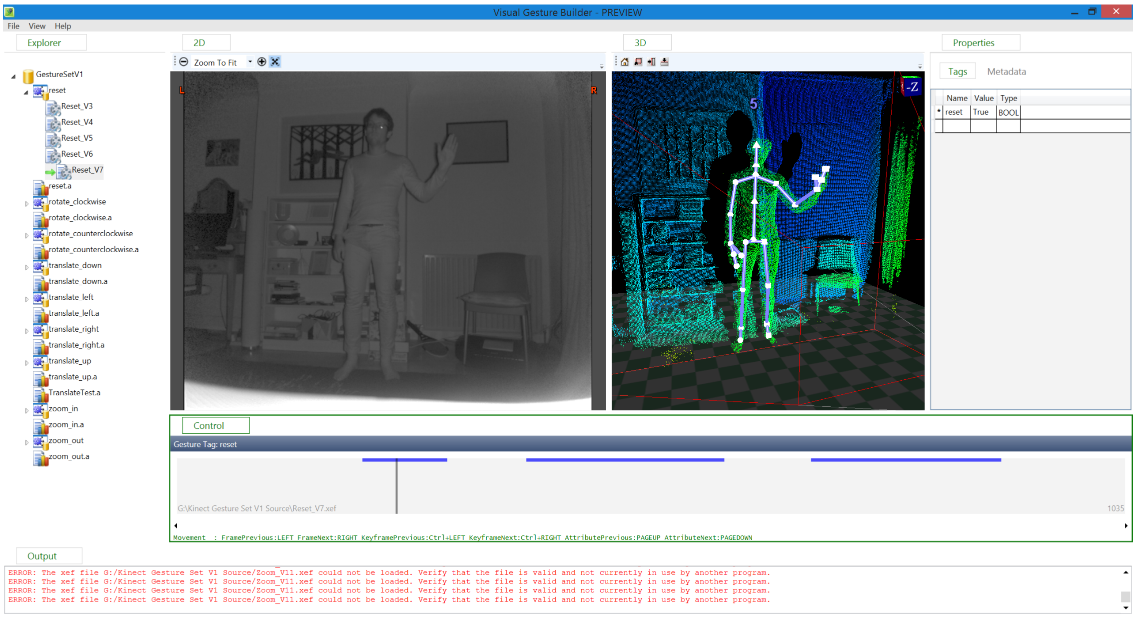


Figure A.3: Left: A live feed from the Kinect sensor is displayed in monitor mode showing a segmented depth image as well as a detected skeleton with a closed fist detected. Right: A live feed from the Kinect sensor showing the 3D point cloud of depth data as well as a detected skeleton with an open palm detected.



Name	Value	Type
reset	True	BOOL

```
Movement  : FramePrevious:LEFT FrameNext:RIGHT KeyframePrevious:Ctrl+LEFT KeyframeNext:Ctrl+RIGHT AttributePrevious:PAGEUP AttributeNext:PAGEDOWN
```

```
ERROR: The xef file G:/Kinect Gesture Set V1 Source/Zoom_V11.xef could not be loaded. Verify that the file is valid and not currently in use by another program.
ERROR: The xef file G:/Kinect Gesture Set V1 Source/Zoom_V11.xef could not be loaded. Verify that the file is valid and not currently in use by another program.
ERROR: The xef file G:/Kinect Gesture Set V1 Source/Zoom_V11.xef could not be loaded. Verify that the file is valid and not currently in use by another program.
ERROR: The xef file G:/Kinect Gesture Set V1 Source/Zoom_V11.xef could not be loaded. Verify that the file is valid and not currently in use by another program.
```

Figure A.4: Visual Gesture Builder with the GestureSetV1 project opened. Each gesture was created using VGB Gesture Wizard. The blue lines in the timeline represent segments of footage that have been tagged as being the trained gesture.

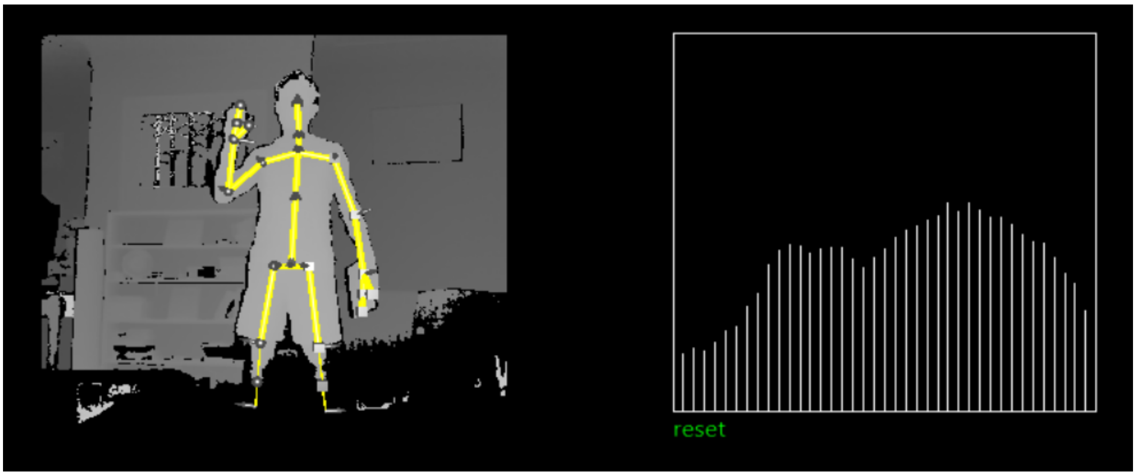
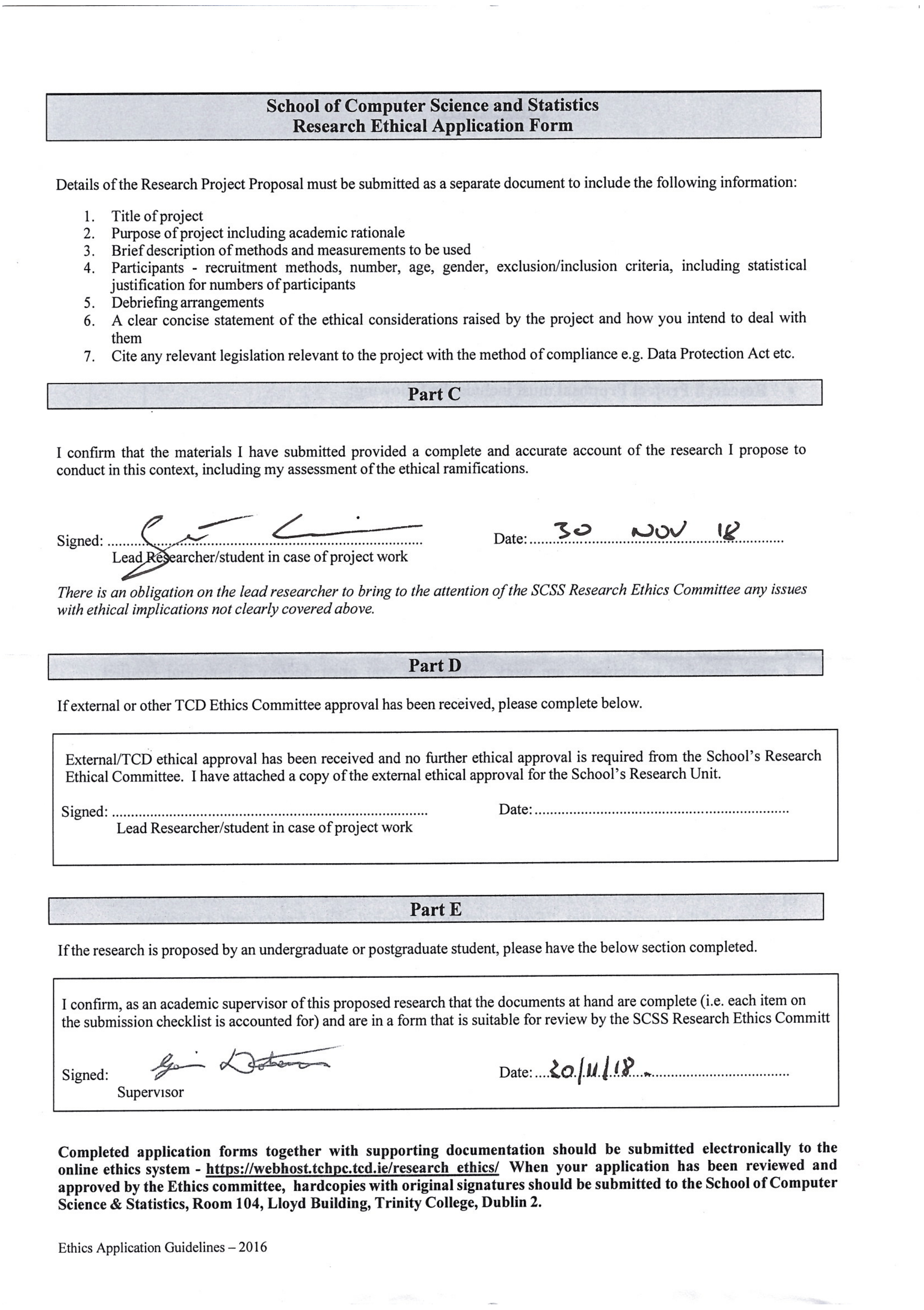


Figure A.5: Close-up view of the Visual Gesture Builder Viewer showing confidence values over time.

Appendix B

Reseach Ethical Application Forms



School of Computer Science and Statistics
Research Ethical Application Form

Details of the Research Project Proposal must be submitted as a separate document to include the following information:

1. Title of project
2. Purpose of project including academic rationale
3. Brief description of methods and measurements to be used
4. Participants - recruitment methods, number, age, gender, exclusion/inclusion criteria, including statistical justification for numbers of participants
5. Debriefing arrangements
6. A clear concise statement of the ethical considerations raised by the project and how you intend to deal with them
7. Cite any relevant legislation relevant to the project with the method of compliance e.g. Data Protection Act etc.

Part C

I confirm that the materials I have submitted provided a complete and accurate account of the research I propose to conduct in this context, including my assessment of the ethical ramifications.

Signed: .... Date: 30 NOV 18
Lead Researcher/student in case of project work

There is an obligation on the lead researcher to bring to the attention of the SCSS Research Ethics Committee any issues with ethical implications not clearly covered above.

Part D

If external or other TCD Ethics Committee approval has been received, please complete below.

External/TCD ethical approval has been received and no further ethical approval is required from the School's Research Ethical Committee. I have attached a copy of the external ethical approval for the School's Research Unit.

Signed: .... Date: ....
Lead Researcher/student in case of project work

Part E

If the research is proposed by an undergraduate or postgraduate student, please have the below section completed.

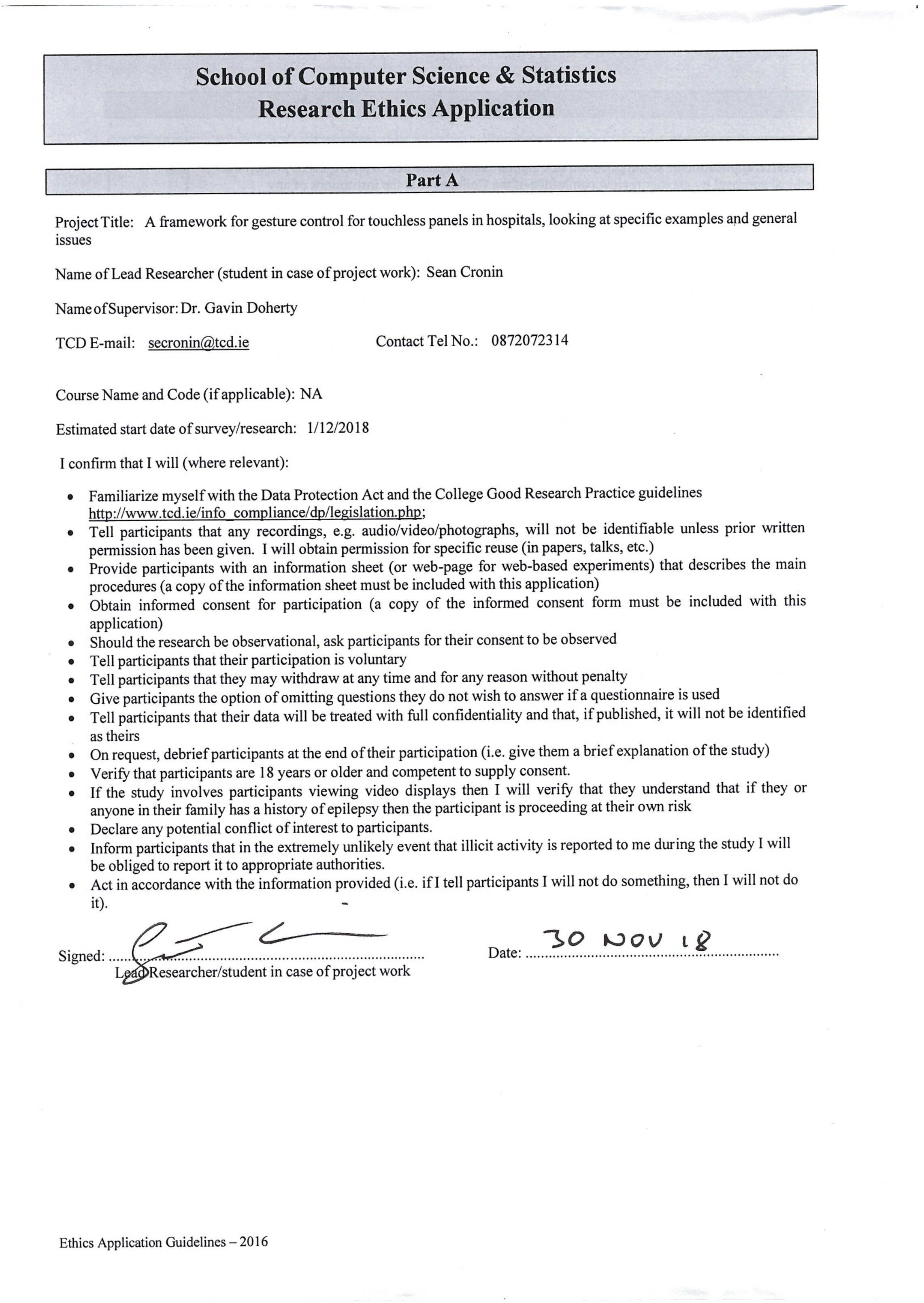
I confirm, as an academic supervisor of this proposed research that the documents at hand are complete (i.e. each item on the submission checklist is accounted for) and are in a form that is suitable for review by the SCSS Research Ethics Committ

Signed: Date: 20/11/18
Supervisor

Completed application forms together with supporting documentation should be submitted electronically to the online ethics system - https://webhost.tchpc.tcd.ie/research_ethics/ When your application has been reviewed and approved by the Ethics committee, hardcopies with original signatures should be submitted to the School of Computer Science & Statistics, Room 104, Lloyd Building, Trinity College, Dublin 2.

Ethics Application Guidelines – 2016

Figure B.1: Research ethical application form (signed)



School of Computer Science & Statistics
Research Ethics Application

Part A

Project Title: A framework for gesture control for touchless panels in hospitals, looking at specific examples and general issues

Name of Lead Researcher (student in case of project work): Sean Cronin

Name of Supervisor: Dr. Gavin Doherty

TCD E-mail: secronin@tcd.ie Contact Tel No.: 0872072314

Course Name and Code (if applicable): NA

Estimated start date of survey/research: 1/12/2018

I confirm that I will (where relevant):

- Familiarize myself with the Data Protection Act and the College Good Research Practice guidelines http://www.tcd.ie/info_compliance/dp/legislation.php;
- Tell participants that any recordings, e.g. audio/video/photographs, will not be identifiable unless prior written permission has been given. I will obtain permission for specific reuse (in papers, talks, etc.)
- Provide participants with an information sheet (or web-page for web-based experiments) that describes the main procedures (a copy of the information sheet must be included with this application)
- Obtain informed consent for participation (a copy of the informed consent form must be included with this application)
- Should the research be observational, ask participants for their consent to be observed
- Tell participants that their participation is voluntary
- Tell participants that they may withdraw at any time and for any reason without penalty
- Give participants the option of omitting questions they do not wish to answer if a questionnaire is used
- Tell participants that their data will be treated with full confidentiality and that, if published, it will not be identified as theirs
- On request, debrief participants at the end of their participation (i.e. give them a brief explanation of the study)
- Verify that participants are 18 years or older and competent to supply consent.
- If the study involves participants viewing video displays then I will verify that they understand that if they or anyone in their family has a history of epilepsy then the participant is proceeding at their own risk
- Declare any potential conflict of interest to participants.
- Inform participants that in the extremely unlikely event that illicit activity is reported to me during the study I will be obliged to report it to appropriate authorities.
- Act in accordance with the information provided (i.e. if I tell participants I will not do something, then I will not do it).

Signed: .... Date: 30 NOV 18

Lead Researcher/student in case of project work

Ethics Application Guidelines – 2016

Figure B.2: Research ethics application

Appendix C

User Study: information sheet for participants

INFORMATION SHEET– Interviews

You are invited to participate in this interview study about usage of Picture Archiving and Communication Systems (PACS). The study aims to support development of new interfaces to Picture Archiving and Communications System (PACS).

This study is being carried out by Sean Cronin in part fulfilment of the requirements of a postgraduate research degree in Computer Science at Trinity College Dublin. We are interested in recruiting healthcare workers for the interviews who are actively involved in using PACs.

We expect the interviews to take 15-20 minutes. We do not anticipate any risks to participants.

The interview will involve discussing work practices with respect to PACS, and we expect that this will in require approximately 15-20 minutes of your time. We would like to audio record the interview (we will transcribe the conversation and delete the audio afterwards), but this is optional. Recordings will be kept on a secure server in Trinity College Dublin until transcription is completed satisfactorily. In keeping with TCD guidelines on good research practice, transcriptions will be retained for 10 years following completion of the study for research validation purposes.

A paper record associating recordings with individual contact details will be kept separately from the recordings so that you will have the option to withdraw should you wish to do so.

Your participation is entirely voluntary, and you can withdraw at any time prior to publication without penalty. You don't need to provide a reason for your decision to do so. In order to withdraw, simply email the researcher giving the email you used to sign up with, indicating that you wish to withdraw. It will not be possible to removed from the study following publication of the results, but recordings can still be deleted.

We plan to publish the results of our research in academic journals and conference proceedings. We will do this in a way which does not identify you, or any other individual participant. No audio or video recordings will be made available to anyone other than the immediate research team at Trinity College Dublin, nor will any such recordings be replayed in any public forum or presentation.

Copies of all publications resulting from the study will be available free of charge on request.

While it is unlikely that illicit activities would be disclosed, if you do so, we would be obliged to report the to the appropriate authorities.

We may use direct quotations from participants. If you wish to review transcriptions for validity then simply let the researcher know you wish to do this.

Conflict of interest:
This is not an independent evaluation. The researchers, who are also the software developers, are conducting this research to explore the usefulness of this form of interaction.

Appendix D

User Study: informed consent

INFORMED CONSENT FORM - Interview

Background of study:
The purpose of the study is to understand Picture Archiving and Communication Systems (PACS) usage in real-world environments. We would like to interview you, as a health professional with experience of using PACS. The interview will involve discussing work practices with respect to PACS, and we expect that this will require approximately 30-40 minutes of your time. We would like to audio record the interview (we will transcribe the conversation and delete the audio afterwards), but this is optional. The study is being carried out in partial fulfillment of the requirements for a postgraduate research degree in Computer Science at Trinity College Dublin.

By signing this form, I am consenting to participate in the interview, and am making the following declarations:

- I am 18 years or older and am competent to provide consent.
- I have read, or had read to me, a document providing information about this research and this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction and understand the description of the research that is being provided to me.
- I am happy that the whole or some parts of the interview might be audio recorded (this statement should be crossed out if not consenting to audio recording)
- I agree that my data is used for scientific purposes and I have no objection that my data is published in scientific publications in a way that does not reveal my identity.
- I understand that if I make illicit activities known, these will be reported to appropriate authorities.
- I understand that I may stop electronic recordings at any time, and that I may at any time, even subsequent to my participation have such recordings destroyed (except in situations such as above).
- I understand that, subject to the constraints above, no recordings will be replayed in any public forum or made available to any audience other than the current researchers/research team.
- I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.
- I understand that I may refuse to answer any question and that I may withdraw at any time without giving a reason.
- I understand that my participation is anonymous and that no personal details about me will be recorded, with the exception of audio recordings (which will potentially be identifiable as your voice) until they are transcribed and the original audio deleted.
- I understand that I can make subsequent contact with the researcher Sean Cronin (via email: secronin@tcd.ie) if I wish to obtain a copy of any papers derived from the research.
- I have received a copy of this agreement.

I, _______________________________, consent to participate in this study conducted by Sean Cronin in the School of Computer Science and Statistics, Trinity College, Dublin under the supervision of Dr. Gavin Doherty and I confirm that I am over 18 years old.

I, _______________________________, consent to being audio recorded as a part of this interview.

Statement of investigator's responsibility: I have explained the nature and purpose of this research study, the procedures to be undertaken and any risks that may be involved. I have offered to answer any questions and fully answered such questions. I believe that the participant understands my explanation and has freely given informed consent.

RESEARCHERS CONTACT DETAILS:
Sean Cronin, secronin@tcd.ie
School of Computer Science & Statistics, Trinity College Dublin

INVESTIGATOR'S SIGNATURE: **Date:**

Appendix E

User Study: interview questions

Interview Guideline

The conducted interview will be a semi-structured one. The participants will be asked general questions and will be encouraged to develop them by giving more details or their personal experiences. Each question is optional. Participants are free to omit a response to any question.

Background information:

- What is/was your position in the hospital?
- How long have you been working with PACS?

PACs:

- How often do you use PACs?
- Who else uses a PACs routinely?
- What tasks are you aiming to achieve when you use a PACs?
- What features of PACs do you use most often?
- Where do you use PACs most often?
- What kind of problems or challenges do you encounter interacting with PACs?
- How important is maintaining sterility to you when using PACs?
- Do you think it would be useful to be able to operate PACS using gestures?
- Can you foresee any difficulties with doing this?

Training (lack of training, what could be better, what would work for your workflow?)
What would a touchless system need to be for you to be happy to use it?
How many types of pacs? Differences?
EHR PACS integration

Appendix F

Prototype Study: experiment protocol

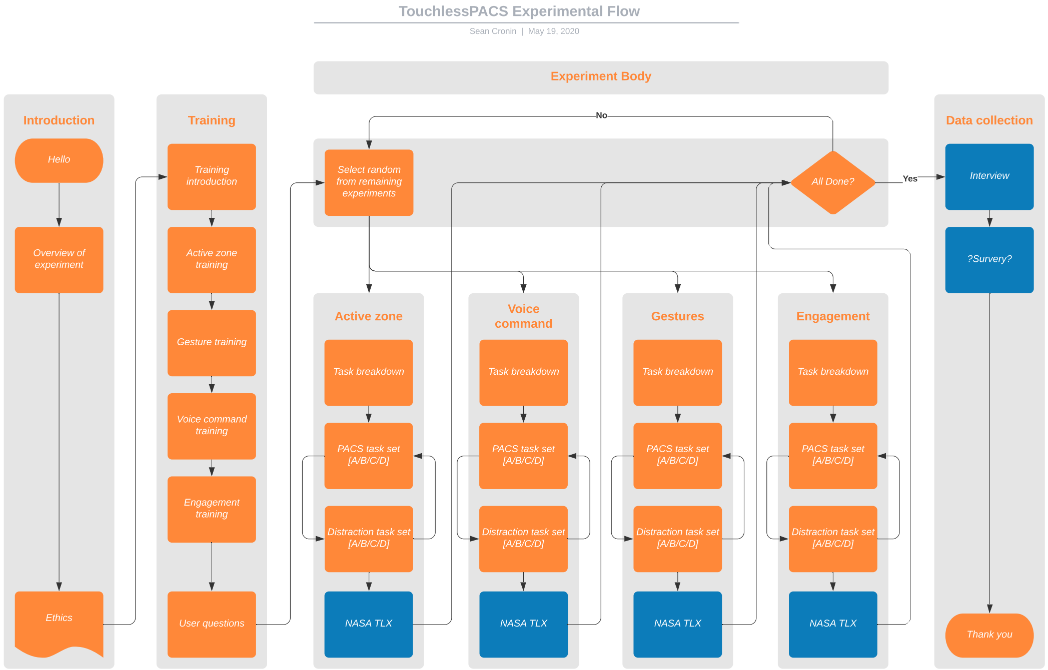
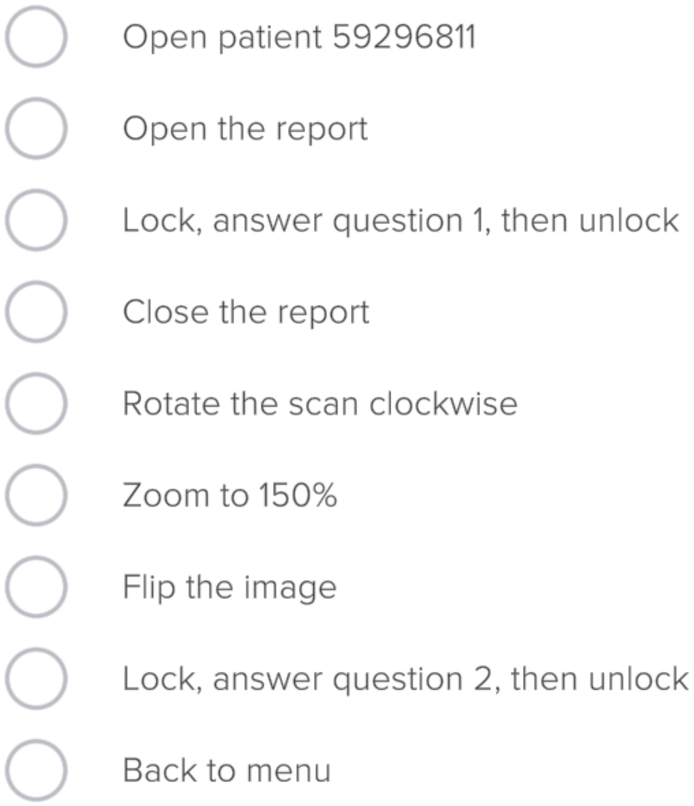


Figure F.1: The experimental session was broken into four primary sections, introduction, training, experiments, and data collection. During the introduction stage, users were introduced to the experiment, its aims, and the experiment protocol. They also completed informed consent forms at this stage. During the training stage, users were able to learn and practice all the interactions they would use with the system, as well as ask any questions the y may have. Within the experiment body, four distinct experiment events occurred (for each of the clutch modalities under investigation). Users were asked to complete a set of tasks (e.g., F.2), both on- and off-screen. At the end of each experiment event, users completed a NASA TLX survey to measure workload. Finally, during the data collection stage, a semi-structured interview was performed to gather qualitative insights into the user experience.



○ Open patient 59296811

○ Open the report

○ Lock, answer question 1, then unlock

○ Close the report

○ Rotate the scan clockwise

○ Zoom to 150%

○ Flip the image

○ Lock, answer question 2, then unlock

○ Back to menu

Figure F.2: An example set of experiment tasks that users were asked to perform.

Appendix G

Prototype Study: information sheet for participants

INFORMATION SHEET– Prototype Evaluation

You are invited to participate in this study on touchless interactions for Picture Archiving and Communication Systems (PACS). The study aims to improve the process of developing such systems. The purpose of the study is to understand Picture Archiving and Communications System (PACS) usage, and operation via touchless interfaces. The study also aims to gather information about touchless interaction methods, focusing on the question of touchless clutching mechanisms.

This study is being carried out by Sean Cronin in part fulfilment of the requirements of a postgraduate research degree in Computer Science at Trinity College Dublin. We will be recruiting both participants with experience of using PACS and without such experience.

First, you will be presented with a regular picture display and navigation system and given an opportunity to familiarize yourself with the system; you may ask any questions at this stage. Next, you will be asked to perform four sets of tasks both on and off-screen. The systems will differ in terms of touchless clutching (locking/unlocking the system). After each set of tasks, you will be asked to fill out a small usability questionnaire. Finally, you will be asked to take part in a short interview and fill out a small satisfaction and usability questionnaire.

We expect the experiment to take 30-40 minutes over the course of a single session. We do not anticipate any risks to participants.

During the time you interact with the tool you will audio recorded, and a person will observe you. Also, your screen will be captured for analysis purposes. The collected data will be analyzed and may be used to improve the tool. Also, some results may appear in scientific publications. Those recordings will be analyzed and the results presented in such a way that it will be impossible to trace the material back to you. The original data (audio and screen recording) will be kept secure and will not be circulated. It will not be used in any oral presentation of the work or in any medium that allows people to freely download individual recordings. A paper record associating recordings with individual contact details will be kept separately from the recordings so that you will have the option to withdraw should you wish to do so.

Your participation is entirely voluntary, and you can withdraw at any time without penalty. You don't need to provide a reason for your decision to do so. In order to withdraw, simply email the researcher giving the email you used to sign up with, indicating that you wish to withdraw. While recordings can be deleted, it will not be possible to be removed from the data underlying the study following publication of the results.

The data will be analyzed in order to gain insight into the effect of various approaches to gestural control on touchless interaction with PACS. We plan to publish the results of our research in academic journals and conference proceedings. We will do this in a way which does not identify you, or any other individual participant. No audio or video recordings will be made available to anyone other than the immediate research team at Trinity College Dublin, nor will any such recordings be replayed in any public forum or presentation of research without anonymization of any individuals portrayed.

At the end of the experiment we will explain more of the reasons behind the experiment and give you a chance to ask questions. Copies of all publications resulting from the study will be available free of charge on request.

While it is unlikely that illicit activities would be disclosed, if you do so, we would be obliged to report the to the appropriate authorities.

We may use direct quotations from participants. If you wish to review transcriptions for validity, then simply let the researcher know you wish to do this.

Conflict of interest:
This is not an independent evaluation. The researchers, who are also the software developers, are conducting this research to explore the usefulness of this form of interaction.

Appendix H

Prototype Study: informed consent

INFORMED CONSENT FORM - Prototype Evaluation

Background of study:
The purpose of the study is to evaluate touchless interfaces to Picture Archiving and Communication Systems (PACS) systems. The evaluation will involve using a computer program via touchless interaction methods, e.g. hand gestures, voice commands, engagement recognition, and position recognition, filling in some questionnaires on my experience, and discussing my actions in a retrospective interview. The study is being carried out in partial fulfillment of the requirements for a postgraduate research degree in Computer Science at Trinity College Dublin.

By signing this form, I am consenting to participate in the evaluation, and am making the following declarations:

I understand that my interaction with the system will be logged and my screen and audio will be recorded.

I understand that the recorded data will be made anonymous and be accessible to a small team of researchers (i.e. a research team of 4 people) for study purposes.

I understand that even though the recorded data will be made anonymous it will not be available to other people outside of the mentioned research team.

I understand that participation in this study is entirely voluntary and that I can withdraw from the study at any time without giving a reason.

I understand that I am free to ask any questions at any time. I am free to withdraw without providing a reason, or to discuss my concerns with the experimenter.

I agree that my data is used for scientific purposes and I have no objection that my data is published in scientific publications in a way that does not reveal my identity.

I understand that I can make subsequent contact with the leading researcher Sean Cronin (via email: secronin@tcd.ie) if I wish to obtain a copy of any papers derived from the research.

In the extremely unlikely event that illicit activity is reported to me during the dialogue I will be obliged to report it to appropriate authorities.

I understand that if I, or anyone in my family, have a history of epilepsy then I am preceding at my own risk.

I understand that the information provided by me will be held anonymously so that it is impossible to trace this information back to me individually. In keeping with TCD guidelines on good research practice, this information will be retained for 10 years for research verification purposes.

I, ________________________________, consent to participate in this study conducted by Sean Cronin in the School of Computer Science and Statistics, Trinity College, Dublin under the supervision of Dr. Gavin Doherty, and I confirm that I am over 18 years old.

I, ________________________________, consent to being audio and screen recorded as a part of this prototype evaluation.

Statement of investigator's responsibility: I have explained the nature and purpose of this research study, the procedures to be undertaken and any risks that may be involved. I have offered to answer any questions and fully answered such questions. I believe that the participant understands my explanation and has freely given informed consent.

Contact details of the Lead Researcher:
Name: Sean Cronin Email: secronin@tcd.ie
Address: School of Computer Science & Statistics, Trinity College Dublin, Dublin 2, IRELAND

Researcher signature: _______________ Date:____________________

Appendix I

Prototype Study: per-modality questions

Questionnaire Guideline

TRINITY COLLEGE DUBLIN

The questionnaire employs the NASA TLX method. Question 7 (based on the work of Tan et al. 2013) aims to bring additional insight by providing insight to perceived system utility.

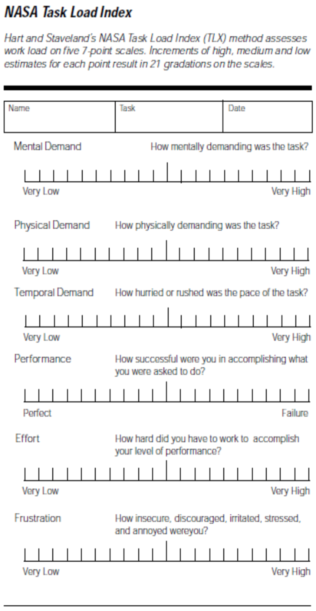
Each question is optional. Feel free to omit a response to any question; however, the researcher would be grateful if all questions are responded to.

NASA TLX:

1. How mentally demanding was using the system/task?
2. How physically demanding was using the system/task?
3. How hurried or rushed was the pace of the task?
4. How successful were you in accomplishing what you were asked to do?
5. How hard did you have to work to accomplish your level of performance?
6. How insecure, discouraged, irritated, stressed, and annoyed were you?

Additional questions:

7. Do you feel that a touchless interface such as the one tested would be useful in your work?



NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Appendix J

Prototype Study: post-experiment questions

Post Experiment Interview Guideline

The conducted interview will be a semi-structured one. The participants will be asked general questions and will be encouraged to develop them by giving more details or their personal experiences. Each question is optional. Participants are free to omit a response to any question.

Background information:

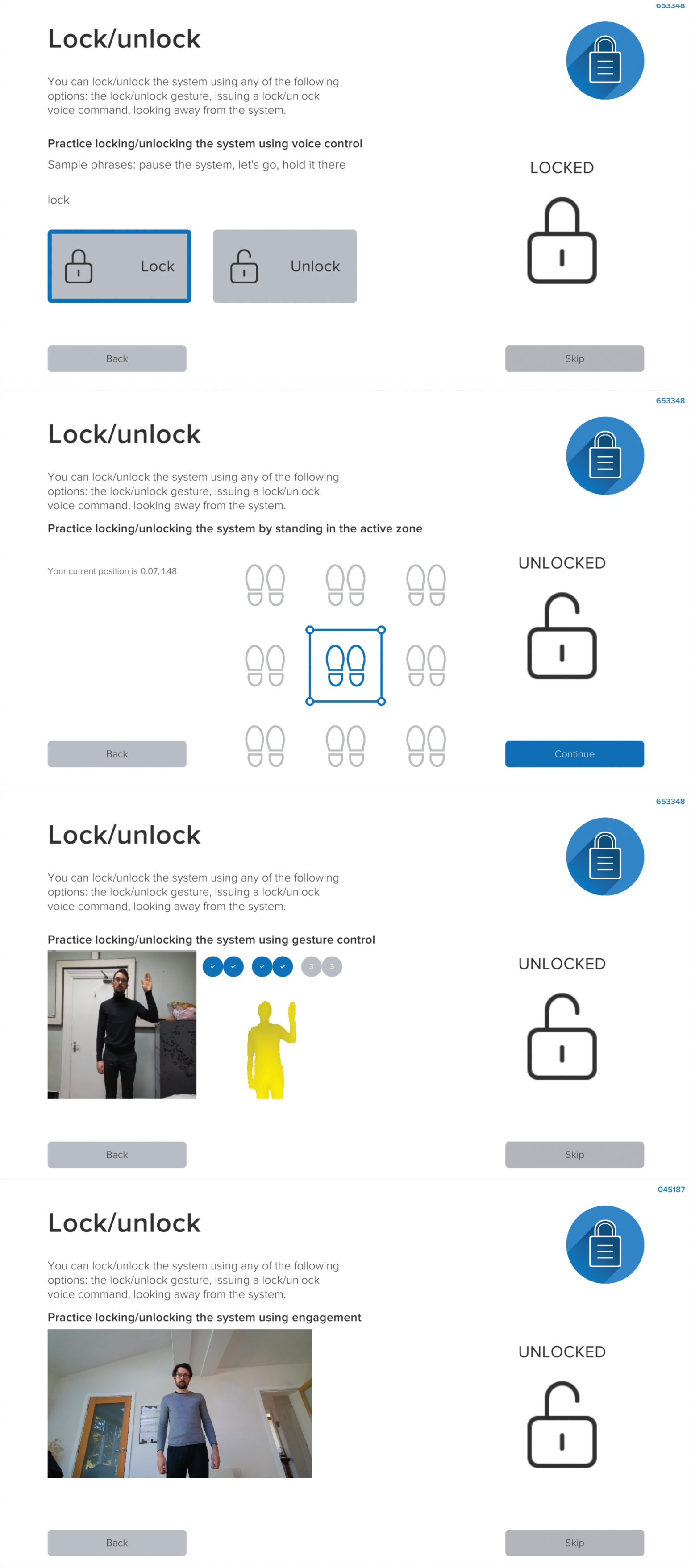
- What is/was your position in the hospital?
- Male/Female?
- How long have you been working with PACS?
- Which is your dominant hand?
- Have you any experience with touchless systems, such as the Xbox Kinect?

Touchless Clutching:

- Overall, how would you rate your satisfaction with the touchless system?
- Overall, did you like or dislike using touchless interaction?
- Did you find it tiring to use touchless control over the duration of the experiment?
- How easy did you find touchless control to use as compared to mouse and keyboard input?
- What issues did you experience using the system?
- Where do you envisage touchless interaction being useful?
- Could you see yourself using touchless interaction?
- What barriers to adoption of touchless interaction do you anticipate?
- Have you any other comments?

Appendix K

Prototype Study: touchless clutching training interface



Lock/unlock

You can lock/unlock the system using any of the following options: the lock/unlock gesture, issuing a lock/unlock voice command, looking away from the system.

Practice locking/unlocking the system using voice control

Sample phrases: pause the system, let's go, hold it there

lock

Lock/unlock

You can lock/unlock the system using any of the following options: the lock/unlock gesture, issuing a lock/unlock voice command, looking away from the system.

Practice locking/unlocking the system by standing in the active zone

Your current position is 0.07, 1.48

Lock/unlock

You can lock/unlock the system using any of the following options: the lock/unlock gesture, issuing a lock/unlock voice command, looking away from the system.

Practice locking/unlocking the system using gesture control

Lock/unlock

You can lock/unlock the system using any of the following options: the lock/unlock gesture, issuing a lock/unlock voice command, looking away from the system.

Practice locking/unlocking the system using engagement

Figure K.1: Each of the four clutching modalities was trained individually to the user's satisfaction.