Developing a high-performance patient-to-image registration system for robotic cochlear implant surgery

Through the human-robot interaction-focused design method leading to a multi-modal, multi-feedback admittance control, touch-based pair-point bone-anchored fiducial registration and its performance measured in accuracy, workload, usability and trust.

Master Thesis Robotics

Susanna Thamar Halman

Delft University of Technology

TU Delft

Developing a high-performance patient-to-image registration system for robotic cochlear implant surgery

Through the human-robot interaction-focused design method leading to a multi-modal, multi-feedback admittance control, touch-based pair-point bone-anchored fiducial registration and its performance measured in accuracy, workload, usability and trust.

by

Susanna Thamar Halman

Instructor:	Luka Peternel
	Arash Arjmandi
Institution:	Delft University of Technology
Place:	Faculty of Mechanical, Maritime and Materials Engineering, Delft
Project Duration:	March 2022 - November 2022.

TU Delft

Abstract

This thesis presents the research to design a successful high-accuracy, sub-millimetric registration method for an autonomous robot equipped to drill bone for cochlear implant (CI) surgery. Its performance, and thus success, is measured in accuracy, workload, usability and trust. While state-of-the-art (STOTA) research lacks the inclusion of human-robot interaction (HRI) related design requirements, this thesis demonstrates its significance. The HRI-focused design led to a successful multi-modal, multi-feedback admittance control, touch-based pair-point bone-anchored fiducial registration method.

A successful high-accuracy registration method is critical for image-guided robotic microsurgery as it contributes directly to its overall surgical precision. The inclusion of HRI-related requirements is concluded to be essential for the success of registration methods for surgical robots due to their inevitable involvement in a hospital workflow consisting of human teams and their lack of autonomous capabilities in all required tasks.

Gaining knowledge through theoretical and empirical research; utilising exclusion and scoring criteria to select the best specific registration method concept; and performing an extensive error- and human-factor analysis to determine the design implementations led to the design of a multi-modal multi-feedback admittance control, touch-based pair-point bone-anchored fiducial registration method. The design allows operators to execute a localisation task in collaboration with the robot. This design concerned a human operator physically guiding the robot end-effector to four bone-anchored fiducials attached to a skull-resembling setup to execute their localisation. The robot control was based on an admittance controller that enables the human to move the robot through the sensed interaction force at the robot's end-effector. Consequently, if the operator exerted a forwarded force on the robot, it moved in that direction proportionally with the preset admittance parameters.

This guidance control contained five modes of control: (1) translational, (2) rotational, (3) fixed, (4) safety and (5) registration mode. Each mode allows the operator to control the robot as needed to achieve high accuracy, usability, trust, and a low workload. To switch between modes, the human operator utilised a foot-operated interface with several buttons, each dedicated to a specific function.

The translational mode had a fixed orientation and permitted free movements of various speeds along translational axes and was used for correctly arranging the end-effector to the fiducial centre in the x-, y- and z-axis. The rotational mode allowed orientation movement while preserving a fixed translational position and was employed for aligning the end-effector perpendicular to the fiducial centre. The fixed mode included a fixed orientation and translational position and was used for filtering out perturbations presented by the human operator. The safety mode allowed the robot to with-tract to a safety position and orientation in case of operator inactivity and was utilised to ensure patient safety. The registration mode allows for saving the end-effector position, hence localising the fiducials.

The feedback provided to the operator consisted of visual feedback, auditory feedback and a step-by-step graphical user interface (GUI). The visual feedback permitted colour coding of the modes of control and safety thresholds regarding force, torque and workspace. This was used to build operator awareness and control large errors introduced by operator input. The auditory feedback included a sound when localisation of the end-effector was completed and was utilised to form operator awareness and prevent timing errors. The GUI contained safety checklists, workflow guidance and operator performance feedback and was used to minimise operator variations.

To compare and benchmark our design, the baseline included a similar STOTA non-HRI focused single-mode no feedback admittance control touch-based pair-point bone-anchored registration method.

This baseline method allowed a human operator to guide the robot end-effector in the translational axis and save fiducial localisation positions in cooperation with a second human operator. The comparison study involved a quasi-experiment of two groups with 13 and 14 participants, respectively. The groups executed repeated fiducial localisation and were compared on their system accuracy, workload, usability and trust.

The system accuracy was measured in fiducial localisation error (FLE) and target registration error (TRE) based on repeated measurements of each fiducial. The workload was estimated quantitatively in average operator force and qualitatively with a NASA TLX survey. Usability was calculated with the Computer System Usability Questionnaire (CSUQ) and System Usability Scale (SUS). The Trust Perception Scale-HRI measured trust. The design achieved the second-highest accuracy from all discovered STOTA registration methods. Furthermore, it reduced the operator workload and improved usability and accuracy compared to a non-HRI-focused design.

This thesis concludes that an HRI-focused symbiotic design provides a successful high-accuracy submillimetric registration method for robotic surgery as measured by accuracy, workload and usability.

To a future where humans and machines live together in harmony.

The comparison study in this thesis is performed in agreement with the TU Delft Human Research Ethics Committee (HREC).

Acknowledgement

This master thesis would, first and foremost, not have been possible without the opportunity provided by Anupam Nayak, founder and CEO of Eindhoven Medical Robotics (EMR), to be part of EMR during this thesis. I also want to voice my gratitude to my supervisors, Luka Peternel and Arash Arjmandi. They have coached me throughout this thesis with numerous discussions, knowledge, insights and feedback. Besides, I thank other employees of EMR that supported me in this project, specifically Chimey, Eiz and Siddarth, who helped me with troubleshooting, reflecting, bug fixing and implementing and connecting hardware implementations. Furthermore, a special thanks to Martin and Bram, who helped perform the error and human-factor analysis and Sean, who acted as an excellent sparring partner. Moreover, I would like to thank all EMR employees for their kindness, making me feel at home, willingness to make time whenever needed and participation in the tests. Special thanks go to David Abbink and Jenny Dankelman of the TU Delft for their time, reflectance and views at the start, to tap into their expertise to instruct me on the topics.

I am grateful for the help, support and backing of my family, friends and girlfriend Jocelyn, who have always believed in me, supported me in different ways and made me laugh, smile and believe in myself throughout this thesis. Specifically, my friends Hugo, Berry, Timo and Pepijn for providing feedback and studying together to make the thesis time feel less lonely and with whom I became good friends during my master's despite Covid.

Nomenclature

Term	Description
Admittance control	A way to control the position of a robot arm based on external force and torque
Autonomous	A robot that acts without human control
Cochlear implant	An electronic device that can help to provide a sense of sound to a person with hearing loss
End-effector	A tool attached to the end of the robot arm
Extrinsic	Not part of the essential nature of a patient
Facial recess	A small recess bounded by the facial nerve and chorda tympani
Fiducial markers	A small object such as a screw, placed in or on the body to mark an area
Registration	Finding a transformation matrix that provides the relationship between image data and the body
Intraoperative	Occurring or performed during the course of a surgical operation
Intrinsic	Naturally belonging to the patient
Invasiveness	The scale that describes the amount, size, and the corresponding burden on the patient related to the nevessity for incision
Kolmogorov-Smirnov test	A non-parametric goodness-of-fit test, used to determine wether distributions differ from a normal distribution
Landmarks	Meaningful locations that can be unambiguously defined and repeatedly located with a high degree of accuracy
Manipulator	An arm-like structure joined to the body of a robot and is used to execute tasks
Mann-Whitney U test	A non-parametric alternative test to determine if 2 groups are significantly different from each other
Mastoidectomy	Surgery to remove cells in the hollow, air-filled spaces in the skull behind the ear within the mastoid bone
Pair-points	A point resembling the same position in two different coordinate frames, creating a pair of two points
Phantom	A specially designed object that is utilised as a “stand-in” for human tissue
Postoperative	During, relating to, or denoting the period following a surgical operation
Preoperative	Denoting, administered in, or occurring in the period before a surgical operation
Stereotactic surgery	A surgical intervention that makes use of a 3D coordinate system to locate targets from the body

Abbreviation	Description
BFR	Bone Fiducial Registration
CI	Cochlear Implant
CT	Computed Tomography
CIS	Computer-Integrated Surgery
CSUQ	Computer System Usability Questionnaire
DH	Denavit-Hartenberg
DOF	Degree-of-Freedom
DTC	Distance-To-Centroid
EA	Error-analysis
EPE	Entry-Point Errors
FEA	Finite Element Analysis
FLE	Fiducial Localisation Error
FM	Fiducial Marker
FRE	Fiducial Registration Error
GT	Ground Truth
GUI	Graphical User Interface
HFA	Human-factor Analysis
HFACS	Human Factor Analysis and Classification System
HITL	Human-in-the-loop
HOTL	Human-out-the-loop
HRC	Human-robot Collaboration
HRI	Human-robot Interaction
ICP	Iterative Closest Point Algorithm
IGS	Image-guided Surgery
IQR	Interquartile Range
LE	Localisation Error
LORA	Levels Of Robot Autonomy
LSR	Laser Surface Registration
MD	Mental Demand
MIS	Minimally Invasive Surgery
MRI	Magnetic Resonance Imaging
OR	Operating Room
PBSM	Physics-based Shape Matching
PC	Point Cloud
PD	Physical Demand
pHRI	Physical Human-robot interaction
RMS	Root Mean Square
RMSE	Root Mean Square Error
SD	Standard Deviation
SLE	Surface Localisation Error
STOTA	State-of-the-art
SUS	System Usability Scale
TCP	Tool Center Point
TRE	Target Registration Error
UI	User-interface

List of Figures

List of Tables

Contents

1. Introduction

This chapter presents the problem underlying the reason for developing an HRI-focused high-performance registration method and introduces the research question to be answered. It briefly provides the essential knowledge and definitions, the thesis goal, scope and requirements, the methodology, the contributions, and its outline.

1.1. Surgical robots for cochlear implant surgery

According to WHO (2021), nearly 80% of people with disabling hearing loss live in low- and middle-income countries. This hearing loss can considerably impact the quality of life, educational levels and productivity. Unaddressed hearing loss is estimated to costs humanity 980 billion USD annually. This problem will expand, as demonstrated in Figure 1.1. Consequently, action is needed. For people with severe hearing loss, a cochlear implant (CI) can be used, which requires surgery. One major barrier to receiving the care needed is the scarcity of facilities, surgeons, expertise and other resources required to execute surgeries (Grimes et al., 2011).

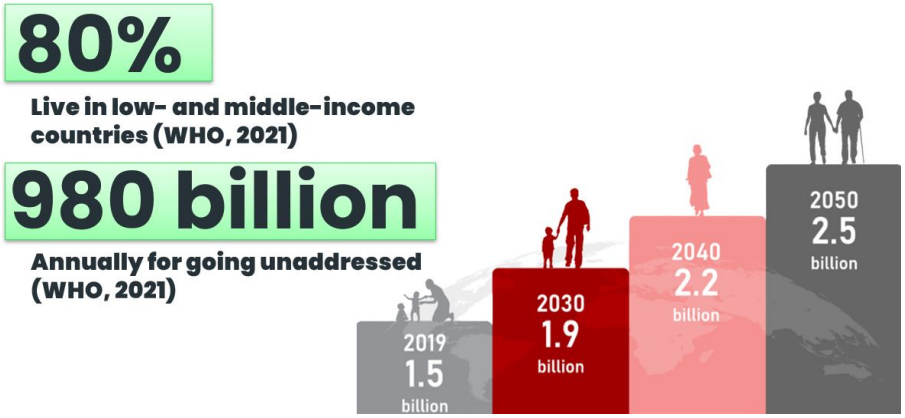


Figure 1.1: Statistics related to hear loss according to WHO (2021).

A potential strategy to overcome this barrier is the deployment of surgical robots. Robots in surgeries have acquired much awareness in recent years due to their precision and tirelessness. One company involved in surgical robots is Eindhoven Medical Robots (EMR). They aspire to provide an autonomous robot for CI surgery to solve the shortage of surgical care.

The trend in robot usage goes hand-in-hand with the trend of minimally invasive surgery (MIS) (J. Liu et al., 2019). Invasive is a scale which describes the amount, size, and corresponding burden on the patient (Alam and Rahman, 2016). MIS creates less trauma and reduces hospital stay and recovery times (Machetanz, Grimm, Wang, et al., 2021; Westebring–van der Putten et al., 2008). MIS faces additional challenges, which can be mitigated with the acquisition of computer-integrated surgery (CIS) or image-guided surgery (IGS) (Taylor, 2006). Surgical robots use IGS by acquiring patient images to conduct surgical plans.

Once the patient arrives in the operating room (OR), the robot must align its surgical plan to the patient's actual position. This process is called registration. Registration is crucial; a robot can perform surgery autonomously exclusively with adequate registration.

1.1.1. Patient-to-image registration

Registration is the procedure of locating the patient with respect to the robot frame, or *"the determination of a relationship between image data and the body"* (Greenwood and Vallee, 2021). A transformation matrix correlates the coordinate frame of the patient images with the coordinate frame of the patient in the OR, hence providing this relationship. Registration attempts to match points, landmarks or point clouds (PCs) between two different coordinate frames and significantly impacts the overall surgical accuracy (Machetanz, Grimm, Wang, et al., 2021).

Figures 1.2 and 1.3 illustrate two examples of registration. In Figure 1.2, one coordinate frame is supplied by a radiographic image and the other by the OR coordinate frame. Those images are rotated and translated to overlay them in the best way possible, minimising the distance error between external markers placed on the patient's head. External markers are placed outside the human body and are assumed to mark the same position in both coordinate frames. Registration that utilises any form of external markers in addition to the patient's anatomy is called extrinsic registration.

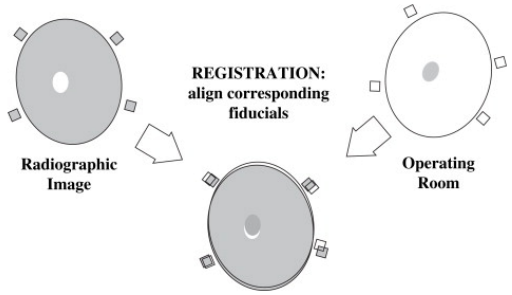


Figure 1.2: An example of registration, finding the transformation, including rotation and translation, between two coordinate frames to overlay them in the best way possible, minimising the distance error between external markers in both frames that resemble the same position ("Image-Guided Technique in Neurotology | Ento Key", n.d.).

In Figure 1.3, a coordinate frame in the computed tomography (CT) image is matched with the OR coordinate frame based on features inside the human anatomy, hence not utilising external additions. Registration that utilises internal anatomical markers or other human characteristics, such as its surface, to match coordinate frames is called intrinsic registration.

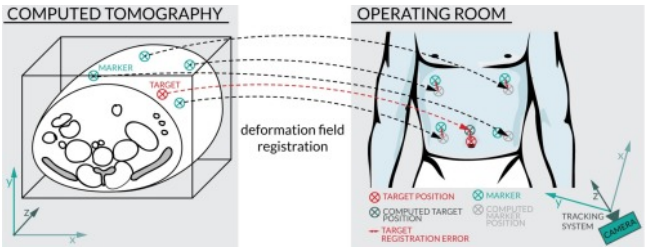


Figure 1.3: An illustration of registration finding the transformation between two coordinate frames using the minimisation of feature positions inside the human anatomy. Green wheel crosses resemble the anatomical markers (2017). The 3D coordinate frame of the CT image is transformed to provide the minimum distance error between these internal markers, represented by the difference between the green and light grey wheel crosses.

Different registration methods have various advantages, disadvantages and accuracies. As the examples display, one can choose either internal or external markers resembling intrinsic or extrinsic registration. Extrinsic registration is the registration process that relies on external entities connected to the patient (Alam et al., 2016). These objects are called external markers or frames. On the contrary, intrinsic registration relies on internal anatomical features. Hence, any addition required to the patient's body resembles extrinsic registration, whereas intrinsic registration only requires the human body to perform registration.

Extrinsic methods are either invasive or non-invasive, where invasive yields higher accuracy but adds to the patient's burden as it requires additional surgical cuts. Intrinsic methods are point-based (landmark-based), surface-based, and voxel-based registration methods. Point-based uses anatomical or geometrical points to match, such as Figure 1.3. Surface-based utilises the human body surface for matching. Voxel-based methods match 3D voxels in both coordinate frames (Alam and Rahman, 2016; Alam et al., 2016). Extrinsic registration is predominantly preferred over intrinsic registration since external objects provide better visibility and are thus more easily distinguished. Further, extrinsic registration is faster as no complex algorithms are demanded (2016). However, extrinsic registration methods require an additional marker which demands more planning and increases the workflow complexity.

CI surgery concerns moving alongside crucial anatomical structures; hence sub-millimetric accuracy is required. The most vital structures are the facial nerve and the chorda tympani nerve, which cross each other in the facial recess. This recess holds a size of about 2.4 - 5.7mm (J. Wang et al., 2020). The drill dimensions demand an additional 0.5mm space, leaving an obligatory accuracy range of 1.9 - 5.2mm, excluding safety margins. Image registration accuracy directly influences the system accuracy. A high accuracy is therefore required to leave the highest possible sub-millimetric accuracy tolerance for the registration itself. Gerber et al. (2013) emphasises the need for high-accurate image-to-patient registration: *"Perhaps the greatest remaining limitation to image-guided robotic microsurgery is a sufficiently accurate and verified patient-to-image registration"*.

In summary, one central issue that should be solved before CI surgical robots can be successful, is developing a high-accuracy, sub-millimetric registration method.

1.1.2. Human-robot interaction-focused design

The existing registration literature deduced that most papers exclusively focus on the technical aspect of registration, summarised in Figure 1.4.

Figure 1.4: A schematic outline of the steps to execute patient-to-image registration. First, locate the landmarks or PC in the image space. Then locating the identical landmarks or another PC in the patient or robot space. Validation of the localisation methods is required, whereafter registration can be executed, resulting in a transformation matrix.

That the robot performs registration or surgery autonomously does not indicate it can execute other tasks autonomously (Beer et al., 2014). Registration is placed inside a more extensive hospital workflow, shown in Figure 1.5. This general hospital workflow incorporates tasks such as fixating the patient to the table, sterilising the patient with drapes, seducing the patient, and checking the patient afterwards. These are all tasks that the robot cannot perform presently.

Figure 1.5: A schematic overview of the steps required to perform patient-to-image registration. Starting with the pre-operative phase, where the diagnosis is constructed. Then the patient image is gathered to conduct a surgical plan. The intra-operative phase initiates with registration, followed by surgery. The post-operative phase consists of check-ups.

Further, robots have weaknesses and strengths that are complementary to human strengths and weaknesses, and automation is known to form additional problems such as overreliance or under-reliance on a system (Crouser et al., 2013; Fitts, 1951). Since the medical field requires minimisation of errors

and failures, a human-in-the-loop (HITL) method is considered best. Then the skills of both the robot and humans can be combined accordingly, mitigating liability-related problems.

Accordingly, the robot should collaborate with a human team to execute CI surgery successfully. This collaboration will introduce different system requirements related to human-induced errors compared to a system without collaboration. Concentrating exclusively on the technical factors restricts the solutions in their abilities to be adopted by hospitals and become successful.

Thus it is essential to incorporate **HRI-related** factors in addition to the **technical** requirements. These factors lack emphasis in STOTA methods even though these instantly influence the accuracy and adaptation of any registration method.

HRI-related challenges
As noted, the robot must collaborate with a team of surgeons and other medical staff to conduct patient-to-image registration, providing the need to optimise HRI. In registration, the human and robot can interact in different ways, and the robot can have different levels of robot autonomy (LORA) in this collaboration (Beer et al., 2014). The type of interaction and the LORA provide different added challenges and, thus, design requirements. The main challenges are acceptation and adaptation, safety and human errors. These challenges should be mitigated in a successful design.

Acceptation and adaptation
Acceptation and adaptation are significant for the registration method. If a well-designed method is hard to adapt and accept, the method is of no use. Factors that influence the acceptation and adaptation of radical technologies such as automation are expressed by Parasuraman and Riley (1997). From these, the most important factors are considered *workload*, *usability* and *trust*.

Workload refers to the quantity of work humans have to do, executing a task together with a robot within available time (Harriott, 2015; Parasuraman and Riley, 1997). Usability is defined as *"the extent to which a product can be used by specific users to achieve specific goals with (i) effectiveness, (ii) efficiency, and (iii) satisfaction in a specified context of use"* (ISO, 2011). Trust in a task with a clear goal is defined as an attitude which retains the belief that the collaborator will perform as expected and can, within the limits of the designer's intentions, be relied on to accomplish the goals (M. Lewis et al., 2018).

Safety
Safety is vital for a robot that collaborates with a human team (Abdelaal et al., 2019). Particularly for surgical robots, safety is essential. Errors can significantly harm the patient or surgical staff when not mitigated correctly. One method to advance robot safety is through its control method.

Human factors
Since humans collaborate with the robot, they will present errors in the workflow and significantly affect the registration outcome. Human errors can be introduced on different levels in the workflow, and the Swiss Cheese model expresses these levels by Shappell and Wiegmann (2000). These errors incorporate both active and latent failures. The latter can go disregarded for a prolonged time.

In summary, another central issue that should be solved before CI surgical robots can be successful is the lacking inclusion of registration design requirements that target HRI-related challenges.

1.2. Problem Statement and Goal

In summary, the CI surgery robot of EMR requires the development of a state-of-the-art **sub-millimetric accuracy** registration method that regards both **technical** and **human-robot interaction-related** factors. The **goal** of this thesis is ***designing an HRI-focused, sub-millimetric accuracy patient-to-image registration method for robotic CI surgery that achieves high-performance measured in accuracy, workload, usability and trust.***

1.2.1. Thesis scope and requirements

Due to time limitations, several choices had to be made:

1. The design excludes the image localisation part of registration, solely focussing on robot localisation
2. The deformable surface-based and voxel-based registration methods were excluded due to their lack of application
3. Systems performance is measured through accuracy, workload, usability and trust. These were believed to be the most important and well-known metrics relevant to this thesis
4. No excessive and methodological literature research has been performed
5. The thesis excludes latent human factor errors as influencing factors on the system's performance

Requirements for this thesis stem from the medical application which demands extensive justifications for design choices. Consequently, a clear reasoning path should be followed in designing the registration method, evaluating different concepts before settling on one final design. The chosen path pursues the double diamond design method described by Design Council (2019). Another requirement is the inclusion of clinical and general factors as design criteria, such as effective control, minimising the footprint, limiting its maximum force and speed, costs, maintenance, number of additional parts and installation time. This thesis will regard them as HRI-related requirements. Hence a final registration method will be selected from a literature search, utilising exclusion and scoring criteria to select the best specific registration method concept; and performing an extensive error- (EA) and human-factor analysis (HFA) to determine the design implementations.

There are different concepts for the word error, but in this project, the term error is used broadly, which means both a technical error or failure and human mistakes during the registration process. Hence it could be defined as "any concept that leads to lower system performance". The term "failure" is mainly used related to human-induced errors. Patient localisation is referred to in this thesis as physical space localisation, robot localisation or patient localisation.

1.3. Research Question

The corresponding research question states

"How can a human-robot interaction-focused method be designed for robotic cochlear implant surgery that acquires high system patient-to-image registration performance, measured in accuracy, workload, usability and trust?"

This research question is divided into two sub-questions:

Q1:*"What specific registration method concept performs best according to technical and human-robot interaction criteria for high-accuracy image-to-patient registration for robotic cochlear implant surgery"*

Q2:*"Does the proposed human-robot interaction-focused design improve the human-robot (system) performance in executing image-to-patient registration, measured in accuracy, workload, usability and trust ?"*

1.4. Methodology

Q1 has been answered by gaining knowledge through theoretical and empirical research; utilising exclusion and scoring criteria to select the best specific registration method concept; and performing an extensive error- and human-factor analysis to determine the design implementations. The resulting design of a multi-modal multi-feedback admittance control, touch-based pair-point bone-anchored fiducial registration method allowed operators to perform localisation tasks with the robot collaboratively. This design concerned a human operator physically guiding the robot end-effector through admittance control with five different control modes. The feedback provided to the operator consisted of visual feedback, auditory feedback and a step-by-step graphical user interface (GUI).

Q2 has been answered by comparing the above design to a baseline. The baseline included a similar STOTA non-HRI focused single-mode no feedback admittance control touch-based pair-point bone-anchored registration method. The comparison study involved a quasi-experiment of two groups, and both design performances are quantified by testing it along the most critical factors that influence its performance: *accuracy*, *workload*, *usability*, and *trust*.

The system accuracy was measured in fiducial localisation error (FLE) and target registration error (TRE) based on repeated measurements of each fiducial. The workload was estimated quantitatively and qualitatively through average force and a NASA TLX survey, respectively. Usability was calculated with the Computer System Usability Questionnaire (CSUQ) and System Usability Scale (SUS). The Trust Perception Scale-HRI measured trust.

1.5. Contributions

This thesis delivers a clear engineering and scientific contribution. Its engineering contribution lies in designing a STOTA registration method incorporating HRI-related factors and acquiring sub-millimetric registration accuracy. These design performances have yet to be perceived in literature.

The scientific relevance of this thesis lies in addressing the literature gap concerning human influence on registration performance and including HRI-related factors in registration. Additionally, it provides a new understanding of registration accuracy and all factors influencing this accuracy by performing an EA and HFA. Lastly, it contributes to the HRI field by applying an HFA in the field of medical robots.

1.6. Thesis outline

This chapter introduces the problem underlying this thesis and provides the thesis research question, goal, scope and requirements. It also delivered a short preface to the methodology, which will be elaborated on in Chapter 2. The background information, including theoretical and contextual background, is discussed in Chapter 3. The answer to Q1 is then examined in Chapter 4. The final design implementations are described in Chapter 5. The quasi-experiment baseline comparison study results are summarised in Chapter 6. In Chapter 7, an extensive discussion of the results, design and project is supplied along with recommendations, and Chapter 8 concludes this thesis by answering the stated research questions.

2. Methodology

This chapter presents the methodology followed to answer the research question.

The research question, *"How can a human-robot interaction-focused method be designed for robotic cochlear implant surgery that acquires high system patient-to-image registration performance, measured in accuracy, workload, usability and trust?"*, was answered by dividing the thesis into two main phases, the designing phase and the testing phase.

The first phase answers sub-question Q1: *"What specific registration method concept performs best according to technical and human-robot interaction criteria for high-accuracy image-to-patient registration for robotic cochlear implant surgery?"*. This phase is executed using the double-diamond design method and divided into a research phase and a design phase, as pictured in Figure 2.2. It answers Q1 by gaining knowledge through theoretical and empirical research, utilising exclusion and scoring criteria to select the best specific registration method concept design; and performing an extensive EA and HFA to determine the design implementations.

The second phase, the testing phase, provides an answer to the second sub-question Q2: *"Does the proposed human-robot interaction-focused design improve the human-robot (system) performance in executing image-to-patient registration, measured in accuracy, workload, usability and trust?"* . It encloses a comparison study that involved a quasi-experiment of two groups of human operators that perform registration with two distinct designs. The design performances are quantified by testing the most critical elements that influence its performance: *accuracy*, *workload*, *usability*, and *trust*.

Consequently, this thesis consists of two separate primary phases: the designing phase and the testing phase. Which can yet be diverged into three phases: research, design and testing, summarised in Figure 2.1. The phases and how they answer the questions are clarified in more detail below.

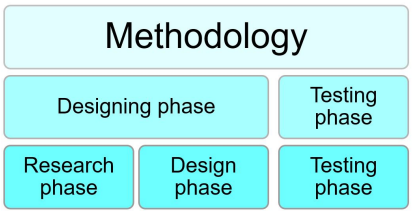


Figure 2.1: The methodology is separated into a designing phase and a testing phase. The designing phase pursues the double diamond structure and is additionally diverged into research and design phases.

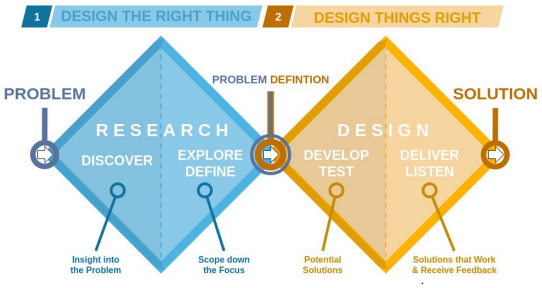


Figure 2.2: An recap of the double diamond model defined by Design Council (2019), which is a design process that consists of a research phase which contains divergent thinking and a design phase which contains convergent thinking. The research phase is the first diamond and guarantees that the right thing is designed, and the design phase ensures that the design is performed correctly (2019).

2.1. Designing phase

To answer Q1 and provide a final best registration design, the double-diamond design process was used as described by Design Council (2019) and shown in Figure 2.2. The double diamond is a design process consisting of a research phase incorporating divergent thinking and a design phase containing convergent thinking. The research phase is the first diamond and guarantees that the right thing is designed, and the design phase ensures that the design is performed correctly (2019). The double diamond was applied and led to the method illustrated in Figure 2.3, where the research and design phases are split into four and three distinct steps, respectively.

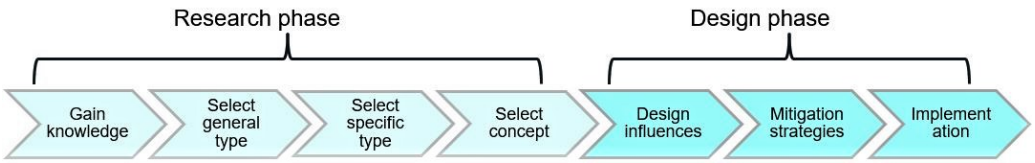


Figure 2.3: The various steps in the research and design phase according to the double diamond design method.

2.1.1. Research phase

The research phase presents the best concept that can be elaborated in the design phase. The research phase is separated into a discover phase followed by a defining phase (2019). The discover phase, step 1 in Figure 2.3, is utilised to acquire knowledge that can be applied in the defining phase, steps 2 till 4 in Figure 2.3, to determine the best specific registration method concept.

Discover phase: Gain knowledge

Knowledge was collected on registration methods, STOTA methods, and the different **technical** and **human-robot interaction-related** factors that can impact the design. The discovery phase consisted of two searches, executed theoretical and empirical, as depicted in Figure 2.4.

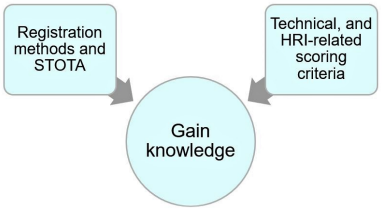


Figure 2.4: The two searches conducted in the discover phase to acquire the necessary knowledge to design a successful registration method that includes technical and HRI-related factors.

Information was gathered with the aid of Google Scholar, PubMed, Scopus, Mendeley, and Research Rabbit and through conversations with employees of EMR. A thorough search protocol on registration methods and STOTA, including the scope, search queries and inclusion and exclusion criteria, is provided in Appendix A. Due to the scope of this master's thesis, this search was not conducted.

Define phase: Select one final concept

The defining phase selects one final concept from the accumulated knowledge. This selection procedure regards two types of registration methods: general and specific. The specific registration type reflects how a general registration type can be performed in detail. For instance, surface-based registration would be a general type. However, within surface-based registration, there are numerous ways to gather the PCs, such as active tracking, laser surface scanning or using tracked pointers. The selection procedure delivered a general registration type, from which three specific registration types were selected, which led to four different concepts. Then five stakeholders scored these with scoring criteria to conclude the best concept. The defining phase was split into three steps: general, specific,

and concept selection, as portrayed in Figure 2.5. The applied criteria were drafted from the collected knowledge. From these criteria, the ones that have specific restrictions in this particular situation are utilised to equip exclusion criteria. This method is illustrated in Figure 2.6.

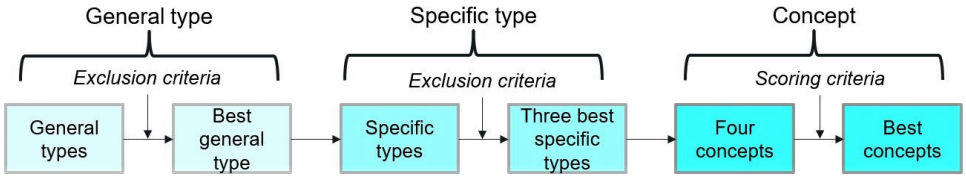


Figure 2.5: The define phase methodology to select the best concept. This phase is separated into three phases: general type, specific type and concept. The best general registration method was selected in the first phase by applying exclusion criteria to the known methods. In the second phase, exclusion criteria were applied to the specific types to find the three best specific types. Four concepts were drafted based on the three best specific types in the concept phase. Then five stakeholders scored these with scoring criteria to conclude the best concept.

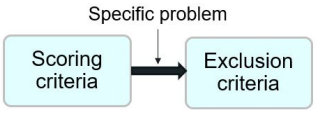


Figure 2.6: Formulating the exclusion criteria from the scoring criteria and the specific contextual background.

2.1.2. Design phase

The design phase concludes the best specific registration design implementations and delivers the final design. This phase is divided into a development phase followed by a deliver phase (2019). This design was yielded by formulating mitigation strategies for possible active errors and failures that can limit the design's success measured in accuracy, workload, usability and trust. Therefore the design influences were explored, from which mitigation strategies were formulated, scored and selected to mitigate external influences that negatively impact the design and enhance favourable effects. Ultimately, this led to design implementations, as shown in Figure 2.7.

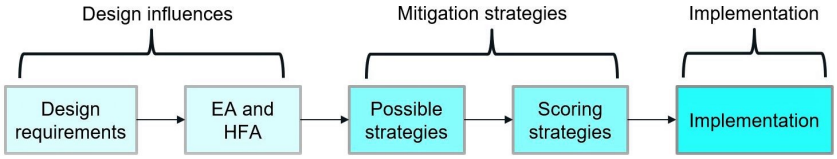


Figure 2.7: The design phase methodology to select the best concept, separated into design influences, mitigation strategies and implementations. The design influences were compiled by stating design requirements and performing an extensive EA and HFA on active errors and failures. The mitigation strategies were gathered by drafting many strategies and scoring them according to their advantages and risks. Implementation performs the final design implementations.

Develop: Provide multiple design options

The design influences were gathered by stating design requirements and performing an extensive EA and HFA on active errors and failures, as summarised in Figure 2.8. Many possible mitigation strategies for the encountered errors and failures were formulated.

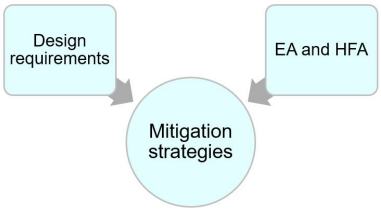


Figure 2.8: The develop phase accumulates mitigation strategies by stating design requirements and performing an extensive EA and HFA on active errors and failures.

Deliver phase: Implement and finalise the design
The mitigation strategies were scored according to their advantages and risks and conformed to the constructed technical and HRI-related factors, as illustrated in Figure 2.9. The selected implementations were enforced, finalising the design and answering Q1.

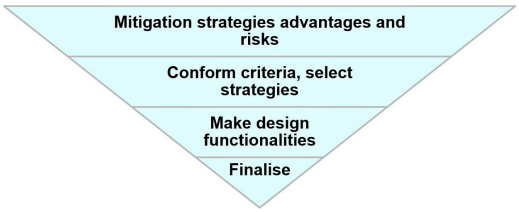


Figure 2.9: The deliver phase steps to scope down to one final design. The mitigation strategies were scored according to their advantages and risks and conformed to the constructed technical and HRI-related factors. These were translated to design functionalities and, lastly, implemented.

2.2. Testing phase

The first sub-question led to a suggested multi-modal multi-feedback admittance control, touch-based pair-point bone-anchored fiducial registration method design. Hence Q2 is answered by a comparison study that involved a quasi-experiment of two groups of 13 and 14 participants, respectively, comparing the design to a baseline. This baseline contained a similar STOTA non-HRI focused single-mode no feedback admittance control touch-based pair-point bone-anchored registration method. The groups executed repeated fiducial localisation of four bone-anchored fiducials attached to a skull-resembling setup. The design performances were quantified by measuring their *accuracy*, *workload*, *usability*, and *trust*. With system accuracy measured in FLE and TRE, workload estimated quantitatively in average operator force and qualitatively with a NASA TLX survey, usability measured with the CSUQ and SUS scores and trust measured with the Trust Perception Scale-HRI score as summarised in Figure 2.10.

	Metric 1	Metric 2
Accuracy	FLE	TRE
Workload	Average force	NASA TLX
Usability	SUS	CSUQ
Trust	Trust Perception Scale-HRI	

Figure 2.10: Summary of the measured metrics in this thesis related to the four categories accuracy, workload, usability and trust.

These four metrics result in four sub-sub-questions:
Q2.1: "Does the use of multi-feedback multi-modal admittance control improve the **accuracy**, measured by the FLE and TRE, in executing touch-based pair-point bone-anchored fiducial registration?"
Q2.2: "Does the use of multi-feedback multi-modal admittance control decrease the **human workload**, measured quantitatively by the average operator force and qualitatively by the NASA TLX survey, in executing touch-based pair-point bone-anchored fiducial registration?"
Q2.3: "Does the use of multi-feedback multi-modal admittance control improve the human-robot (system) **usability**, measured by the CSUQ and SUS scores, in executing touch-based pair-point bone-anchored fiducial registration?"
Q2.4: "Does the use of multi-feedback multi-modal admittance control improve the human **trust** in the system, measured by the Trust Perception Scale-HRI score, in executing touch-based pair-point bone-anchored fiducial registration?"

2.2.1. Task

The task in this quasi-experiment is the touch-based pair-point bone-anchored fiducial localisation task in 3D robot space. The key objective of the task is locating four fiducial screws eight times each. These screws are attached to an adult skull-resembling wooden setup as depicted in Figure 2.11. This task was executed using admittance control of the robot manipulator, which enables the human to move the robot through the sensed interaction force at the robot's end-effector. Consequently, if the operator exerted a forwarded force on the robot, it moved in that direction proportionally with the preset admittance parameters. The detailed block, fiducial screws and registration bit dimensions are located in Appendix B.

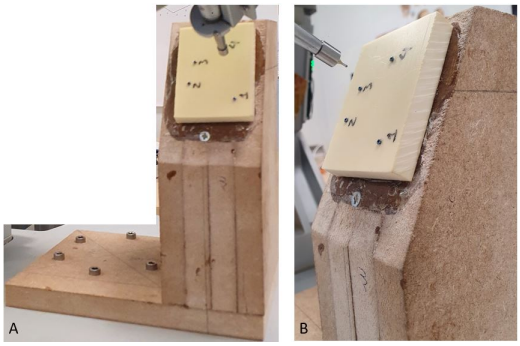


Figure 2.11: Setup of fiducials on a wooden block with dimensions similar to the human head. The detailed dimensions of the block can be found in Appendix B.

Admittance control for the new design was accomplished with the control loop described in Figure 5.7 and the gains defined in Table 5.3. The control loop described in Figure 3.28 was utilised for the baseline design. Setting up the robot with the correct Simulink real-time model and correctly positioning it to execute the tasks was performed by the researcher and is described in Appendix B.

2.2.2. Experimental conditions

The experimental conditions are required for the reproducibility of the experiment. Table 2.1 summarises the experimental conditions.

Table 2.1: Experimental conditions in the baseline comparison quasi-experiment between design 1 and design 2.

Comparison	Design 1	Design 2
Admittance control	Single-mode, translation only. Shown in Figure 3.28	Multi-modal, translation, rotational, fixed, safety and registration modes. Shown in Figure 5.7
Admittance control gains	K_t = 4.0 K_{setp} and K_{seth} = [0.05 0.05].	Depends on mode. K_t = 0.0-4.0, K_{setp} and K_{seth} = [0.0-0.2 0.0-0.2] and K_t = 0.0-0.7 Shown in Table 5.3.
GUI and feedback	No feedback. Shown in Figure 2.12	Multi-feedback consisting of GUI, auditory and visual feedback. Summarised in Figure 5.12 and Table 5.5
Operators	Two operators: one moving the robot, the other saving positions.	One operator and one supervisor. Operator moving the robot and saving positions.
Training	Steps: procedure and general terminology, roles and responsibilities, workflow, safety buttons, robot errors and case-studies. Summarised in Appendix C	Steps: procedure and general terminology, roles and responsibilities, workflow, safety buttons, robot errors and case-studies. Summarised in Appendix C
Task	Localising four fiducials 8x	Localising four fiducials 8x
Metrics to measure	3D position of fiducials, force, time, heading angle, CSUQ, SUS, Trust, NASA TLX.	3D position of fiducial, force, time, heading angle, CSUQ, SUS, Trust, NASA TLX, modes, gains.
Survey	Six sections: description, trust CSUQ, SUS, NASA TLX and end. Shown in Appendix D.	Six sections: description, trust CSUQ, SUS, NASA TLX and end. Shown in Appendix D.

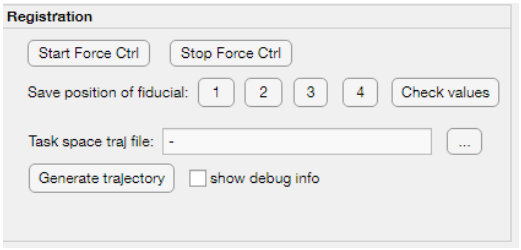


Figure 2.12: The GUI presented to the operators acting with the baseline design

2.2.3. Variables to be assessed

The variables to be assessed are summarised in Figure 2.10. Each is described in detail. Statistical tests are used to provide conclusions for the quasi-experiment. All metrics are tested for normal distribution with a Kolmogorov-Smirnov test, and assuming their non-normal distributed data, the Mann-Whitney U test was performed. This test holds the null hypothesis that the medians of two data sets are equal. A p-value less than 0.05 is considered statistically significant, so the null hypothesis should be rejected.

Accuracy

The dependent variables under assessment are the Fiducial Localisation Error (FLE) and Target Registration Error (TRE) which are explained in more detail in Chapter 3.2.

FLE

The FLE is calculated by examining the robot tool centre position (TCP) and the fiducial centre's ground truth (GT) position. The GT of the fiducial can not be determined; consequently, the average positional value of each fiducial measurement was embraced as its GT. Calculating the FLE with

$$FLE = \begin{bmatrix} FLE_x \\ FLE_y \\ FLE_z \end{bmatrix} = \begin{bmatrix} TCP_x \\ TCP_y \\ TCP_z \end{bmatrix} - \begin{bmatrix} F_{cx} \\ F_{cy} \\ F_{cz} \end{bmatrix} \tag{2.1}$$

with TCP_x, TCP_y and TCP_z the TCP measurements for the x-, y-, and z-coordinates for one specific measurement for one fiducial, and F_{cx}, F_{cy}, and F_{cz} the centroid position in x, y and z of that specific fiducial, representing the GT. Then the vector length is calculated as

$$|FLE| = \sqrt{(FLE_x)^2 + (FLE_y)^2 + (FLE_z)^2} \tag{2.2}$$

The FLE was calculated according to these formulas for each of the eight fiducial measurements, providing eight FLE values per participant per fiducial. Thus 32 FLE values per participant. Figure 2.13A depicts the measurement of FLE as the difference between the GT centroid position of the fiducial and the measurement of the TCP, and Figure 2.13B shows multiple FLE measurements surrounding one GT fiducial.

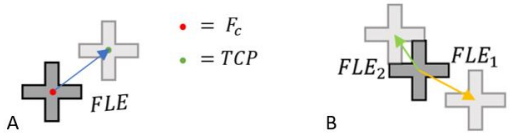


Figure 2.13: FLE metric to measure localisation performance. A depicts the measurement of FLE as the difference between the GT centroid position F_c of the fiducial and the measurement of the TCP, and B shows multiple FLE measurements surrounding one GT fiducial.

TRE

For the TRE, the difference between two targets, also called point targets (PT), in two different coordinate frames is calculated. The target point resembles the round window location in this thesis, as

this is a critical structure in CI surgery. Figure 2.14 summarises this thesis's steps to calculate the TRE. It requires three frames, with frame 0 actual CT data with fiducial placement equivalent to the fiducial placement in this experiment. Frame 1 resembles the fiducial configuration on the block measured with a specialised microscope. Frame 2 resembles the measured fiducial positions during the localisation task as their GT values for each participant.

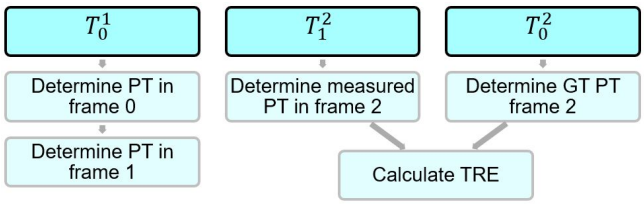


Figure 2.14: Schematic overview of the steps taken to calculate TRE. With frame 0 real CT data, fame 1 resembles the fiducial configuration measured with a specialised microscope, and frame 2 the fiducial positions during the localisation task. The transformation matrices are T_0^1 between frame 0 and 1, T_1^2 between frame 1 and 2 and T_0^2 between frame 0 and GT positions of frame 2.

First, the PT position is extracted from the actual CT data. Then using the DetermineHomogeneousTransformationMatrix script, the transformation matrix T_0^1 is determined from which the PT in frame 1 is calculated. Second, the transformation matrix T_1^2 between frame 1 and a combination of the measured positions in frame 2 is determined from which the measured PT in frame 2 is calculated. The transformation matrix T_0^2 between frame 0 and the participant's GT fiducial positions in frame 2 is calculated. This is used to calculate the GT PT position in frame 2. The GT PT position is then subtracted from the measured PT position in frame 2 to determine the TRE value. This is repeated for 8 to the power of 4 different combinations of fiducial measurements per participant. More details are supplied in Appendix B.

Human workload

The human workload is measured both quantitatively and qualitatively.

Qualitative measure: NASA TLX

The NASA TLX, NASA Task Load Index, created by Hart and Staveland (1988), is a multi-dimensional survey-based measure of workload (Grier, 2015). This survey scores workload based on six sub-scales: mental demand, physical demand, temporal demand, frustration, effort, and performance. Each is scaled from 0 to 100 by marking the desired position and weighted based on participant questions related to the perceived contribution to the workload. From each sub-scale, the mean was calculated with

$$\text{mean} = \frac{\text{Sum of all question points}}{\text{Number of question points}} \tag{2.3}$$

These mean values were weighted through multiplication with its tally and dividing by 15, the sum of all weights. The resulting weighted mean is the overall workload, ranging from 0 (low) to 100 (high). More details of the NASA TLX are given in Appendix B.

Quantitative measure: Average force

The operator effort is measured quantitatively by the average force. The higher the average force, the higher the workload. The average force was calculated from the total force over the total time:

$$WL = \frac{\sum_{t_{\text{begin}}}^{t_{\text{end}}} F_t}{t_{\text{end}} - t_{\text{end}}} \tag{2.4}$$

with WL the workload for one participant, t_{end} the end-time of the localisation task and t_{begin} the beginning time, and F_t the force magnitude measurement as $|F|$ at some point in time t. Then the WL for each group of participants is analysed and compared to the WL of the other participants.

Usability
Usability analysis can be performed in various ways, such as interviews, behaviour analysis and through questionnaires. For this research, two questionnaires were employed: the Computer System Usability Questionnaire (CSUQ) and the System Usability Scale (SUS).

CSUQ
CSUQ is a 19-item survey that measures usability specifically for human-computer interaction. It is created by J. R. Lewis (1995) and holds a 0.95 overall Cronbach's alpha score. The 19 questions were scored along a bipolar 7-point Likert scale. Afterwards, participants are requested for three negative and three positive aspects of the system. The mean scores were calculated with Equation 2.3. The complete questionnaire is demonstrated in Appendix B.

SUS
The SUS is a 10-item survey designed by Brooke et al. (1996) and shown in Appendix B. The scale ranges from 1 to 5, and these scores are normalised from 0 (poorest rating) to 4 (best rating). To calculate the participant SUS score, the odd questions are scored positively, whereas the even questions are scored negatively. These are then summed and multiplied by 2.5:

$$\begin{aligned} SUS = 2.5(20 + SUM(SUS01, SUS03, SUS05, SUS07, SUS09) \\ - \text{SUM}(SUS02, SUS04, SUS06, SUS08, SUS10)) \end{aligned} \tag{2.5}$$

The final value is a reflection of the system's usability. Its relevance is based on the distribution of all scores, similar to a grading system. A SUS score of 74 entails that the usability is 70% better than all products tested. The top 10% of the scores are above 80.3. A SUS of 68 is average and reflects a C, whereas a SUS of 51 or lower reflects the bottom 15%, an F.

Trust
To measure human trust in robots, the Trust Perception Scale-HRI was developed by Schaefer (2016). It is applicable across all robot domains. The Trust Perception Scale-HRI is a 40-item survey with a 14-item sub-scale developed based on six dimensions, as pictured in Appendix B, Figure B.12. The scale was designed as a post-interaction measure to compare changes in trust across multiple conditions. Hence, the survey should be filled in instantly after the interaction. Five items have to be reverse coded, indicated in Figure B.12 with an "a". Then the items are summed and divided by the total number of items, as Equation 2.3, acquiring an overall trust score between 0 and 100%. In this measure of trust, other factors like perceived risk and reliability are included and will not be measured individually.

2.2.4. Variables to be controlled
The variables that will be measured are dependent; thus, other variables should be controlled. These should be held as consistently as possible throughout the experiments. The variables that influence the experiment outcome are elaborated in Chapter 5.1.2. The variables are distinguished into kinematic, non-kinematic and human-related errors. It is assumed that the non-kinematic errors were the same for each participant, thus averaging out for both groups. This also applies to the higher level Human Factor Analysis and Classification System (HFAC) Cheese-model levels as described in Chapter 3.4.1.

Concerning the kinematic errors, the registration bit, fiducials, fiducial amount, fiducial placement and robot were the same throughout all the experiments. Their dimensions and details are provided in Appendix B. Fixating the block to the robot base plate excludes table fixation requirements. Since the human localisation error is a measure of a relative error with respect to some GT, the exact position change is not crucial for this experiment.

Other variables that influence the outcome are related to individual participant characteristics. Therefore, participants must provide their age group and scale their experience with robots. The groups

were diverged to resemble close equivalence.

2.2.5. Hypotheses

In answering Q2, the hypothesis for the sub-sub-questions are provided first.

For Q2.1, the hypothesis is that multi-modal multi-feedback admittance control will improve accuracy by reducing the FLE and, thus, TRE. This reduction is anticipated since the human is equipped with localisation feedback to enhance its performance. Besides, employing multiple modes offers more intuitive control of the robot manipulator. Consequently, humans are foreseen to retain more mental capacity for accurately localising the fiducials. Further, the implemented functionalities mitigate the unsafe acts of the operator to a minimum, thus reducing the number of failures that can occur. Hence the hypothesis states H_1 = **Multi-modal multi-feedback admittance control will improve the accuracy by decreasing the FLE and thus TRE.**

For Q2.2, the hypothesis is that multi-modal multi-feedback admittance control will lower the human workload since the operator will have a more smooth motion of the robot manipulator, demanding less force. Additionally, the mental capacity is anticipated to be lower due to the intuitive control in line with human localisation techniques and the elimination of inter-communication with another operator. Thus the hypothesis states H_2 = **Multi-modal multi-feedback admittance control will decrease the human workload measured by the average force and NASA TLX survey.**

For Q2.3, it is envisioned that multi-modal multi-feedback admittance control will have a higher usability score than the baseline design due to its intuitiveness. In addition, the feedback is anticipated to increase the participant's understanding of the process. Understanding is a fundamental principle for usability according to Wilkinson et al. (2021). Hence the hypothesis states H_3 = **Multi-modal multi-feedback admittance control will improve the human-robot (system) usability, measured by the CSUQ and SUS scores.**

For Q2.4, it is expected that multi-modal multi-feedback admittance control will improve human trust in the system since essential factors that influence trust are understanding, communication, and feedback (Schaefer, 2016). Thus the hypothesis is H_4 = **Multi-modal multi-feedback admittance control will improve the human trust in the system, measured by the Trust Perception Scale-HRI.**

Thus answering Q2, it is predicted, according to all four sub-sub-questions and their hypothesis, that the multi-modal multi-feedback admittance control will enhance the overall system performance as measured by accuracy, workload, usability and trust. Expressing the hypothesis H_0 = **Multi-modal multi-feedback admittance control will improve the human-robot (system) performance measured by accuracy, workload, usability and trust.**

2.2.6. Ethical considerations

This experiment involves humans, thus requesting ethical considerations. Consequently, a consent form, a data management plan and a checklist for human research were used and approved by the TU Delft ethics committee. All are provided in Appendix E.

3. Background information

This chapter describes the most important background information related to the thesis. First, information related to the registration is given followed by HRI-related challenges. Then the contextual background is provided.

Theoretical information

This part describes the theoretical knowledge underlining this thesis, including registration, its accuracy, different methods, STOTA literature, human-robot interaction and its related challenges.

3.1. Image registration

For robots to execute surgery autonomously, image registration is needed (Panara et al., 2021). Image registration is mapping the features or coordinate space of one image with the features or coordinate space of another image and is thus a spatial transform (Alam et al., 2016; Cao et al., 2016; SickKids, n.d.). Also defined as: "*the process of locating the patient with respect to the robot and correlating the patient location with the pre-operative scan used for planning the robot trajectory*", or: "*the determination of a relation between image data and the body*" (Eggers et al., 2006; Fitzpatrick and West, 2001; Greenwood and Vallee, 2021; Machetanz, Grimm, Wang, et al., 2021). In this project, the image data or pre-operative scan is acquired using a CT scan. Image registration can be referred to as patient-to-robot registration, image-to-physical registration, coordinate space registration, and patient-to-image registration. Other terms are summarised in Figure A.1. This thesis uses the term registration.

Registration is critical in robotic surgery since it significantly affects the total system accuracy. The main stages for registration are (1) Feature detection, (2) Feature extraction, (3) Transformation, and (4) Optimisation (Alam et al., 2016; J. Liu et al., 2019; Zitová and Flusser, 2003).

Feature detection and Feature extraction

The essence of feature matching is detecting distinctive objects, named fiducial markers, in both frames (Wyawahare et al., 2009). Once these features are detected, a cost function, such as the sum of squared differences, the sum of absolute differences, normalised cross-correlation, or normalised correlation coefficients, finds a similarity measure of these in both spaces (Alam et al., 2016; W. Liu et al., 2009; Oliveira and Tavares, 2014). The specific features vary per registration method.

Transformation

An optimal coordinate transform must be specified to convert the coordinates from different coordinate systems (Eggers et al., 2006; J. Liu et al., 2019; M. Shah, 2014). This transform consists of a rotation matrix (B) and a linear translation (y) (Eggers et al., 2006). To convert points Q_i from coordinate system Q to points P_i in coordinate system P gives the following equation:

$$y(Q_i) = B * Q_i + y = P_i \tag{3.1}$$

Registration is finding the most suitable transform matrix by optimising an energy function, or maximising the similarity metrics, given as:

$$\epsilon = F(I_t I_s \cdot M) + R(M) \tag{3.2}$$

with the transform matrix M, I_s the source image reflecting one coordinate frame and I_t the target image representing the other coordinate frame. F is the matching criterion, a quantification of the

alignment of both images after transformation. The term R regularises transformation (J. Liu et al., 2019). So, in mathematical terms, registration aims to determine the rotation and transformation between the image data and the patient's body with high accuracy (Eggers et al., 2006; M. Shah, 2014).

Optimization
In the optimisation step, the total degree of similarity is calculated (Alam et al., 2016).

Types of registration methods
Registration methods are classified based on different criteria. This thesis utilises the classification of Alam et al. (2016), where the primary division is in intrinsic or extrinsic registration. Extrinsic means the process depends on external objects or features affixed to the patient (2016). Intrinsic techniques use internal patient landmarks, or features (Alam and Rahman, 2016; J. Liu et al., 2019). The various registration types are referred to as registration methods, types or techniques.

Use of registration methods
Image registration can be utilised throughout the entire clinical practice (2019). It can be employed diagnostically by overlaying the patient image with images of a specific condition. It can be used during surgery by overlaying the scheduled path with real-time images of the actual path (2019). It can likewise be utilised for surgical evaluation by comparing pre-operative and post-operative patient images (2019). This thesis merely considered registration for aligning patients and image coordinate frames in the intraoperative phase.

The diverse registration methods can be conducted manually or automatically (Greenwood and Vallee, 2021). Manual registration is mainly obtained by pair-point fiducial matching and does not need additional imaging, demanding less radiation exposure to the patient and surgical team (2021). Most automatic methods use intraoperative imaging and reference arrays that can be identified by a camera (2021). In manual and automatic registration, optimisation is paramount, consisting of coarse and fine optimisation steps (J. Liu et al., 2019). Coarse optimisation constructs an initial low-accuracy estimate for the alignment. Fine optimisation searches for more accurate results within specific ranges (2019).

3.1.1. Extrinsic registration methods

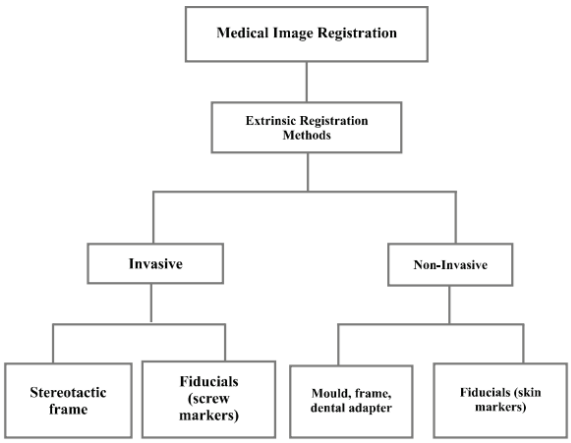


Figure 3.1: Classification of extrinsic registration methods (Alam et al., 2016). In extrinsic registration methods, the registration depends on fiducial markers connected to the patient (J. Liu et al., 2019). The attachment of these markers can be done either in an invasive way or a non-invasive way. Invasive is the necessity for incisions, and invasiveness is the scale that describes the amount, size, and the corresponding burden on the patient (Alam and Rahman, 2016). The invasive extrinsic registration methods are divided into stereotactic frames or fiducials affixed to the patient's bone as registration features. The non-invasive extrinsic registration methods are separated into mould, frame or dental adapters or fiducials affixed to the outside of the skin as features for registration

Extrinsic registration relies on distinctly visible and detectable external fiducial markers attached to the patient (J. Liu et al., 2019). These methods are relatively fast and easy, do not require complex algorithms and can predominantly be automated (Alam et al., 2016). Extrinsic methods are organised as invasive or non-invasive, as shown in Figure 3.1 (2016). Invasive is the necessity for incisions, and invasiveness is the scale that describes the amount, size, and the corresponding burden on the patient (Alam and Rahman, 2016). The extrinsic registration methods with their advantages and disadvantages are summarised in Table 3.1.

Table 3.1: Summary of extrinsic registration methods and their advantages and disadvantages. I = invasive, NI = non-invasive.

Method	Description	I/NI	Advantage	Disadvantage
Stereotactic frame registration	The frame, also called a stereotactic frame is fixed to the patients head with screws and provides a coordinate system in the physical space that is used for registration by knowing the exact position of the surgical tool with respect to the frame and the patients surgical area	I	• High-accuracy Mechanical • support for surgery • Head fixation	• Highly invasive • Can limit the surgical area
Fiducials (Screw markers) registration	Fiducial markers are attached to the patient in the form of artificial objects that are clearly visible on the preoperative image. The same fiducial markers in both coordinate frames are aligned, which is called pair-point registration and requires at least three positions that are linearly independent	I	• Simple No need for • computational algorithms • High-accuracy	• High-level planning is required Markers have to be • inserted with anesthesia • Additional imaging needed
Mould, Frame and Dental Adapter registration	Mould, head-holder frame and dental adapter are objects that are tightly attached to the patient and contain landmarks that are used for registration in a pair-point wise registration	NI	• Non-invasive • Fast and simple No additional • optimisation or computation	• Low repeatability • Low accuracy
Fiducials (Skin markers) registration	Fiducial markers are attached to the patient skin that are clearly visible on the preoperative image. The same fiducial markers in both coordinate frames are aligned, which is called pair-point registration and requires at least three positions that are linearly independent	NI	• Non-invasive • Fast and simple No additional • optimisation or computation	• Skin movement • Low accuracy • Additional costs and effort

Invasive image registration

Invasive registration uses invasively placed markers, thus requiring an incision, for example, screwing markers on the patient's head. This robust registration type provides practical and specific information about the patient's anatomy. It can be performed in either frame-based or frame-less conditions, separating it into the stereotactic frame or fiducial methods (Cardinale et al., 2017).

Stereotactic frame registration

The term stereotactic is a technique directing the tip of an instrument employing coordinates from medical imaging so that a precise location can be reached. The stereotactic system consists of a reference frame, a technique for image acquisition, and a mechanism for the proper direction of surgical devices (Alam et al., 2016). The frame, called a stereotactic frame, is shown in Figure 3.2. It is fixed to the patient's head with screws and provides a coordinate system in the physical space by knowing the exact position of the surgical tool with respect to the frame and the patient's surgical area (2016; Eggers et al., 2006). The frame has shortcomings, one being its invasiveness. Besides, it has to be mounted before surgery and can not be removed until after surgery, presenting a burden to the patient (Alam et al., 2016; Eggers et al., 2006). Lastly, it can restrict the surgeon's workspace for particular brain regions (2006). The advantages are its accuracy and the fact that it can give mechanical support for IGS to select certain positions along paths (2006).

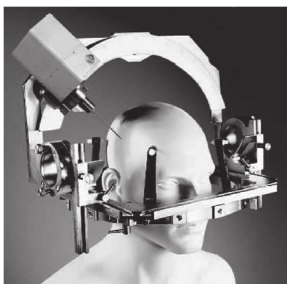


Figure 3.2: An example of a stereotactic head frame fixed to the patient's head with screws. It provides a coordinate system in the physical space used for registration (Alam et al., 2016; Eggers et al., 2006; Zanotto et al., 2011). The position of the surgical equipment is known exactly with respect to the head frame.

Fiducials (Screw markers) Image registration
In this method, named Bone Fiducial Registration (BFR), fiducial markers are attached to the patient. These markers are salient artificial objects of various shapes and materials, as shown in Figure 3.3 (Alam et al., 2016; Eggers et al., 2006). Registration is established by aligning the same fiducial markers in both coordinate frames, called pair-point registration, which requires at least three linearly independent markers to compute the transformation matrix (Alam et al., 2016).

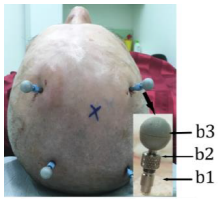


Figure 3.3: An example of invasive fiducial markers for registration (2018). Four fiducial markers are screwed inside the patient's bone. These markers consist of three parts: b1) the screw element that is self-drilling and self-tapping. b2) A base that can be removed to equip more patient comfort between surgery and pre-operative imaging. b3) A reflective sphere is used as the fiducial marker in the images.

This type of registration is straightforward and does not need complex computational algorithms. It is the most accustomed method due to its proven accuracy (Eggers et al., 2006). The downside is its invasiveness, additional imaging, hence additional costs and radiation, and the necessity of high-skill planning to position the fiducial screws and adjust the workflow (Alam et al., 2016; Eggers et al., 2006). The markers must be inserted with local anaesthesia, which can lead to complications (2006). The fiducial positions should remain invariant between pre-operative imaging and intraoperative registration (2006). Consequently, the region of the skin must be selected with care since it can lead to swelling and shifting (2006). Furthermore, the markers should be placed near the surgical area to prevent undesirable irradiation while avoiding restrictions to the surgeon's workspace (2006).

In contrast to a frame-based method, the geometry of the markers consistently differs, demanding marker localisation. This can be done in real-time with a tracker or once before the surgery (2006). The manual localisation of the fiducials in both the image space and physical space is time-consuming and error-prone, thus introducing automated methods (2006). These depend on the automatic identification of markers in the image coordinate frame and can be more exact than the manual process and more robust to lower image resolutions (2006). Localising in the physical space remains mostly a manual process (2006).

Non-invasive image registration
The less invasive a surgery, the faster a patient's recovery time (Alam et al., 2016). Non-invasive methods use additional equipment or objects as distinctive markers. They face several challenges, limiting their adoption. One of the major challenges is the provision of interactive image guidance

during surgery (2016). Non-invasive techniques are preferred over invasive ones once their accuracy is adequate due to their high efficiency and wide adoption in clinical applications (2016). These methods further diverge into Mould, Frame and Dental Adapter, and Fiducials (Skin markers) techniques. Excluding frame adapters, they utilise the pair-point-based registration principle.

Mould, Frame and Dental Adapter registration

Mould, head-holder frame and dental adapter are objects tightly secured to the patient during IGS (2016). These objects all have fiducial markers rigidly attached to them. For example, a dental adapter is held by the patient's teeth and places markers outside the mouth. These objects are secured to the patient's head prior to imaging and re-attached in the identical position before the surgery to perform registration (Eggers et al., 2006). Head-holders are similar to stereotactic frames yet non-invasive (2006). Dental adapters are connected to the upper jaw as shown in Figure 3.4 (2006). Malleable masks are fitted to the patient's face as shown in Figure 3.5 (2006).

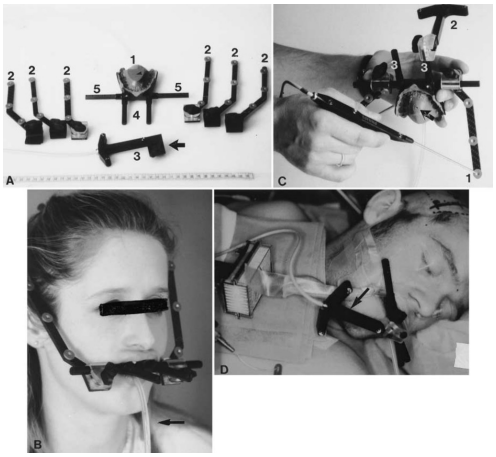


Figure 3.4: An example of non-invasive dental adapter for registration (2000). The dental adapter is positioned inside the patient's mouth. The reflective fiducial markers are attached to this mouthpiece and established outside the mouth, next to the patient's head, so they are visible for registration. A) Shows the different pieces of the dental adapter. B) Reveals the dental adapter with the mouthpiece inside the patient's mouth and the markers visible exterior. C) Shows the reflective markers used for registration. D) Displays a patient in the OR using the dental adapter(2000).

This method is efficient since it requires no complex optimisation or computation (Alam et al., 2016). Regardless, patient movement or dislocation of markers has a significant consequence on the registration accuracy and, thus, the surgical accuracy (2016). Section 3.1.3 elaborates on registration accuracy.

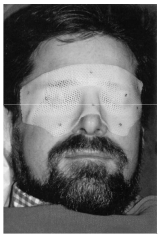


Figure 3.5: An example of non-invasive malleable mask for registration (2001). Before the registration procedure, the mask is positioned on the patient, and the fiducial markers contact the patient's skin. The mask is cleared before the surgery (2001).

Fiducials (Skin markers) registration

This registration type places artificial external markers on the patient's skin, as depicted in Figure 3.6 (Alam et al., 2016). This method works similarly to invasive markers, except being non-invasive. Their position can move slightly with respect to the skin, giving an accuracy of approximately 2mm

(Eggers et al., 2006). Using invasive or non-invasive markers comes with additional costs and effort. Therefore, markerless techniques are likely needed for more widespread use of IGS (2006).

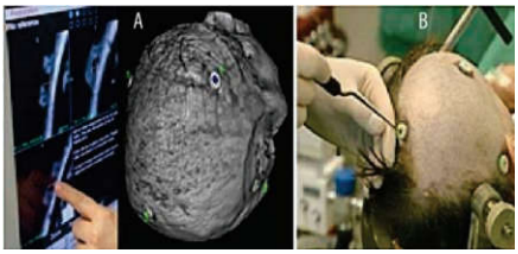


Figure 3.6: The use of non-invasive artificial external markers for IGS (2016). Markers are attached to the patient skin. The left side of the Figure illustrates the fiducials in the image coordinate frame. The right side of the Figure portrays the fiducials in the real world, which can be localised with a localisation device (2016).

Evaluation of extrinsic registration methods

The extrinsic methods, summarised in Table 3.1, were evaluated extensively by Alam et al. (2016) based on the following metrics: accuracy, efficiency, reliability, robustness, optimisation procedure, transformation, error detection, target localisation, computation, clinical use, modality and support or availability of software tools. Details about the metric definitions are expressed by Alam et al. (2016). The outcomes are shown in Figure 3.7.

Parameters	Extrinsic Registration Methods			
	Invasive Registration		Non-Invasive Registration	
	Stereotactic Frames	Fiducials (Screw Markers)	Mould, Frame, Dental Adapter	Fiducials (Skin Markers)
Accuracy	Very High, because registration cues can be taken from a device that is built expressively to provide such cue	Perform accurate registration, but requires high skill from surgeons	Less accurate then invasive registration	Less accurate due to motion of the skin during surgery
Efficiency	Less efficient in head frame placement	Efficient, because there is no need for complex optimization algorithms	Quick and fast then invasive	Quick and fast then invasive
Reliability	More reliable because neither the anatomy nor the pathology are involved in registration	Reliable but small surgical risk associated with their use	Reliable	Not reliable due to elasticity in human skin
Robustness/ Stability	Provide robust basis for registration	Mechanically stable	More stable and robust	Less robust due to the independent motion of the markers on the skin
Optimization Procedure	Not complex but inconvenient	Simple and no need for complex optimization procedures, since the registration parameters can often be computed explicitly	Simple, because registration parameters are mostly computed explicitly	Simple, because registration parameters are mostly computed explicitly
Transformation	Relationship between skull and brain remain rigid during surgery	Often restricted to rigid (Translation and Rotation)	Often restricted to rigid (Translation and Rotation)	Often restricted to rigid (Translation and Rotation)
Error Detection	Can easily detect errors introduced as a result of the mathematical operations	Can check and detect errors for co-registration	Can detect target registration errors and fiducial localization errors	Easily detect fiducial localization errors
Target Localization	Requires high spatial Accuracy	Sophisticated markers and algorithms are available which can quickly and accurately localize targets	Provide precise target localization and patient setup	subject to localization errors
Computation/Automation	Automatic and explicit	Automatic and explicit	Automatic and explicit	Automatic and explicit
Clinical use/Applications	General surgery, orthopedic, neurosurgical, craniofacial and otolaryngology	Biopsies , Catheter Insertions, Gamma Knife Surgery, Injections and Aspirations, Orthopedic and facial surgery	Biopsy, thermal ablation, endoscopy, and laparoscopy	Orthopedic and facial surgery nasal sinus surgery and neuronavigation
Modality	Well visible and accurately detectable in all of the modalities	Visible and can be detected in both mono and multimodal	Visible and can be detected in both mono and multimodal	Well visible and accurately detectable in all of the modalities
Software Tools Availability	Nero surgical planning software tools are available	Available	Available	Available

Figure 3.7: Analysis on the different types of extrinsic registration methods performed by Alam et al. (2016) based on the following metrics: accuracy, efficiency, reliability, robustness, optimisation procedure, transformation, error detection, target localisation, computation, clinical use, modality and support or availability of software tools.

3.1.2. Intrinsic registration methods

Intrinsic registration utilises points, curves, landmarks or surfaces from the patient without the addition of any external objects (Alam and Rahman, 2016). Intrinsic methods are organised into three classifications: point-based (landmark-based), surface-based, and voxel-based methods (2016). These are further organised into subcategories and shown in Figure 3.8 and summarised in Table 3.2.

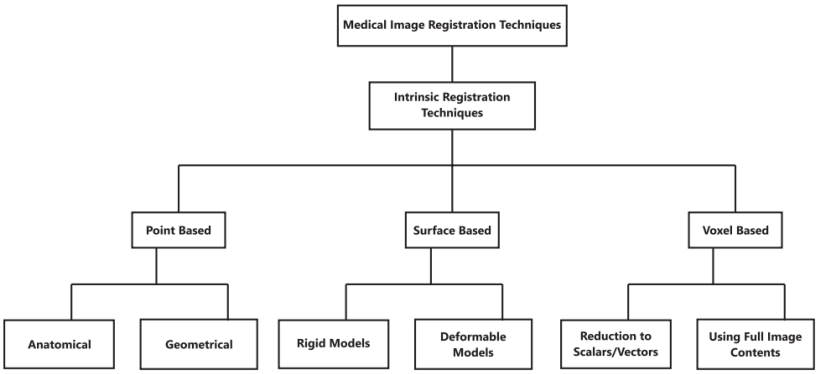


Figure 3.8: Classification of distinct intrinsic registration methods (2016). In intrinsic registration methods, the registration relies on features and anatomical information extracted from the patient. No supplementary objects are used. These methods can be further separated into point-based (landmark-based), surface-based and voxel-based methods (2016). Point-based uses specific points or landmarks inside the patient as markers with pair-point-based matching. Surface-based methods use surfaces of PCs to achieve registration. Voxel-based methods match coordinate systems based on image intensities (2016).

Table 3.2: Summary of intrinsic registration methods, their advantages and disadvantages. I = invasive, NI = non-invasive.

Method	Description	I/NI	Advantage	Disadvantage
Point or landmark based registration	Geometrical or anatomical landmarks are used that represent the same feature in different images or coordinate frames, using the pair-point principle to find a transformation matrix	NI	• Non-invasive • No additional markers • High speed and efficiency • No additional imaging	• Landmarks shift • Depends on selection and algorithm • Can be time consuming
Surface or segmentation based registration	Two large datasets of points, called Point Clouds (PCs) that describe the same surface are used and with ICP the distance between the two PCs are minimised to find a transformation matrix	NI	• Non-invasive • No additional markers • No need for anatomical data • Easy to automate	• Limited accuracy • Time consuming
Voxel-based registration	This method uses image intensities or brightness values to find the transformation matrix	NI	• Non-invasive • Fast and easy implementation	• High computational costs • Only for image-to-image registration

Point- or landmark-based registration

This registration type uses points or landmarks portraying the same feature in different coordinate frames, using the pair-point-based principle displayed in Figure 3.9 (2016). The selected points, being intrinsic, can be either anatomical or geometrical. Both equip the benefit of not needing markers and their non-invasiveness (Eggers et al., 2006). The supplementary benefit of anatomical structures is their known identification (Alam and Rahman, 2016). Yet, their accuracy is restricted as some anatomical landmarks can shift (Eggers et al., 2006).

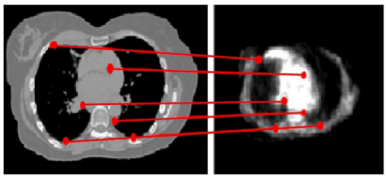


Figure 3.9: Registration based on landmarks. Small landmarks on the right side of the Figure are compared to the image coordinate frame on the left side. These landmarks are used for the pair-point-based registration to find a transformation matrix between the image and real-world coordinate frames (2016).

For geometrical points, the registration uses the measurement of key points and their position in the images (2016). Figure 3.10 offers an example. The accuracy depends on the selected points, their identification, and the image quality (2016). Geometrical point-based registration consists of four steps: identification of landmark points, development of mesh model, image transformation, and interpolation (2016).

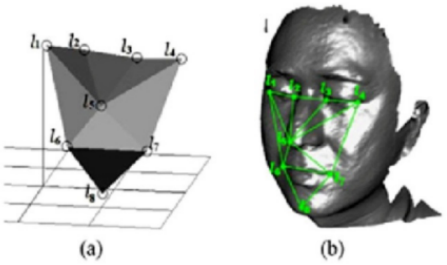


Figure 3.10: Registration based on geometrical points chosen from the image, on the left side (a) and the patient on the right side (b). This registration process consists of four steps: identification of landmark points, development of mesh model, image transformation, and interpolation (2016). The landmark points are identified as l_1 till l_8. (b) illustrates the constructed mesh model with the same landmarks. Then a transformation matrix is calculated based on this pair-point-based principle, followed by an interpolation to optimise (2016).

The selection and extraction of landmarks can be executed manually, semi-automatic or fully automatic (2016). The accuracy of manual methods is highly dependent on the experience and accuracy of the user (2016). On the other hand, fully automated methods depend on the precision and optimisation of the algorithms (2016). Overall, the landmark-based registration techniques can be performed with high speed, and efficiency (2016).

Surface- or segmentation-based registration

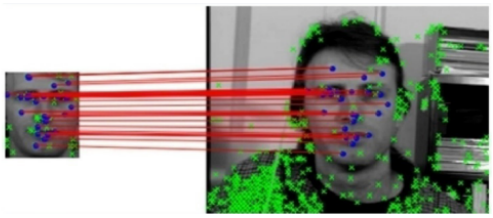


Figure 3.11: An example of surface-based registration matching two PCs. PCs are gathered with a laser surface scanner. The green crosses symbolise the detected feature points. The most critical area is selected in both images to match with ICP. The blue dots represent the feature points that are being matched. Then a transformation matrix is calculated based on those PCs. No points in the PCs will contain identical positions (2016).

This method uses two large datasets of points, called Point Clouds (PCs), representing the same surface. These datasets possess no points that exactly match in both coordinate frames (Eggers et al., 2006). Rather it views the surface as a whole, and a surface-matching algorithm finds the finest match between both PCs. The best-known algorithm, Iterative Closest Point (ICP), calculates the distance

between all points from both PCs and minimises the square errors between them (MRPT, 2022). It uses gradient descent to find the minimum error between the two surfaces in an iterative way which relates to an optimum that aligns both surfaces as in Figure 3.11 (Eggers et al., 2006; MRPT, 2022). Predominantly a surface scanner gathers the PCs, called laser surface registration (LSR). Other methods include tracked pointers or other trackers. The number of collected points varies between 40 and 200 for trackers, and 100,000 for laser surface scanners (Eggers et al., 2006).

This registration type is classified in either rigid or deformable models (Alam and Rahman, 2016). Rigid models preserve the shape, distances and angles among points of the images in transformation (2016). Deformable or non-rigid model registration, as shown in Figure 3.12, does not maintain the distances and angles among points and is proven valuable in the analysis of soft tissues and deformable organs (2016; Oliveira and Tavares, 2014). Deformable registration provides a more accurate model of human anatomy than rigid registration (J. Liu et al., 2019). Nonetheless, the efficiency of deformable methods is lower, and their computational costs are higher compared to rigid methods. As deformable models is not required for CI surgery, this will not be considered in this thesis.

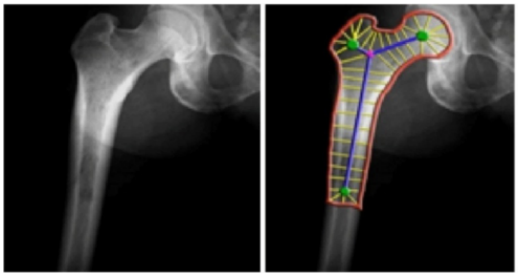


Figure 3.12: An example of deformable model registration. The left side shows the image data before registration. The right side shows the registered image, where contours are applied to see the exact and relative positioning of the data (2016). Deformable registration can account for local anatomical discrepancies and find the optimal transformation matrix between two coordinate frames.

This registration method is particularly successful when images hold low information about human anatomy (2016). Another advantage is that no auxiliary marker is needed (Eggers et al., 2006). However, the accuracy is limited, and data collection can be time-consuming (2006).

Voxel-based registration

This method operates image intensities or brightness values to find the transformation matrix, as shown in Figure 3.13 (Alam and Rahman, 2016). This is likewise called intensity-based registration and is organised into two types: reductive or complete image content-based registration (2016). Yet, voxel-based registration is exclusively used for image-to-image registration and will not be assessed further in this thesis.

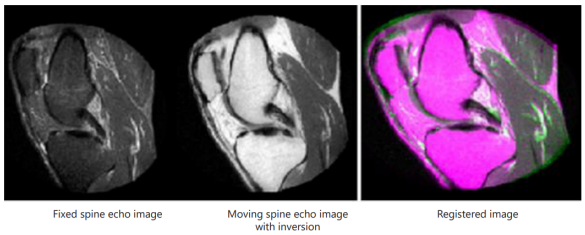


Figure 3.13: An example of voxel-based registration. The left side reveals a fixed image, whereas the middle shows a moving image of the same structure. The right side shows the registered image obtained through voxel-based registration. Hence matching voxels with the same intensity values to find a transformation matrix. Grey areas in the right Figure indicate the voxels with the same intensity values in both images. The difference in intensity values between both images is indicated with magenta and green areas (2016).

Evaluation of intrinsic registration methods

The diverse intrinsic registration techniques, summarised in Table 3.2, were studied based on the metrics stated by Alam and Rahman (2016) resulting in the table shown in Figure 3.14 and 3.15. Landmark-based methods are superior to other methods since they are less complex and bear low costs, high reliability and validation, good optimisation, easy error detection, and a wide application with multiple software tools available for support (2016).

Parameters	Intrinsic Registration Techniques					
	Landmark based (point based) registration		Segmentation based/surface based registration		Intensity (voxel) based registration	
	Anatomical	Geometrical	Rigid models	Deformable models	Reductive registration	Using full image contents
Accuracy	Accuracy depends on precise correspondence between landmarks	Accuracy relies on the accurate localization of a sufficient number of corresponding landmarks in all modalities.	High accuracy is only possible if the pre-segmentation step is performed precisely.	Provide very high accuracy when the images are precisely pre-processed and well suited for inter-subject registration.	Low accuracy due to strong dependence on the intensity variations	Low accuracy due to strong dependence on the intensity variations
Efficiency	Require more processing time due to the physical contact with multiple points and the creation of large number of exposure by the surgeons	Require more processing time due to the physical contact with multiple points and the creation of large number of exposure by the surgeons	Show high efficiency when registered structures are clearly visible and easy to segment.	Less efficient due to internal energy constraints	High, because it can directly operate on image gray values and does not require preprocessing and user interaction	High, , because it can directly operate on image gray values and does not require preprocessing and user interaction
Reliability	Excellent reliability and validity	Excellent reliability and validity	Precise pre-segmentation step is essential for reliable registration.	More reliable then rigid models	Robust and versatile	Robust and versatile
Robustness/ stability	Main advantage of using landmarks is robustness	Main advantage of using landmarks is robustness	Gives robust results	Robust and elegant	Highly stable and robust of all registration methods	Highly flexible and robust
Optimization procedure	Simple and efficient due to the availability of limited number of points in the registration process	Simple and efficient	Difficult due to several alteration in image parameters	Computationally efficient due to separate alteration in image parameters	For optimization need a priori knowledge of the nature of registered images	For optimization need a priori knowledge of the nature of registered images
Transformation	- Non rigid and low dimensional - Good feature alignment - Faster to compute the mapping transformation then surface based or intensity based registration	- Rigid and low dimensional - Good feature alignment - Faster to compute the mapping transformation then surface based or intensity based registration	Rigid transformation, with only six degrees of freedom.	Non rigid	Rigid, non-rigid and global transformations,	Rigid, non-rigid and global transformation
Error detection and calculation	Point based algorithms can easily identify and calculate errors such as Target registration error (TRE and fiducial localization error (FLE) due to the availability of well known points	Point based algorithms can easily identify and calculate errors such as Target registration error (TRE and fiducial localization error (FLE) due to the availability of well known points	Easy due to small amount of change in angle	Difficult error detection, measurements and comparisons due to high degree of freedom	Detection and computation of mean square error is difficult due to low resolution and small variations in image intensities.	Detection and computation of mean square error is difficult due to low resolution and small variations in image intensities.
Target localization	Difficult, due to improper visibility of target lesion	Difficult due to surrounding critical organs and normal tissues	Simple due to dependence on less number of points for correspondence	Easy but has some limitations in a situation such as patient motion and breathing	Difficult and subject to inaccuracies, uncertainties and measurement errors	Difficult and subject to inaccuracies and measurement errors
Computation/ automation	Intensive computation and automatic landmark selection is a challenging task	Intensive computation	-Commonly automated but presegmentation step is usually executed semiautomatically -Intensive computation	-Commonly automated but presegmentation step is usually executed semiautomatically -Intensive computation	Computationally expensive as they operate on the full image content but Implemented in an automatic fashion	Computationally intensive but Implemented in an automatic fashion

Figure 3.14: Analysis on the different types of intrinsic registration methods performed by Alam and Rahman (2016) based on the following metrics: accuracy, efficiency, reliability, robustness, optimisation procedure, transformation, error detection, target localisation and computation (2016).

Parameters	Intrinsic Registration Techniques					
	Landmark based (point based) registration		Segmentation based/surface based registration		Intensity (voxel) based registration	
	Anatomical	Geometrical	Rigid models	Deformable models	Reductive registration	Using full image contents
Clinical use/ applications	Bimaxillary surgery	Used for brain, breast, chest, liver, kidneys diagnoses	Lung cancer radiotherapy	-Track them on-rigid motion of the heart, the growing tip of neuritis, motion of erythrocytes. -Used to locate structures in the brain, register images of the retina, vertebra and neuronal tissue	Mainly used in the re-alignment of scintigraphic cardiac studies	Spline surgery, coronary angiography, morphometry
Modality	CT, MR, PET, CT-MR	CT, MR, PET, CT-MR	CT,MR, PET, SPECT	X-Rays, CT, MR and ultrasound	CT, MR, PET	Mostly restricted to mono-modality
Software tools availability	Photoshop and Matlab landmark/fiducial based registration toolbox	Photoshop and Matlab landmark/fiducial based Registration toolbox	Drop, VRMesh software	Virtual Grid, Bellevue city, WA	Virtual Grid, Bellevue city, WA	Maxilim software
Drawbacks	- Difficult registration of point matching constraints - Produce erroneous results due to the fewer definable points. - Local discontinuities - Require automatic and high-accuracy matching of a large number of landmarks	- Time consuming identification of landmarks. - Less robust under severe geometric distortions or Incomplete matching. - Local discontinuities - Landmark extraction is always prone to error.	- Number of 2D views, angle between views, view angle relative to anatomical objects, co-registration error between views, noise in the images, and image distortion. -Registration accuracy is limited to the accuracy of the segmentation step.	-Very sensitive to parameters -poor convergence to boundary concavities -Unable to capture object curves in some medical image segmentation - Infinite number of possible non-rigid transformations of the templates - Cannot handle topological changes during model generation or evolution	- Sensitive to missing image data - Limited to image to image registration - Often susceptible to any physical differences in the images being matched	- Cannot cope with large geometric deformations -Very limited use in time constrained applications such as intra-operative 2D and 3D registration

Figure 3.15: Analysis on the different types of intrinsic registration methods performed by Alam and Rahman (2016) based on the following metrics: clinical use, modality and support and availability of software tools (2016).

3.2. Registration accuracy

Registration finds a transformation matrix that arranges two different coordinate systems to match as closely as possible (Alam et al., 2016; Eggers et al., 2006; Fitzpatrick and West, 2001; Greenwood and Vallee, 2021; Machetanz, Grimm, Wang, et al., 2021). What this matching entails depends on the chosen registration method. For surface-based registration, this means matching PCs, whereas pair-point registration entails matching points. The registration accuracy measures how well these points or surfaces are matched. The registration accuracy is measured in different metrics, such as Localisation Error (LE), Fiducial Registration Error (FRE) and Target Registration Error (TRE). These metrics and their meaning differ slightly per registration method. LE is related to the error in locating the different landmarks (Fitzpatrick and West, 2001). This translates to a Fiducial Localisation Error (FLE) for pair-point-based registration or a Surface Localisation Error (SLE) for surface-based registration.

This section focuses primarily on pair-point-based registration errors to provide an example of registration accuracy. For pair-point-based registration, registration is finding a transformation matrix that positions n number of pair points in two different coordinate systems so that they match as closely as possible (Alam et al., 2016; Eggers et al., 2006; Fitzpatrick and West, 2001; Greenwood and Vallee, 2021; Machetanz, Grimm, Wang, et al., 2021). The registration accuracy for pair-point-based registration is described in FLE, FRE and TRE. The calculations in this section are primarily based on the paper of Shamir et al. (2011).

3.2.1. Fiducial Localisation Error

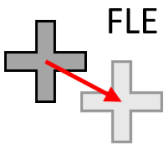


Figure 3.16: Illustration of the FLE vector for one fiducial marker. Dark grey illustrates the actual fiducial position, whereas light grey represents the localised position. The Euclidean distance between those two positions is the FLE vector, represented by the red vector.

The FLE, also named localisation error or localisation registration error, is the error that occurs in localising a fiducial. FLE is the difference, or vector, between the actual and localised fiducial position, shown in Figure 3.16. The localised position is also referred to as the measured position. The vector size relates to the FLE magnitude (2011). FLE can be estimated by calculating the average vector between repeated localisation positions and their centroid of one fiducial point (W. Liu et al., 2009).

The FLE distribution can be either isotropic/anisotropic, homogeneous/heterogeneous and unbiased/biased (Shamir et al., 2011). Isotropic distribution holds an equal FLE in all directions, whereas anisotropic distribution contains a variable FLE in different directions (2011). Homogeneous distribution relates to multiple fiducials where the FLE is distributed similarly for each fiducial, whereas heterogeneous distribution allows each fiducial to differ in its FLE value. Lastly, the FLE distribution can have an expected value of zero (unbiased) or not zero (biased).

Matching four fiducial pair points in two coordinate frames, A and B, delivers their actual coordinates as $A = \{a_1, a_2, a_3, a_4\}$ and $B = \{b_1, b_2, b_3, b_4\}$, assuming A is the real-world coordinate frame and B the imaging coordinate frame. Defining their measured positions as

$$P(A, L_A) = \{p_i | p_i = a_i + \vec{l_{a_i}}\}_i, \tag{3.3}$$

and

$$Q(B, L_B) = \{q_i | q_i = b_i + \vec{l_{b_i}}\}_i. \tag{3.4}$$

with $L_A = \{\vec{l_{a1}}, \vec{l_{a2}}, \vec{l_{a3}}, \vec{l_{a4}}\}$ and $L_B = \{\vec{l_{b1}}, \vec{l_{b2}}, \vec{l_{b3}}, \vec{l_{b4}}\}$ the set of FLE vecors for each fiducial. Thus P and Q are the set of measured points in either the real-world or imaging coordinate frame shown in Figure 3.17 in light red and light orange for coordinate frame A and B, respectively. Figure 3.17 shows the actual fiducial locations in grey and the FLE vectors per fiducial. The FLE occurs in both coordinate frames, A and B. The transformation matrix is calculated based on the localised fiducial locations by finding the minimised root mean square distance or error (RMSE) between the localised points in both frames (Robillard, 2019; Shamir et al., 2011). Thus registration tries to overlap the measured fiducial positions in the world frame $P(A, L_A)$ with the measured fiducial positions in the image frame $Q(B, L_B)$ by rotating and translating their positions. It is called rigid transformation if the shape and size of the fiducial distances remain the same in both coordinate frames.

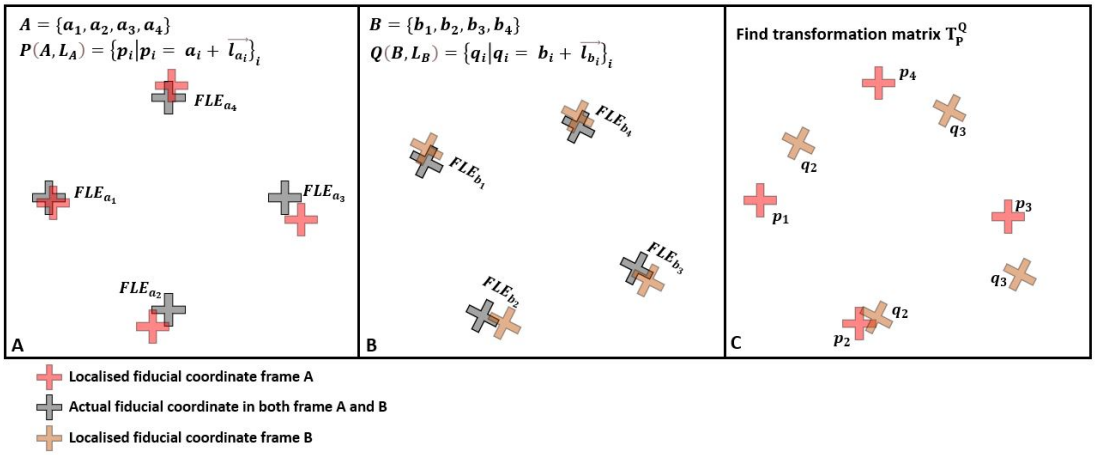


Figure 3.17: Example of FLE vectors in two different coordinate frames, A and B, and how they influence the registration. **A)** shows the real-world coordinate frame with the fiducials shown in dark grey and their measured fiducial points, shown in light red, with the FLE vectors for each fiducial as the distance between the measured and actual fiducial positions in the real-world coordinate frame. **B)** shows the imaging coordinate frame with the fiducials, shown in dark grey and their measured fiducial points in light orange, with the FLE vectors for each fiducial in the imaging coordinate frame. **C)** shows the measured fiducial positions from both coordinate frames, which require re-arranging to find the transformation matrix T_P^Q that minimises the RMSE between the four measured fiducials in each frame.

T_P^Q is the rigid transformation that minimises the RMSE between corresponding measured fiducial point pairs in the different coordinate systems $\{(p_i, q_i)\}_i$, following:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(p_i - q_i)^2}{n}}. \tag{3.5}$$

Figure 3.18 shows the RMSE minimisation between the localised fiducials in two coordinate frames, where the blue vectors indicate the error between fiducial locations related to $p_i - q_i$ in Equation 3.5. The minimal RMSE provides a transformation matrix T_P^Q, consisting of a rotational and a transnational part in the form of

$$\begin{aligned} H &= \begin{bmatrix} R & d \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} n_x & s_x & a_x & d_x \\ n_y & s_y & a_y & d_y \\ n_z & s_z & a_z & d_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned} \tag{3.6}$$

with H the homogeneous transformation matrix, R the rotational part and d the translation. More details on the algorithm that finds this transformation matrix is given in Appendix H. After registration, two other errors occur, the FRE and TRE.

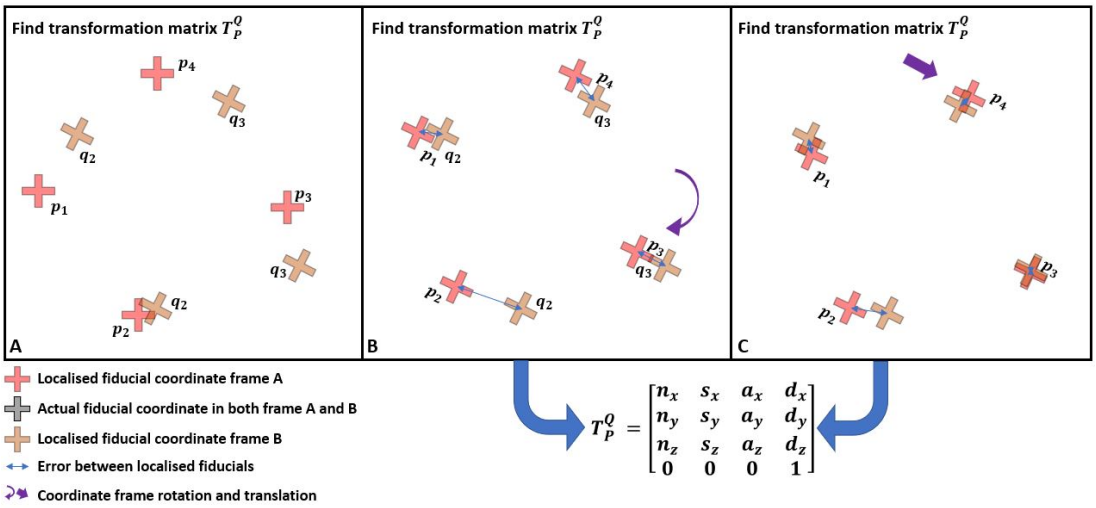


Figure 3.18: An example of finding the transformation matrix T_P^Q between two coordinate frames that minimises the RMSE **A)** shows the localised fiducial locations, $P(A, L_A) = \{p_i | p_i = a_i + \vec{l_{a_i}}\}_i$ and $Q(B, L_B) = \{q_i | q_i = b_i + \vec{l_{b_i}}\}_i$, from both coordinate frames. **B)** rotates coordinate frame A to decrease the error difference between the localised fiducials indicated by blue vectors. **C)** translates the already rotated coordinate frame A to minimise the error difference indicated by the blue vectors. The rotation and translation contribute to the Transformation matrix T_P^Q.

3.2.2. Fiducial Registration Error

The FRE is the root mean square (RMS) distance between the location of the fiducial points on images and their physical location after registration (2011). Figure 3.19 explains the FRE. Suppose the discovered transformation matrix between two coordinate systems is re-applied to one of these coordinate systems. In that case, it should yield the exact same coordinates for the fiducials as localised in the other coordinate frame. Thus if the transformation matrix is applied to the real-world frame fiducial locations, $P(A, L_A)$, it should provide the exact localised coordinates in the image frame $Q(B, L_B)$.

The discrepancy between this expected value and the yielded value is the FRE calculated as

$$F\vec{R}E_i = T_P^Q p_i - q_i, \tag{3.7}$$

Normalising the FRE vector provides the FRE value of fiducial pair point i as

$$FRE_i = |F\vec{R}E_i|. \tag{3.8}$$

Total FRE is calculated from the RMS of the n separate FRE vectors, in this case, four, as

$$FRE(P, Q, T_P^Q) = \sqrt{\sum_{i=1}^{n} \frac{FRE_i^2}{n}}. \tag{3.9}$$

FRE can also be defined as the mean, variance or 95th percentile of all FRE vectors, but this thesis defines it as the RMS. The FRE or another variant of the FRE is mainly displayed on IGS systems to indicate how well the transformation fit was ("Image-Guided Technique in Neurotology | Ento Key", n.d.; Robillard, 2019). Nevertheless, a misconception is that low FRE means high accuracy registration. FRE does not relate to the TRE, which is the most critical metric for registration accuracy; FRE increases with the number of matching points, whereas TRE decreases with the number of matching points ("Image-Guided Technique in Neurotology | Ento Key", n.d.).

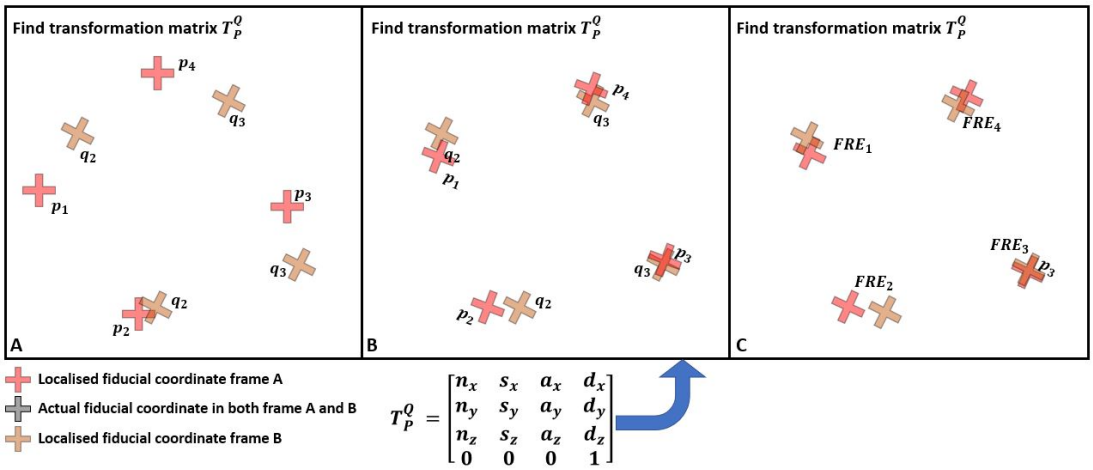


Figure 3.19: Explanation of FRE value in pair-point-based registration. **A)** shows the localised fiducial positions $P(A, L_A) = \{p_i | p_i = a_i + \vec{l_{a_i}}\}_i$ and $Q(B, L_B) = \{q_i | q_i = b_i + \vec{l_{b_i}}\}_i$from coordinate frames A and B in light red and light orange, respectively. **B)** shows the resulting fiducials positions of coordinate frame A after applying the known transformation matrix T_P^Q. **C)** shows the FRE values measured by $F\vec{R}E_i = T_P^Q p_i - q_i$

3.2.3. Target Registration Error

The TRE is the error in a specific target point, which is not one of the fiducial points, after registration (Cohen et al., 2015; W. Liu et al., 2009; Robillard, 2019). It is the most clinically significant metric (Gerber et al., 2013). The TRE cannot be measured since the target point lies inside the anatomy, consequently estimating it from other measurements. TRE correlates to the FLE according to Fitzpatrick (2009), Gerber et al. (2013), and Oliveira and Tavares (2014):

$$\left\langle |\,\mathrm{TRE}(x)|^2 \right\rangle = \frac{1}{N}\left(1 + \frac{1}{3}\sum_{k=1}^{3}\frac{d_k^2}{f_i^2}\right)\left\langle \mathrm{FLE}^2 \right\rangle \tag{3.10}$$

providing the TRE at a point (x). With d_k the distance of x from principal axis k of the fiducial configuration, f_k^2 the mean of squared distances of the fiducials from that axis and N the number of

fiducial points (Fitzpatrick, 2009).

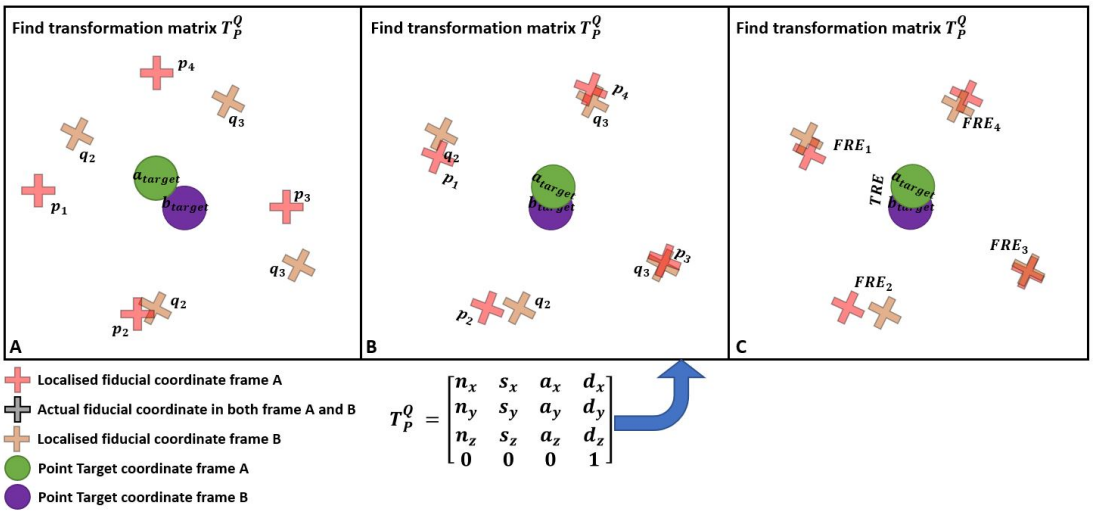


Figure 3.20: Explanation of TRE value in pair-point-based registration. **A)** shows the localised fiducial positions $P(A, L_A) = \{p_i | p_i = a_i + \vec{l_{a_i}}\}_i$ and $Q(B, L_B) = \{q_i | q_i = b_i + \vec{l_{b_i}}\}_i$ in light red and light orange and PT locations a_{target} and b_{target} in green and purple from for coordinate frames A and B, respectively. **B)** shows the resulting fiducials positions and target position of coordinate frame A after applying the known transformation matrix T_P^Q. **C)** shows the FRE and TRE values measured by $\vec{FRE_i} = T_P^Q p_i - q_i$ and $\vec{TRE_{target}} = T_P^Q a_{target} - b_{target}$

The target point or point target (PT) is located at a_{target} and b_{target} for coordinate frames A and B, respectively. Any PT can be used as long as it is similar in both frames and will yield different results. This thesis uses the round window as the PT due to its clinical relevance in CI surgery. Like FRE, the TRE is estimated by re-applicating the transformation matrix to either coordinate frame and summarised in Figure 3.20. As the PT can not be measured in the real-world coordinate frame, it has to be estimated. The TRE vector is the vector or euclidean distance between the PT location of both coordinate frames after the transformation (Cohen et al., 2015):

$$\vec{TRE_{target}} = T_P^Q b_{target} - a_{target}. \tag{3.11}$$

Normalising the vector provides the scalar error value

$$TRE(T_P^Q, b_{target}, a_{target}) = |\vec{TRE_{target}}|. \tag{3.12}$$

Providing TRE as one number limits its value as a statistical distribution. An average TRE of 1.0mm deviates for different occasions (“Image-Guided Technique in Neurotology | Ento Key”, n.d.). Mathematically calculating the TRE can be done in numerous ways, one shown by Fitzpatrick and West (2001), which measures the maximum TRE every 20 or so measurements. Besides, the TRE changes according to the PT position relative to the fiducial markers (Eggers et al., 2006).

3.3. State-of-the-art registration

Figure 3.21 provides a summary of the found STOTA registration methods given in Appendix F. This thesis focuses solely on the real-world coordinate frame, disregarding the image-based localisation in the STOTA papers. No methodological literature search was performed. Thus this information might lack some STOTA papers. Appendix A describes the steps for an elaborated literature search.

The paper from Gerber et al. (2013) achieves the lowest TRE of 0.101 ± 0.040mm, utilising a semiautomatic fiducial detection technique for pair-point-based bone-anchored fiducial registration. Multiple papers show a relation between the distance-to-centroid (DTC) and the TRE for external fiducials, in-

trinsic markers and surface points. Hence, the points used for registration should be placed so that their centroid is close to the PT. Another relationship is between the TRE and the number of markers, where the TRE decreases with increasing markers.

Most extrinsic registration methods rely on BFR, as it is the most accurate one (Su et al., 2022). However, J. Wang et al. (2020) mentions that BFR increases the patient's surgical trauma and the likelihood of infection, which is not beneficial to postoperative rehabilitation. Non-invasive methods try to counteract these disadvantages but yield low accuracy. The dental split acquires the highest accuracy, but markers deviate with an average displacement of 0.03mm and 0.07 degrees (2020). The inaccuracy of skin markers comes from the marker's size, skin mobility, and markers falling off or slipping over time (Güler et al., 2013; Mongen and Willems, 2019; Omara et al., 2014).

Most intrinsic registration methods rely on LSR as they are fast, easy to use, require no additional planning, and can be automated easily. LSR generate accuracies around 1.0 - 2.5mm. However, the LSR method accuracy highly depends on the available imaging data and can lead to an increase in intraoperative preparation time (Machetanz, Grimm, Wang, et al., 2021). Besides, clinical studies with LSR show additional problems related to the non-rigidity of surfaces, such as facial expressions (Omara et al., 2014). The anatomical point-based registration methods range more extensive as their results depend significantly on the chosen number of landmarks. Güler et al. (2013) is the sole paper mentioning the human influence on the FLE and states a higher human localisation error in anatomical markers compared to fiducial screws (2013). No geometrical point-based registration methods were found.

Extrinsic

Invasive	Metric	Accuracy value	Literature
Stereotactic frame	TRE:	1.5 ± 0.6 mm 1.6 ± 0.8 mm	Machetanz et al. (2021)
	TRE:	0.76 ± 0.34 mm	Cardinale et al. (2017)
Fiducial screw markers	FLE:	0.005 ± 0.004 mm excl. tracking error 0.046 ± 0.029 mm incl. tracking error	Gerber et al. (2013)
	TRE:	0.101 ± 0.040 mm incl. tracking error	
	TRE:	0.94 ± 0.06 mm	Soteriou et al. (2016)
	TRE:	0.7 ± 0.5 mm 1.5 ± 0.8 mm	Machetanz et al. (2021)
	TRE:	0.7 ± 0.3 mm periorbital 0.8 ± 0.3 mm viscerocranium 1.1 ± 0.4 mm neurocranium	Luebbers et al. (2008)

Non-invasive	Metric	Accuracy value	Literature
Mould, frame, dental adapter	Dental split TRE:	1.1 ± 0.3 mm periorbital 1.3 ± 0.4 mm viscerocranium 2.3 ± 0.5 mm neurocranium	Luebbers et al. (2008)
	Dental split TRE:	0.471 ± 0.276 mm entry point 0.671 ± 0.268 mm cochlear target point	Wang et al. (2020)
	Dental split TRE:	0.73 ± 0.25 mm	Labadie et al. (2004)
	Dental split TRE:	0.55 ± 0.28 mm	Ledderose et al. (2012)
	Head band TRE:	1.46 ± 0.15 mm	Wellborn et al. (2017)
	Granular cap TRE:	0.56 - 1.40 mm	Wellborn et al. (2017)
	LED mask TRE:	0.92 ± 0.13 mm	Grauvogel et al., (2017)
Fiducial skin markers	TRE:	2.49 ± 1.07 mm	Woerdeman et al. (2007)
	TRE:	2.49 ± 0.86 mm seven markers	Mongen and Willems (2019)

Other	Metric	Accuracy value	Literature
Automatic InstaTrak	TRE:	1.41 ± 0.04 mm	Soteriou et al. (2016)
Ultrasound	TRE:	0.3 - 5.1 mm	Varma and Eldridge (2006)
Dental split and	No results	No results	Opdenakker et al. (2017)
Neuroclate	TRE:	0.67 ± 0.29 mm	Cardinale et al. (2017)
STAMP method	TRE:	1.2 ± 0.12 mm	Oka et al. (2014)
Dental split with two fiducial screws	TRE:	0.6 ± 0.2 mm periorbital 0.8 ± 0.3 mm viscerocranium 1.2 ± 0.5 mm neurocranium	Luebbers et al. (2008)

Intrinsic

Point-based	Metric	Accuracy value	Literature
Anatomical	FLE: TRE:	1.6 - 3.0mm nine landmarks 4.1 ± 1.6mm nine landmarks	Shamir et al. (2009)
	TRE:	5.15 ± 0.66 mm four landmarks 4.37 ± 0.73 mm five landmarks	Soteriou et al. (2016)
	TRE:	0.51 ± 0.28 mm eight landmarks EAC and ME 5.4mm four landmarks MAS	Schneider et al. (2018)
	FLE: TRE:	1.6 - 5.5mm five - eight landmarks 3.5 ± 0.17mm five-eight landmarks	Omara et al. (2014)
	FLE:	0.61mm plastic skull 0.9mm anatomic specimen	Güler et al. (2013)
	TRE:	0.93 ± 0.31mm six landmarks	Sun et al. (2013)
Geometrical		No papers were found	

Surface-based	Metric	Accuracy value	Literature
Rigid model	TRE:	1.59 ± 0.14 mm	Soteriou et al. (2016)
	TRE:	1.0 ± 0.4 mm periorbital 1.2 ± 0.4mm viscerocranium 1.1 ± 0.4mm neurocranium	Luebbers et al. (2008)
	TRE:	1.2 ± 0.3mm	Fan et al. (2020)
	TRE:	3.0 mm	Su et al. (2022)
	TRE:	0.36 ± 0.13 mm 150 points MAS, EAC, ME 2.3 mm 50 points EAC	Schneider et al. (2018)
	TRE:	2.3 mm for 40 and 60 point PCs	Cao et al. (2016)
	TRE:	1.25 ± 0.64 mm	Ke et al., (2016)
	TRE:	5.35 ± 1.64 mm	Mongen and Willems (2019)
	TRE:	1.25 ± 0.64 mm	Ballesteros-Zebadúa et al. (2016)
	TRE:	1.7 ± 0.9 mm	Machetanz et al. (2021)
	TRE:	1.44 ± 0.24 mm	Marmulla et al. (2005)

Other	Metric	Accuracy value	Literature
physics-based shape matching	FLE	0.43, 0.32, and 0.46 mm	Suwelack et al. (2014)
Point-to-plane	Improv	0.05 to 0.2mm	Park and Subbarao (2003)

Figure 3.21: The found STOTA literature is separated per registration method, and their accuracy is given. Colour coding provides an intuitive overview of the literature that could suffice sub-millimetric accuracy, with red having accuracies larger than 1.2mm, orange being around 1.0mm and green being lower than 0.9mm.

3.3.1. Conclusion

In conclusion, registration can be divided into extrinsic or intrinsic registration based on the different features used for registration. The extrinsic methods can be further divided into invasive and non-invasive methods, whereas intrinsic methods are further divided into point-based, surface-based and voxel-based. The latter is solely used for image-to-image registration thus excluded for this thesis. The invasive extrinsic registration methods acquire high-accuracy but are invasive and require more planning. The non-invasive extrinsic methods are fast, simple and non-invasive but yield low accuracy and repeatability. The point-based intrinsic registration is non-invasive, highly efficient and does not need additional markers but yields low accuracy and has its results highly depending on specific selection. Surface-based registration is non-invasive, fast, east to automate but yields limited accuracy and is time consuming. The gold standard in registration accuracy is fiducial marker registration using bone markers (Eggers et al., 2006), this also follows from Figure 3.21.

3.4. Human-robot interaction

Human-robot interaction (HRI) is a research field considered with the understanding, designing and evaluation of robotic systems that interact with humans in performing specific tasks (R. R. Murphy et al., 2010; Sheridan, 2016). In HRI, the term operator refers to humans that perform tasks jointly with a robot. This interaction can occur in different ways, from robots that assist in carrying heavy weights to robots that answer questions, operating individually or in collaboration. The robot can have different levels of autonomy (LORA). This thesis focuses on physical HRI in the medical field.

3.4.1. HRI challenges

In HRI, challenges arise with human complexity, and the number of possible interactions (Vasconez et al., 2019). A subset of these challenges is (1) Acceptation and adaptation, (2) Safety, and (3) Human-factors. This thesis includes both robot-specific and general automation challenges, and these terms are used interchangeably.

Acceptation and adaptation

Radical technologies such as robots are less easily accepted compared to incremental innovations (Beer et al., 2014). Accepting and adapting robots into society is complex and requires additional research. Beer et al. (2011) states that factors such as appearance, structure, autonomy, control, and interfacing impact the acceptance and adaptation of robots. Parasuraman and Riley (1997) describes elements that influence the adaptation and acceptance of automation, as shown in Figure 3.22, such as a human's attitude towards automation, the mental workload, cognitive overhead, trust, individual confidence, machine accuracy, and perceived risk. This thesis studies the factors trust and workload in detail and adds usability.

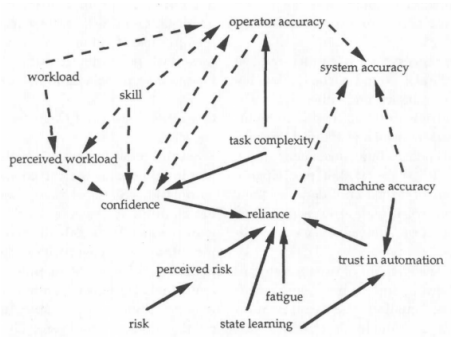


Figure 3.22: An outline of various factors impacting the adaptation and acceptance of automation, including trust, perceived risk, system accuracy and workload. Some of these factors affect each other, denoted by the arrows. The solid arrows depict relations verified by data, whereas the dotted lines are hypothesised connections (1997).

Trust

Trust is critical in adopting and accepting robots in society, especially in high-risk situations (Hancock et al., 2011). When trust decreases, operators are more likely to intervene in the robot task, called disuse. Disuse is the underutilisation of the automated system and forms unreliability (M. Lewis et al., 2018; Parasuraman and Riley, 1997). For illustration, in self-driving cars, the driver turns off or ignores all warnings of automated systems. Disuse primarily occurs at the onset of new interactions with automation (1997). Hence, the number of alarms should be restricted and false alarms should be avoided (1997).

When humans' trust is too elevated, it likewise creates unreliability. In this case, humans are over-reliant on automation (Lee and See, 2004; Parasuraman and Riley, 1997). In self-driving cars, for instance, humans start performing other activities and ignore potential hazards. This over-reliance is misuse and can occur in both trained and untrained operators (1997). Some strategies that oppose misuse are: making state indicators sufficiently salient and requiring active operator involvement in the process (1997).

Another disbalance in trust is in the matter of abuse, which is the inappropriate application of automation (Lee and See, 2004; Parasuraman and Riley, 1997). In self-driving cars, that would lead to putting cars in extreme conditions, such as people jumping in front of it. One prominent contributor to the abuse of automation is the absence of a human-oriented approach in the design phase of automation (1997). The role and responsibilities of the human must be defined based on capabilities rather than a by-product of automation.

Factors that influence trust are clustered by Schaefer (2016) in Figure 3.23, extending the three-factor model of Lee and See (2004) with extensive literature research. The nature of trust can differ for different LORA, and varying levels of trust correlate positively with varying LORA.

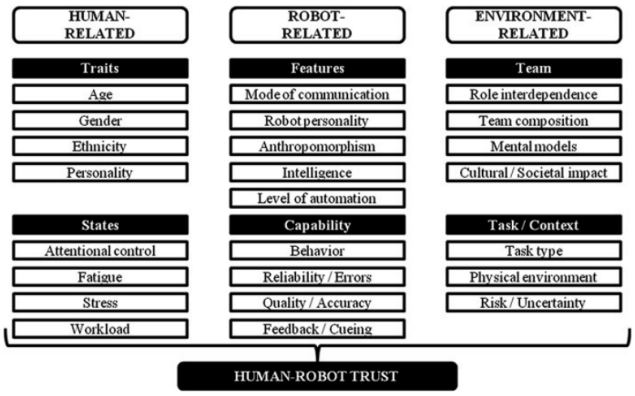


Figure 3.23: Factors that affect trust in HRI, based on the Three Factor Model of Human-Robot Trust of Lee and See (2004), with three main categories of factors: human-related, robot-related and environment-related. Schaefer (2016) extended these by performing an extensive literature review. For human-related factors, this supplied the sub-categories of traits and states. For robot-related factors, this provided the sub-categories of features and capabilities. For environment-related factors, this equipped the sub-categories of a team, and task/context (2016).

There are two types of interaction between humans and robots: performance-based or social-based (M. Lewis et al., 2018). In performance-based, there is a task with a precise performance goal. Trust can have multiple definitions; in performance-based interaction, an explicit definition is" an attitude which encloses the belief that the collaborator will perform as expected and can, within the limits of the designer's intentions, be relied on to accomplish the goals" (2018). Many definitions of trust reflect upon the role of expectations or attitudes of humans towards favourable responses (Heineman, 1984; Lee and See, 2004). These expectations entail predictability and include persistence, technical compe-

tency, and fiduciary responsibility (Heineman, 1984; M. Lewis et al., 2018). In Human-out-the-loop (HOTL), humans lose insight into the robot's actions and might lose their capacity to diagnose the problem and react accordingly (Beer et al., 2014; Parasuraman and Riley, 1997). Hence, HRI designs should be equipped with capabilities for humans to understand the robot's actions through software and interfacing (De Santis et al., 2008).

Expectations and trust can transform over time. Prematurely in the relationship, the trust in the system is predominantly based on its predictability (M. Lewis et al., 2018). When interaction with automation or robots takes place over a prolonged time, the human can convert trust into something called faith (2018). Thereby creating generalisations about the system and trust in the future behaviour of the system (2018).

Another noteworthy factor is system reliability. If the operator's trust is high, it creates a more reliable system, and consequently, the operator becomes more reliable. If the operator trust is low, more violations can occur, and the reliability declines (2018). This varying level of reliability is related to disuse and misuse.

So, human trust should be well-calibrated for the optimal performance of an HRI system. Therefore, Lee and See (2004) states that one should aspire to an interpretable, simplified and understandable automation, robot or algorithm, achievable through purpose, process, and performance. The purpose is to gain knowledge of what automation is presumed to do. The process is the system's functionality, and performance is how the robot performs in a certain task (2004; M. Lewis et al., 2018). Nevertheless, due to the robot's adaptability to new situations, trust can become increasingly complex.

Workload

Another influential factor in acceptation and adaptation of robots is workload. The workload is the amount of work humans have to do, performing a task with a robot within available time (Harriott, 2015; Parasuraman and Riley, 1997). This workload is mainly divided into mental and cognitive; regardless, other types of workload also exist, such as temporal demand (Hart and Staveland, 1988). HRI affects multiple human cognitive and perceptual resources such as their short-term memory, working memory, visual (seeing) senses, auditory (hearing) senses, and tactile (touch) senses (Prewett et al., 2010). Subjective measures of workload decrease with higher LORA (Steinfeld et al., 2006). The workload is a well-known metric in HRI, and human performance is dependent on workload (Parasuraman and Riley, 1997; Steinfeld et al., 2006). When the workload is excessively high or low, this could impact human performance and contribute to a loss of situational awareness. Situational awareness denotes "the ability to perceive, comprehend, and project the state of an environment" (2006), and when this awareness decreases, this could lead to dangerous circumstances. The workload can be measured either objectively or subjectively, and one of the most common methods is the NASA TLX survey (Harriott, 2015).

To make a new robot easy to adopt, it should have a fairly low workload without reducing situational awareness. A factor that can reduce situational awareness is transmitting large amounts of information to the same sensory modalities, which can cause interference (Hart and Staveland, 1988).

Usability

Usability is a fundamental metric for product development. ISO (2011), defines usability as *"Usability is the extent to which a product can be used by specific users to achieve specific goals with (i) effectiveness, (ii) efficiency, and (iii) satisfaction in a specified context of use"*. Effectiveness relates to the number of errors and successfully achieved activities. Efficiency is related to task time, effort, fatigue, and workload (Chacón et al., 2021). Satisfaction relates to the human subjective opinion,

whereas usability is related to user experience. User experience is the human experiences evoked while interacting with a technology (Chowdhury et al., 2020). The user experience can be positive when particular anticipated goals are met (2020). To quantify human satisfaction, usability testing scores are developed to compare diverse designs (Charlton and O'Brien, 2019).

Safety
Safety is critical, particularly in physical HRI (pHRI) (Abdelaal et al., 2019). In the scientific community, safety standards for pHRI need to be better-defined (De Santis et al., 2008). One of the most significant safety hazards is the chance of collision. The higher the robot energy, the more severe harm can be induced (Abdelaal et al., 2019).

All aspects such as mechanical, electrical and software should be considered to equip safety (De Santis et al., 2008; Okamura et al., 2010). The material design should limit sharp edges, excessive weight and inertia, for instance, by adding compliant elements (De Santis et al., 2008). However, these elements reduce the stiffness and affect a robot manipulator's accuracy and task performance. Other possible improvements are the addition of sensors and hardware and software safety procedures (2008).

Another way to increase robot safety is by employing the correct control method, which can enhance performance, reliability and safety (2008). These control methods can lead to a more intuitive pHRI, lowering the risk of collisions and damage.

Human factors
The prominent role of humans in HRI demands analysis of introduced errors with the Human Factor Analysis and Classification System (HFACS) (Shappell and Wiegmann, 2000). This framework represents human errors on four levels of failure, which united lead to substantial errors. The levels are (1) Organisational influences, (2) Unsafe supervision, (3) Precondition for unsafe acts, and (4) Unsafe acts of operators. Since each level impacts the subsequent, this is called the Swiss cheese model as portrayed in Figure 3.24 (2000). This framework regards both active and latent failures. The latter can go overlooked for a prolonged time and were disregarded before the Swiss cheese model despite their substantial role in accidents. This thesis only includes the level of unsafe acts of operators; others are explained in Appendix G.

Unsafe acts of operator
This level consists of the active failures at the operator's direct actions. These failures diverge into two categories (1) Errors and (2) Violations.
Errors
Errors are unintended behaviours and depict operators' mental or physical activities that fail to fulfil their intended outcome. Humans make errors by nature. Errors are split into three sub-categories:

- Decision errors
  Errors occur when the operator's actions proceed as intended, yet the plan proves inadequate for the desired end state. These are called "honest mistakes" since they associate to lacking appropriate knowledge or choosing inadequately with the best intentions. These are split into procedural errors, poor choices, and solving errors. Procedural errors appear when a situation is misdiagnosed and the wrong procedure is obeyed. Problem-solving errors surface in situations demanding novel solutions (2000).
- Skill-based errors
  These errors occur when a skill requiring little focus is overlooked, mainly due to attention or memory failure. Skill-based errors can also be technique errors; two people with similar training can still perform the same task with a different technique and have distinct errors.
- Perceptual errors
  Errors emerge when operators' sensory input is degraded, and a decision is founded on

that input. There is a disparity between the operator's perception and reality. To limit these errors, the operator must rely more on internal mechanisms rather than external cues

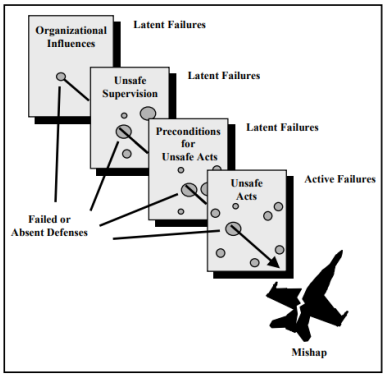


Figure 3.24: A schematic summary of the Swiss Cheese model utilised in the HFACS to investigate errors introduced by human factors. This model differentiates four levels of failures: organisational influences, unsafe supervision, preconditions for unsafe acts, and unsafe acts of operators. Each level impacts the successive level. The first three levels are latent failures, meaning they could have gone unnoticed for a prolonged time before provoking accidents. The last level consists of so-called active failures, which instantly lead to accidents. When all levels of failures are crossed via failed or absent defences, accidents will appear (2000).

Violations

Violations are intended to disregard the rules and regulations and are split into two sub-categories:

- Routine

  These violations are habitual actions of the operator tolerated by the supervisor as certain limits are tolerated, such as speeding at 5 km/h.
- Exceptional

  These violations arise non-regularly and are not tolerated or habitual. This behaviour is predominantly not explainable and consequently tough to anticipate

It is essential to consider that the Swiss Cheese model is theoretical and lacks details in its application for real-world situations. The model was primarily created and adopted by the aviation industry. It has been and could be applied to other industries, but it requires identifying specific holes in the system.

Contextual information

This part describes the contextual information underlying this thesis including the cochlear implant surgery and EMR robot.

3.5. Cochlear implant surgery

A cochlear implant (CI) is a medical instrument positioned inside the cochlea (NIDCD, 2021). The cochlea is a spiral structure of three fluid-filled canals in the inner ear, as shown in Figure 3.25 A. Hair cells in the cochlea move in reaction to sound frequencies, triggering nerve cells that convert these movements to electrical impulses sent to the brain, allowing the perception of sound (Hawkins, 2020).

If the hair cells, nerves or part of the brain is not functioning correctly, a CI could supply a sense of sound (Mudry and Mills, 2013; NIDCD, 2021). Figure 3.25 B displays a CI, consisting of a speech processor behind the ear that transforms sounds into electrical signals and sends them to the implant (Mudry and Mills, 2013). The implant is attached to a long electrode array reaching inside the cochlea, which triggers nerve cells when signalled (NIDCD, 2021). Figure 3.26 demonstrates the CI surgery in steps. CI surgery can be performed in various ways; the chosen approach is the round window approach (Jiam and Limb, 2016). The goal of this surgery is (1) placing a receiver and (2) locating

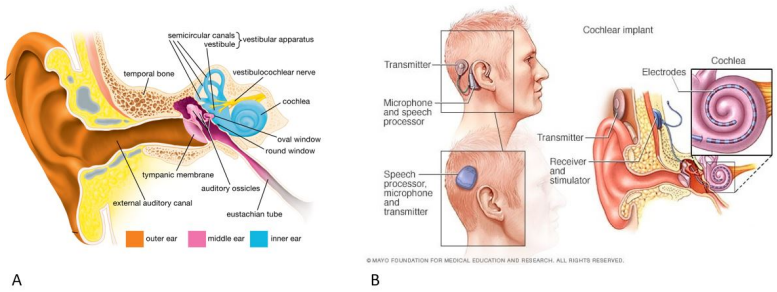


Figure 3.25: A) Schematic view of the human ear, where the cochlea and semicircular canals are displayed in the inner ear (Hawkins, 2020). The cochlea is a spiral structure of three fluid-filled canals, called the semicircular canals, in the inner ear, portrayed in blue. **B)** Schematic overview of a CI (Clinic, n.d.). A sound processor intercepts sounds, sending electrical signals to the transmitter, further forwarding the signal to a receiver in the middle ear. This receiver transmits the electrical signals down the electrode inside the cochlea, where it stimulates hair cells (n.d.).

the round window to insert an electrode.

First, creating an incision, then removing the hollow bone behind the ear through mastoidectomy, drilling layer for layer to reach the middle ear. Then a well is drilled so that the electrodes coming from the cochlea can be transferred to the receiver (Medicine, n.d.). Drilling continues until reaching the round window, removing the covering membrane and drilling its overhang to open it, allowing insertion of the electrode. Lastly, placing the receiver and connecting it to the electrode in the well (Clinic, n.d.; Medicine, n.d.). This explanation equips a general overview and lacks details for surgeons. The surgeons are guided by their perceptions, such as sight and touch, to execute surgery.

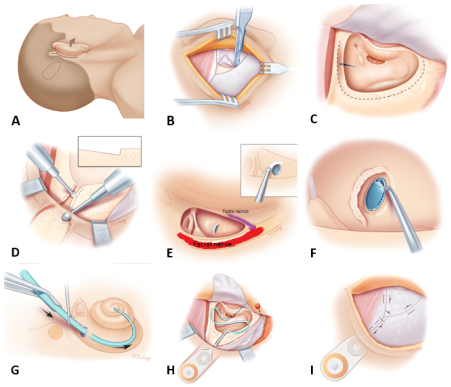


Figure 3.26: The steps for CI surgery according to Medicine (n.d.). **A)** Folding the ear forward. **B)** Creating an incision. **C)** Removing the hollow bone with a mastoidectomy to reach the middle ear. **D)** Drilling a well that allows the receiver placement and connection to the electrode later. **E)** Drilling until reaching the round window. **F)** Removing the round window covering membrane and overhang, opening it with a slight incision. **G)** Inserting the electrode into the cochlea. **H)** Positioning the receiver and connecting it to the electrode in the previously drilled well. **I)** Closing the incisions.

Major surgery challenges occur due to various nerves surrounding the round window. One is the facial nerve, illustrated in Figure 3.26 E in red. Another is the chorda tympani nerve (taste nerve), also shown in Figure 3.26 E in purple. These nerves create the so-called facial recess. This recess equips an average adult size of about 2.4 - 5.7mm, demanding high-accurate surgery to avoid damage to these nerves (J. Wang et al., 2020). Other critical structures are semicircular canals, the bone that isolates the brain from the mastoid cavity, and the sigmoid sinus, a large blood vessel (Hawkins, 2020). CI surgery can cause residual hearing loss due to high forces, cochlea trauma, fast electrode injection and acoustic trauma from drilling (Panara et al., 2021). An experienced CI surgeon can perform surgery in 80 to 90 minutes per surgery from incision to closure (2021). The accepted standard for facial nerve paresis for CI surgery is less than 0.1% (2021).

3.5.1. Surgical robot for the CI surgery

Robots in CI surgery can prevent residual hearing loss since they can improve accuracy and suppress trauma and high forces (2021). Robots do not rely on perception but rather on a pre-planned pat, requiring an extensive pre-operative phase identifying anatomical structures and fiducial markers (Caversaccio et al., 2017). The intra-operative workflow likewise requires adaptation as the robot performs the surgery after registration, and human assistants will perform patient fixation and preparation.

3.6. EMR robot

The EMR robot, in Figure 3.27, is a position-controlled 5 DOF robot with five motor modules and seven distinct coordinate frames running on Simulink Real-Time. The robot holds a drill at its end-effector that can be changed for a registration tool during registration. The position-control loop is switched to a admittance control-loop in registration.

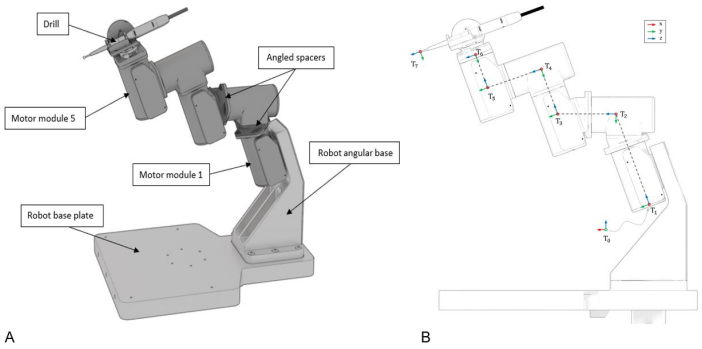


Figure 3.27: The 5 DOF EMR robot with A) showing the different parts of the robot such as the robot base plate, the motor modules, angles spacers and drill. B) Shows the different coordinate frames of the robot.

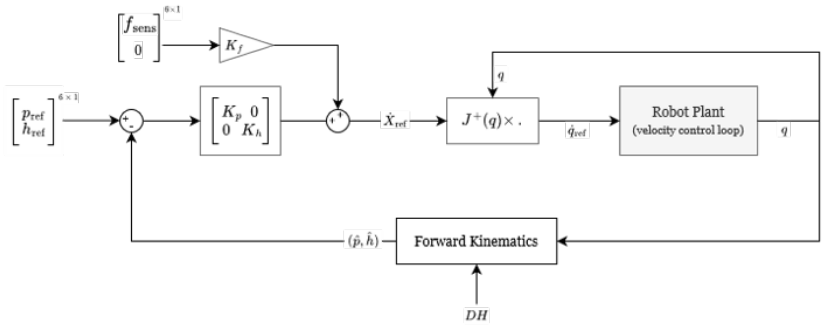


Figure 3.28: The EMR control loop with $\dot{x}_{ref}$ the reference endpoint velocity, q the joint angles, $\dot{q}_{ref}$ the reference joint velocities, $\hat{p}$ the estimated end-effector position, $\hat{h}$ the estimated end-effector heading angle, p_{ref} the reference end-effector position, h_{ref} the reference end-effector heading angle, f_{sens} the force sensor reading, J the Jacobian matrix, DH the denavit hartenberg paramters, K_f the translational force gain, K_p the set point positional gain and K_h the set point heading gain.

4. Concept selection

This chapter describes the selection procedure leading to the touch-based bone-anchored fiducial pair-point-based registration with the admittance control concept.

The best concept for registration in the given context is touch-based bone-anchored fiducial pair-point-based registration with admittance control due to its high accuracy, accessible robot workspace, robustness, low complexity, low maintenance, high patient safety, lack of intellectual property problems, possible usage of CT modality and limited development time.

An extensive selection procedure realised this result. This selection procedure regards two types of registration methods: general and specific. The general types are the ones provided by Alam et al. (2016) and Alam and Rahman (2016) given in Figures 3.1 and 3.8. The specific registration type reflects the details of performing a general registration type. For instance, surface-based registration would be a general type. However, within surface-based registration, there are numerous ways to gather the PCs, such as active tracking, laser surface scanning or using tracked pointers. The selection procedure delivered a general registration type, from which three specific registration types were selected, which led to four concepts, explained in Section I.1 and Appendix I. Then four stakeholders scored these with scoring criteria to conclude the best concept. The defining phase splits into three steps: general, specific, and concept selection, as portrayed in Figure 4.1. The applied criteria were drafted from the collected knowledge and displayed in Figure 4.3. Section 4.1.1 discusses the criteria in detail. From these criteria, the ones that have specific restrictions in this particular situation are utilised to equip exclusion criteria as shown in Figure 4.2 and discussed in Section 4.1.2.

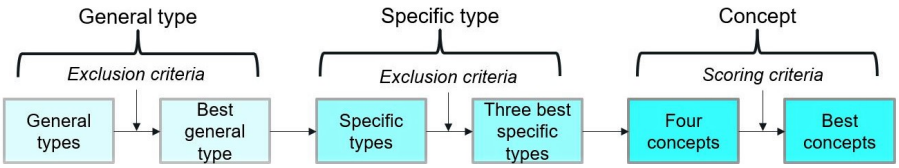


Figure 4.1: The methodology to select the best concept. Finding the general and specific method by applying exclusion criteria. Four concepts were drafted and scored by four stakeholders to conclude the best concept.

Exclusion criteria		
Criteria	**Meaning**	**Reason**
Sub-millimetric accuracy	Target registration error of < 1.0mm	The cochlear implant surgery crosses the facial recess, which has an average gap of 1.9mm
No limits to the robot workspace	Registration methods that provide limits to the robot workspace are excluded	The EMR robot can not work around an additional device and requires its complete workspace to execute the surgery
High robustness	Registration methods that have low robustness are excluded	The EMR robot can not adjust its plan as a response to perturbations as it executes a pre-programmed plan and lacks intelligence
Low system complexity	Registration methods that have a high amount of technical complexity and components are excluded	Middle- and low-income countries demand low system complexity
Low system maintenance	Registration methods that have a high amount of required maintencance are excluded	Middle- and low-income countries demand low system maintenance
High patient safety	Registration methods that pose additional safety threats to the patient are excluded	The medical application demands high patient safety
Limited training	Registration methods that require a high amount of training are excluded	Middle- and low-income countries demand limited training required
No intellectual property limits	Registration methods that are patented are excluded	EMR has to follow laws and regulations and can not adopt patented solutions
Only CT imaging modality	Registration methods that rely on other imaging modalities are excluded	The path-planning of the EMR robot is optimised for CT imaging, and low- and middle-income countries have higher accessibility to CT scans compared to other imaging modalities
Restricted development time	Registration methods that demand a development time > 4 months are excluded	The thesis project has a time limitation

Figure 4.2: The exclusion criteria, drafted from the scoring criteria and the contextual background as shown in Figure 4.2. The criteria are used to select the best general and specific type registration method.

Criteria	
Technical	**Meaning**
Accuracy	Depends on the errors, this has to be elaborated in a detailed error-analysis
Error detection	How easily can the different crucial errors be detected and measured
Efficiency	Computational efficiency achieves the result with little to no additional effort or energy. Limit the amount of time or memory required for calculations, the amount of computational effort
Level of automation	How autonomous is the method
Robustness	Small pertubations have little effect on the outcome
Robot workspace	How the method affects the robot workspace
Registration time	Time it takes to finish the registration procedure
Repeatability	Results reproduced from the same procedure should lie in a small range from each other
Human factors	**Meaning**
Ease of use	How the user (surgeon or other hospital employees) perceives how easily the procedure can be finished
Training needed	How fast the user can learn to use the technique
Human error	Decision error, skill-based error, perceptual error and how they influence the end-result
Influence of operator state	How much the user state can influence the end-result
Reproducibility	How easy it is for the user to reproduce the same procedure
Adaptation in workflow	How does the procedure influence the current workflow, and if additions to the current workflow are needed
Workload	How focused the user must be to provide good end-result
Clinical	**Meaning**
Invasiveness	The amount of additional stress for the patient by additional incisions etc.
Footprint	The amount of space it takes in the OR
Patient-friendly	If the patient has to wear some attribute between the pre-operative and intraoperative phases
General	**Meaning**
Costs	Additional costs for purchase, production and maintenance
Complexity	The number of interacting elements and technical complexity
Safety procedure	The ease of providing a safety procedure that is not relying on a single point of failure
Re-registration possible	The ease of re-registration when needed
Installation time	The time it takes for the first installation and initialisation
Additional parts	How much additional parts are needed
IP	Intellectual property, is this method already patented
Maintenance	The amount of maintenance needed

Figure 4.3: The scoring criteria used to select a registration method. Categorised into technical, human factors, clinical, and general

The best general registration type is the extrinsic invasive fiducial screw marker type, named the bone-anchored fiducial pair-point-based registration. The first four exclusion criteria applied to the general methods yield this result in Figure 4.4. Section 4.2 discusses the details of these results.

Extrinsic				
Invasive	Accuracy	Workspace	Robustness	Complexity
Stereotactic frame				
Fiducial screw markers				
Non-invasive	Accuracy	Workspace	Robustness	Complexity
Moulds				
Frame				
Dental adapters				
Fiducial skin markers				

Intrinsic				
Point-based	Accuracy	Workspace	Robustness	Complexity
Anatomical				
Geometrical				
Surface-based	Accuracy	Workspace	Robustness	Complexity
Rigid models				

Figure 4.4: Scoring the general extrinsic and intrinsic registration methods along the first four exclusion criteria from Figure 4.2. Red and green indicate the exclusion and inclusion of the method based on the criterion, respectively. Orange denotes a criterion that is neither met nor failed, thus not including nor including the method. This analysis concludes extrinsic, invaisve, fiducial screw markers registration is the best general method.Also called the bone-anchored fiducial pair-point-based registration method.

The three best specific types were touch-based, camera-based and head-fixation frame-based. Applying all exclusion criteria to specific bone-anchored fiducial pair-point-based registration methods delivers this result in Figure 4.5. Section 4.3 clarifies the scores and specific types in more detail.

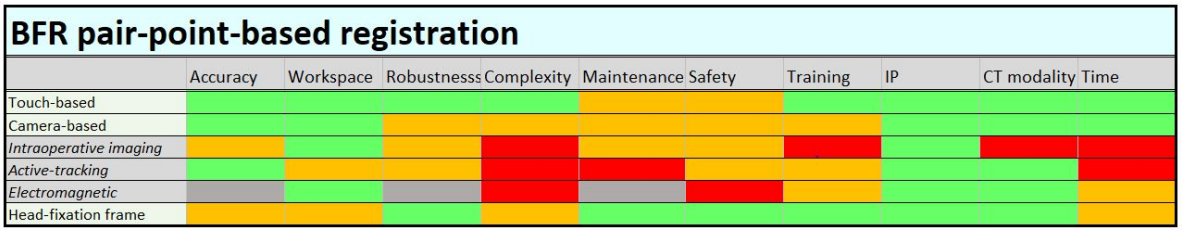


Figure 4.5: Scoring the specific bone-anchored fiducial pair-point-based registration methods along all exclusion criteria from Figure 4.2. Red and green indicate the exclusion and inclusion of the method based on the criterion, respectively. Orange denotes a criterion that is neither met nor failed, thus not including nor including the method. This analysis concludes touch-based, camera-based and head-fixation frame-based the best methods as they do not fail any criteria.

The best concept, touch-based bone-anchored fiducial pair-point-based registration with admittance control, was uncovered by drafting four concepts from the three specific types and scoring these. Four stakeholders scored the concepts along the scoring criteria in Figure 4.3. Section I.1 and Appendix I provide more details of the concepts.

4.1. Scoring and Exclusion criteria

The scoring and exclusion criteria were founded on theoretical and empirical gained knowledge and included both technical and HRI-related criteria. First, conducting the scoring criteria, then applying the thesis context and selecting those with specific restrictions generates the exclusion criteria. This process is clarified in Figure 2.6. Additional knowledge used is provided in Appendix G.

4.1.1. Scoring criteria

The scoring criteria were categorised into four different categories: (1) Technical, (2) Human factors, (3) Clinical, and (4) General. Accordingly, these scoring criteria evaluate both technical and HRI-related criteria. The clinical and general categories are likewise considered HRI-related criteria but, for clarity, were categorised separately.

Technical criteria

The technical criteria include ones related to the hardware and software performance of the registration method. One of the most important criteria is the **accuracy**. The TRE is considered an essential measure that should be minimised for accuracy. Besides, **how easily errors** in the registration system can be **detected, calculated and corrected** is crucial as they can lead to an inaccurate registration, hence surgery. The more easily errors can be detected, the easier it is to mitigate them and correct errors before they lead to large errors and complications.

Another criterion is the registration **robustness**, which indicates how well the registration method can resist small perturbations like a different starting position. In addition, the **repeatability** measures how well the registration results can be replicated by obeying the same procedure.

Due to time limitations in the OR as a consequence of hospital time restrictions and the patient being on anaesthesia, **efficiency** and **registration time** are essential criteria. Efficiency is the Computational efficiency to gain the result, meaning the computational time or memory required for calculations. The registration time is the time required for the registration process from start to finish. Further, permitting **re-registration** and to what extent assesses the easiness of repeated registration.
Another technical criterion is the **LORA** of the system, as a balance in automation and human involvement. Some problems require a high LORA, others a low LORA. Hence it should be adapted as needed. This also counts for the **workspace**. Each surgery requires a different workspace and, thus, different requirements.

Human factor criteria
As mentioned, human factors are critical in this thesis. These criteria relate to the challenges in HRI and human errors according to the cheese model mentioned by Shappell and Wiegmann (2000).

An important metric is the **ease of use** or **usability** of the system. As mentioned, this refers to how the user, surgeon or other employees perceive the easiness of finishing the procedure and achieving the stated goal. The usability relates to the **workload**, which describes how much user effort is needed to provide a good result.

In addition, the amount of **human error** introduced in the system is crucial and hence a scoring metric. Human error contains the number of possible decision errors, skill-based errors, and perceptual errors that can occur. The human error relates to the amount of **influence the operator** has on the registration method and outcome. If possible human errors are large, but the influence is small; they balance out. In addition, **reproducibility** measures how easy it is for the operator to reproduce the same results with the same procedure. Similarly, the **amount of training** required to use the registration method is a scoring metric. It indicates low reproducibility if large amounts of training are needed to become a skilled operator.

Lastly, the **adaptation in workflow** measures how the procedure influences the current workflow and the extent of additions required to incorporate it.

Clinical factor criteria
Clinical criteria are related to the robot's medical environment. Since CIS is mainly used for MIS, it is necessary to minimise the **invasiveness** of the registration method. If the surgery is minimally invasive, but the registration method is heavily invasive, this contradicts the requirements. Besides, invasive procedures add to the patient's risk as they can lead to complications. Related to the patient risk is the **patient-friendliness** where some registration methods require patients to have markers attached to their skull between the pre-operative and intraoperative phases. These markers can pose additional complications and discomfort.

Lastly, the **footprint** of the robot and registration method inside the OR is essential since the OR is packed, demanding space efficiency (Wolf and Shoham, 2009).

General factor criteria
General criteria relate to general machine criteria that are not a robot or registration specific. The first criterion is the **costs**; these require adoption according to the targeted customer base. Costs include those related to the registration method's purchase, production and maintenance. The amount of **maintenance** required reflects the amount of maintenance required over time to maintain system accuracy and safety.

In addition, the **complexity** of the registration system relates to how difficult it is to use the system. The higher the complexity, the higher the chance of error occurrence. This also correlates to the **amount of additional** parts, the more parts, the more likely they can get lost or need replacement which can be difficult in resource-deprived surroundings. Another criterion related to this is the **installation time**; longer installation times demand more knowledge and resources. The **safety procedure** measures the ease of supplying a safety procedure that does not rely on a single point of failure.

The last criterion is **intellectual property (IP)**. Intellectual property is vital since methods that are patented can not be used.

4.1.2. Exclusion criteria

Applying the specific context to the scoring criteria yields the exclusion criteria. These criteria are thus context-specific.

Sub-millimetric accuracy
The first **criterion entails** low accuracy due to the surgery at hand. As mentioned, the distance between the facial and chorda tympani nerve is about 2.4–5.7mm. In contrast, the width between the drill boundaries and the nerves should be 0.5mm or less for inserting the electrode (J. Wang et al., 2020), thus excluding methods with a TRE larger than 1.0mm.

No limit to robot workspace As the robot needs to execute a micro-surgery with high accuracy, it demands non-restricted freedom of movement to reach all the required positions and angles, excluding registration methods that restrict this workspace.

High robustness Since the EMR robot lacks machine intelligence, it can not adjust its plan in response to perturbations. Hence, robustness is significant, excluding registration methods with low robustness.
Low system complexity System complexity can lead to additional system failures and errors. It also demands more intellectual resources, which are restricted in the middle- and low-income countries, excluding registration methods with high system complexity.

Low system maintenance System maintenance requires monetary, personnel and intellectual resources, limited available in the middle- and low-income countries, thereby excluding systems demanding high maintenance.

High patient safety As this project entails a medical application, patient safety is crucial. Therefore registration methods that involve low patient safety are excluded.

Limited training As mentioned, the higher the required training, the lower the robustness. Additionally, a large amount of training requires personnel, time and monetary resources, which has limited availability in the middle- and low-income countries, excluding registration methods that require extensive training.

No IP limits EMR is a company that operates according to rules and legislation, thus not allowing patented solutions, hence excluding methods with IP limitations.

Only CT imaging modality The path-planning of the EMR robot is optimised for CT imaging. Furthermore, low- and middle-income countries have higher accessibility to CT scans than other imaging modalities, thus excluding methods that can not operate with CT modalities (N. Shah, 2014).

Restricted development time This thesis project has a time-limited factor; therefore, it can only consider some possible design implementations, excluding methods that demand a development time of >4 months.

4.2. General registration selection

The first four exclusion criteria are applied to the general methods to select the best general registration method for the application. Figure 4.4 summarises the scoring results and concludes invasive fiducial screw markers the best method with no criteria failed. Red and green indicate the exclusion and inclusion of the method based on the criterion, respectively. Orange denotes a criterion that is neither met nor failed, thus not including nor including the method. This research excludes voxel-based and deformable model surface-based registration methods. It was found that non-invasive extrinsic registration methods scored the worst by failing two or more criteria with an exception for dental adapters. Details about the scoring are given in Appendix J.

4.3. Specific registration selection

The specific bone-anchored pair-point-based fiducial registration methods are summarised in Figure 4.6. Touch-based is performed by physically touching the fiducials through an exterior probe or with the robot end-effector. Camera-based uses a camera to determine the fiducial locations. This is also the case for active tracking, but here the locations are tracked in real-time. Intraoperative utilises an imaging modalitity to determine fiducial locations during surgery. The electromagnetic method uses electromagnetic forces to localise the fiducials. Lastly, the head-fixation frame uses two objects, one fixed to the base of the robot and the other one fixed to the patient's head. These two objects can only fit on top of each other in one way.

Scoring each method led to the results shown in Figure 4.5, concluding the three best methods to be touch-based, camera-based and head-fixation frame since they do not fail any criteria. Red and green indicate the exclusion and inclusion of the method based on the criterion, respectively. Orange denotes a criterion that is neither met nor failed, thus not including nor including the method. Details about the scoring is given in Appendix J.

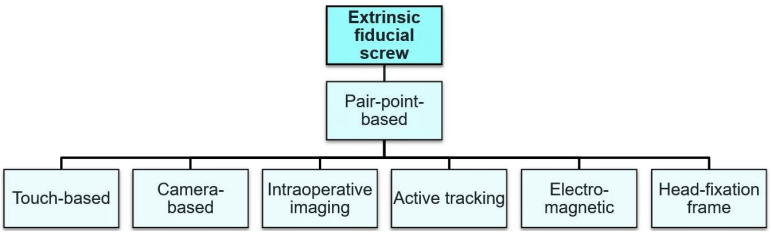


Figure 4.6: The defined specific registration methods for bone-anchored pair-point-based fiducial registration.

4.4. Concept selection

Three methods did not fail any criteria: touch-based, camera-based and head-fixation frame. From these criteria, four concepts were drafted and scored along the scoring criteria: touch-based bone-anchored fiducial pair-point-based with admittance control, touch-based bone-anchored fiducial pair-point-based with tele-operation, camera-based bone-anchored fiducial pair-point-based with calibration balls and head-fixation frame-based registration through ball in grooves. More details are given in Appendix I. The concept scoring uncovers touch-based bone-anchored fiducial pair-point-based registration with admittance control as the best. Admittance control will take place by translating the forces and torques applied by the human operator to the velocity movements of the robot. More about admittance control can be found in Appendix G. In this design, the human operator and the robot will collaboratively execute the localisation task.

4.4.1. Scoring

Scoring the concepts along the scoring criteria in Figure 4.3 on a scale from 0 to 5. Each criteria was scaled from 0 to 4 in importance providing an inverse weight to multiple the given score. The results are provided in Appendix I and yields an average score of 221.25, 218, 179.5 and 212.33 for concepts 1 to 4. Concept 1 scores best, whereas concept 3 scores the lowest.

4.5. Conclusion

This research concludes that touch-based bone-anchored fiducial pair-point-based registration with admittance control is the best registration method that satisfies the scoring and exclusion criteria related to the contextual background with an HRI focus—providing an answer to the first sub-question.

5. Design implementations

This chapter describes the design implementations for multi-modal multi-feedback admittance control to perform touch-based bone-anchored fiducial pair-point-based registration. Further, it documents the related steps that led to these design implementations.

Figure 5.1 displays the chosen implementation functions. These are divided into design functionalities and other implementations. The design functionalities are related to multi-modal admittance control and a multi-feedback interface. The other implementations; training, roles and responsibilities; are not related to direct functionalities but rather to design documentation and to ensuring the design is used correctly. Section 5.3 elaborates on these implementations.

Design Functions
Multi-modal admittance control
Noise filtering
Safety thresholds
Rotational control
Translational control with different speeds
Fixed mode
Safety mode
Switching interface
Registration button
Tuned parameters
Multi-feedback interface
Step-by-step procedure
Checklist, including material checks
Mode feedback
Force exeeding warning
Torque exceeding warning
Robot workspace warning
Sound when position saved
Option for repeated localisation
Extra marker opportunity
Automatically find clustering datapoints
Visual feedback on localisation performance
Filter out double saved data
Averaging of clusters
Filter outliers
Adjustable outlier threshold
Track time
Process bar
Automatically save data

Other implementation
Training, roles and responsibilities
Supervisory role
Training procedure for different roles
Description of roles and responsibilities
Training case-studies
Robot-error protocol

Figure 5.1: The final design implementations split in Design functions and other implementations. The design functions are either related to a multi-modal admittance control or a multi-feedback interface. Colour-coding the functionalities according to their relationships. White multi-modal admittance control related functions illustrate general admittance control loop functionalities, whereas light blue indicates multi-modal functionalities and dark blue parameter tuning functionalities. White multi-feedback interface related functions illustrate GUI functionalities, whereas light blue indicates visual and auditory feedback functionalities. The other implementations are not related to direct functionalities but rather to design documentation, indicated in light blue, and to ensuring the design is used correctly

Therefore it was concluded that a multi-modal multi-feedback design was best for the touch-based bone-anchored fiducial pair-point-based registration with an admittance control concept. This design was conducted to achieve the goal stated in Chapter 1: Designing an HRI-focused, sub-millimetric accuracy patient-to-image registration method for robotic CI surgery that achieves high performance measured in accuracy, workload, usability and trust. Hence the design was built to maximise its accuracy, usability and trust while minimising its workload.

This design was yielded by acquiring mitigation strategies for possible active errors and failures that can limit the design's success measured in accuracy, workload, usability and trust. Section 5.2 de-

scribes these mitigation strategies. Design influences were analysed, from which mitigation strategies were formulated, scored and selected to mitigate external influences that negatively impact the design and enhance favourable effects. Section 5.1 discusses the design influences, leading to the design implementations as demonstrated in Figure 5.2.

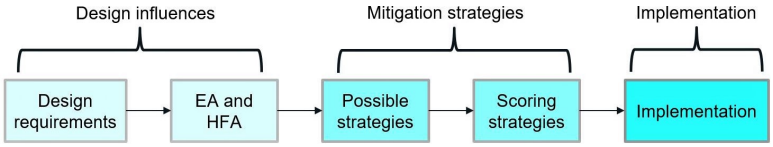


Figure 5.2: The design phase methodology to select the best concept, separated into design influences, mitigation strategies and implementations. The design influences were compiled by stating design requirements and performing an extensive EA and HFA on active errors and failures. The mitigation strategies were gathered by drafting many strategies and scoring them according to their advantages and risks. Implementation performs the final design implementations.

The design influences were gathered by stating design requirements and performing an extensive EA and HFA on active errors and failures. The mitigation strategies were scored according to their advantages and risks and conformed to the design requirements. The selected implementations were enforced, finalising the design and answering Q1.

5.1. Design influences

The design influences consist of requirements that should be met and factors that influence the outcome. The requirements are divided into technical, human-factor, clinical, and general. The primary influencing factors are image localisation, robot localisation and registration algorithm-related factors.

Design requirements	
Technical	**Meaning**
Sub-milimetric accuracy	TRE < 1.0mm
No limits to robot workspace	Robot workspace should not be hindered
Maximum registration time	Maximum registration time of 5 minutes
Data saving	Option to save data so that all movements, forces and errors can be
Minimum amount of buttons	Minimise the amount of buttons or controls required
Human factors / usability	**Meaning**
Intuitive usage	Using the system should come natural to the operator
Engaging interface	The interface should be engaging
Understandable design and GUI	The design and GUI should provide the operator with an understanding of the registration process
Low workload	The design should not require much operator effort
Limited amount of actions	The number of actions required to perform registration should be
Buttons in close proximity	Buttons that are needed should be in close proximity to the operator
Not much training required	The system should be easy to learn for any operator
Low influence of human error	The influence that human error has on the outcome should be limited
Low influence of operator state	The operator state should have restricted influence on the outcome
Minimal change in workflow	The registration design should disrupt the known workflow as little as
Simple feedback	Provided feedback should be easy to understand
Limit sensory overload	Too much feedback to the same sensory modality should be prevented
Clinical	**Meaning**
Minimal invasiveness	The amount of additional stress for the patient should be limited
Small footprint	The design should be limited in its additional footprint inside the OR
General	**Meaning**
No single point of failure	The design should not rely on single point of failure
Safe for any operator	The design should be safe for any operator, regardless of skill
Re-registration possible	Re-registration should be possible when needed
Limited additional parts	How much additional parts are needed should be minimised

Figure 5.3: The design requirements acquired from the scoring and exclusion criteria in Chapter 4.

5.1.1. Design requirements

The goal is to implement admittance control for the EMR robot to execute touch-based bone-anchored fiducial pair-point-based registration with high performance. The scoring and exclusion criteria from Figures 4.2 and 4.3 were employed to assemble design requirements. Not all criteria were included,

as some were insinuated in the chosen concept, such as high robustness. The requirements are divided into technical, human-factor, clinical, and general and summarised in Figure 5.3.

5.1.2. Error- and Human-factor analysis

The TRE is the most important measure of accuracy and system output in registration. The FLE directly influences the TRE. Therefore the factors that impact the FLE are examined. Three main categories influence FLE, as summarised in Figure 5.4:

- *Registration Algorithm*
  The registration algorithm is the algorithm that calculates the final transformation matrix.
- *Image Localisation*
  Image localisation is the process of finding the position of the fiducials in the image coordinate system.
- *Robot Localisation*
  Robot localisation is the process of finding the position of the fiducials in the robot coordinate system.

These categories and influences are extensively analysed in Appendix H, which divides them further into sub-categories. The sub-categories related to CT localisation and robot localisation are split into sub-subcategories as shown in Figures 5.5 and 5.6.

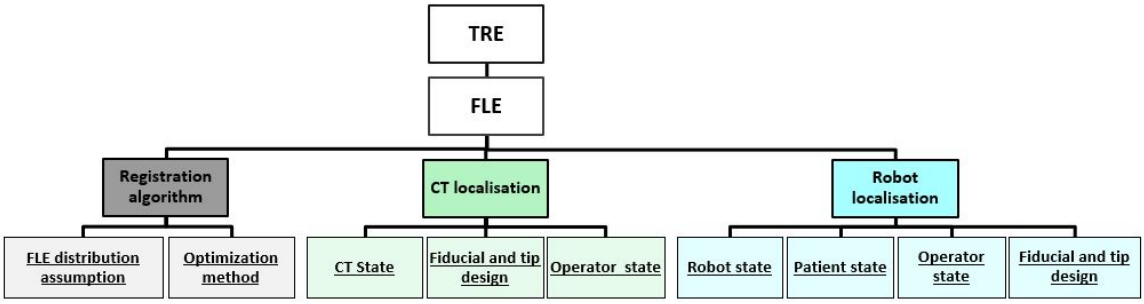


Figure 5.4: An overview of the main categories that influence the TRE in pair-point-based registration. The TRE is directly affected by the FLE (Fitzpatrick, 2009; Gerber et al., 2013; Oliveira and Tavares, 2014). The FLE, in turn, is affected by three main categories: registration algorithm, CT localisation and robot localisation. The registration algorithm is separated into FLE distribution assumptions and optimisation methods. CT localisation is split into CT state, fiducial and tip design, and operator state. Robot localisation is divided into robot state, patient state, operator state, and fiducial and tip design.

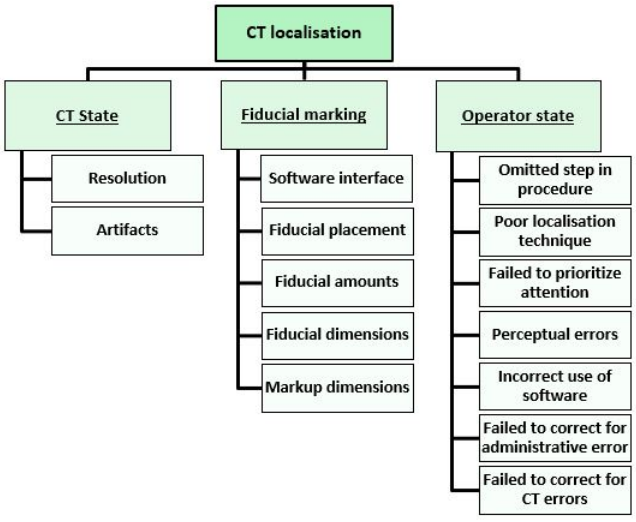


Figure 5.5: An overview of the sub-categories and sub-sub-categories under the CT localisation errors that influence the FLE in pair-point-based registration.

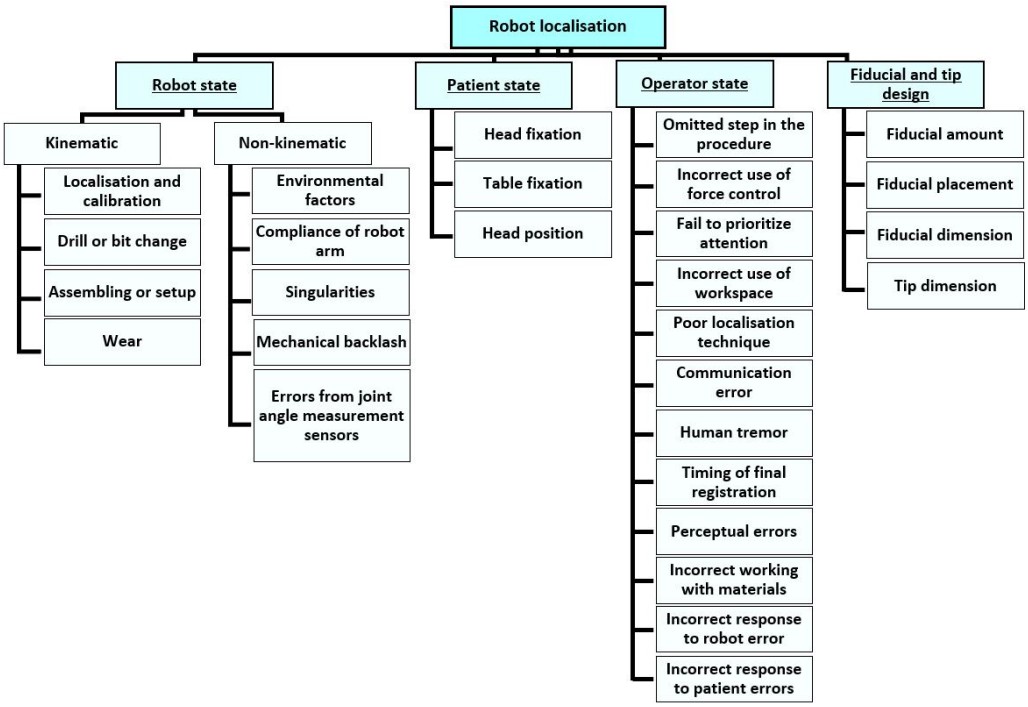


Figure 5.6: An overview of the sub-categories and sub-sub-categories under the Robot localisation errors that influence the FLE in pair-point-based registration.

5.2. Mitigation strategies

Mitigation strategies were conducted and scored from the design requirements and failures associated with the robot localisation operator state. Numerous mitigation strategies were drafted and scored according to their advantages and risk. Their advantages and risks are also related to other errors, such as the robot state.

5.2.1. Possible strategies

Each robot localisation operator state error, as highlighted in Figure 5.6, includes multiple possible strategies. The possible strategies drafted and their descriptions are given in Appendix K.

5.2.2. Scoring strategies

The found strategies are scored on their advantages and risks, as demonstrated in Appendix K. Strategies that overlapped from different errors and provided clear advantages and low risks were implemented, leading to the functionality list shown in Figure 5.1.

5.3. Implementations

The implementations led to a multi-modal multi-feedback admittance control design to achieve touch-based bone-anchored fiducial pair-point-based registration. The implementations of this design are divided into three different categories:

- *Multi-modal admittance control*
- *Multi-feedback interface*
- *Training, roles and responsibilities*

And led to the summary of actions and feedback provided in Table 5.1 with the safety thresholds and warnings in Table 5.2 for force, torque and position. Each category is explained in more detail.

Table 5.1: Summary of foot pedal events and related actions and feedback

Foot pedal	Actions	Feedback
Pedal	**Translational mode** Gains are set, operator can move end-effector in [x y z]. K_t is dependent on pedal input. Force is limited to 8N. Force cut-off value [-1 1]	Visual: • Blue light when within safety threshold force (8N), torque (0.2N) and workspace |100mm| • Red light when outside safety threshold for force (8N), torque (0.2N) or workspace |100mm|
Left btn	**Registration mode** Hold end-effector at its current position. Save datapoint to matrix.	Visual: • Multicolor light • Show position of saved data to supervisor • Add to number of saved datapoints Auditory: • Beeping sound
Right btn	**Rotational mode** Gains are set, operator can change heading angle. Torque is limited to 0.2N. Torque cut-off value [-0.3 0.3]	Visual: • Blue light when within safety threshold force (8N), torque (0.2N) and workspace |100mm| • Red light when outside safety threshold for force (8N), torque (0.2N) or workspace |100mm|
Nothing	**Fixed or Safety mode** Hold end-effector at its current position. After seven seconds of inactivity set gains and move slowly towards safety point. Positional error is limited to 5mm.	Visual: • Green light when within safety threshold force (8N), torque (0.2N) and workspace |100mm| • Red light when outside safety threshold for force (8N), torque (0.2N) or workspace |100mm|

Table 5.2: Different safety thresholds and warnings for three different metrics: force, torque and position. The workspace is not limited with a threshold, but provides only a warning.

Metric	Safety threshold	Warning
Force	Force vector max = 8N	Force vector 8N
Torque	Torque vector max = 8N	Torque vector 8N
Position	Positional error max = 5mm	Workspace vector 100mm

5.3.1. Multi-modal admittance control

Multi-modal admittance control functions relate to three categories: (1) General admittance control loop, (2) Multiple modes, and (3) Parameter optimisation.

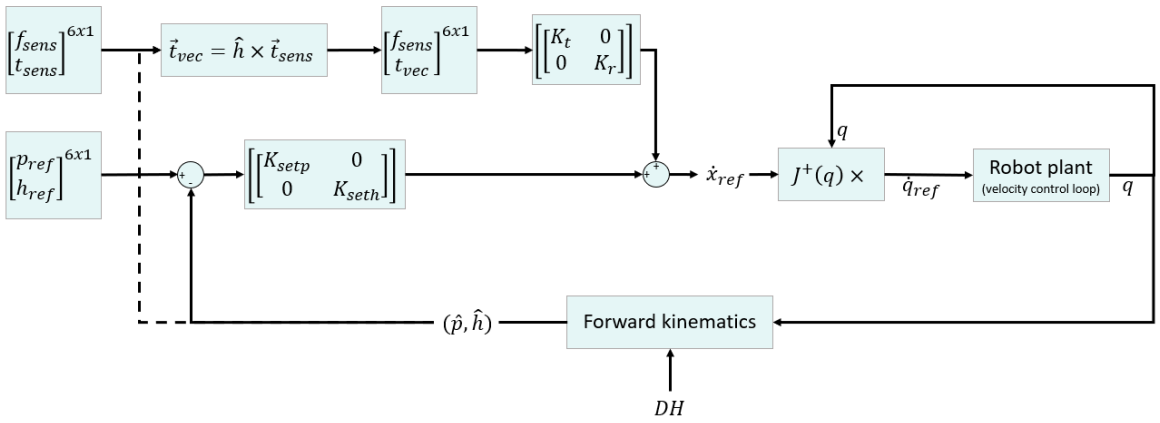


Figure 5.7: The multi-modal admittance control loop with $\dot{x}_{ref}$ the reference endpoint velocity, q the joint angles, $\dot{q}_{ref}$ the reference joint velocities, $\hat{p}$ the estimated end-effector position, $\hat{h}$ the estimated end-effector heading angle, p_{ref} the reference end-effector position, h_{ref} the reference end-effector heading angle, f_{sens} the force sensor reading, t_{sens} the torque sensor reading, t_{vec} the torque vector, J the Jacobian matrix, DH the denavit hartenberg paramters, K_t the translational force gain, K_r the rotational force gain, K_{setpp} the set point positional gain and K_{seth} the set point heading gain.

The control design created to achieve the best system performance in executing touch-based bone-anchored pair-point fiducial registration is multi-modal admittance control, as shown in Figure 5.7. The control loop gains are K_t, K_r, K_{set_p} and K_{set_h}. With K_t and K_r, the translational gain and the rotational gain, respectively. These allow the operator to move the robot end-effector through adequate localisation techniques and admittance control usability. K_{set_p} and K_{set_h} are the positional and heading set point gains, respectively. These gains permit the robot to move to a certain set point in space. The control loop provides multiple modes of control that correspond to different tasks.

General admittance control loop

As mentioned in 3.6, the admittance control loop from the robot looked like Figure 3.28. This project elaborates this control loop with better noise filtering and safety thresholds as summarised in Figure 5.8. More details are provided in Appendix L.

General control loop	
Noise filtering	
Low-pass force filter	Second-order low-pass
Safety thresholds	
Force threshold	8.0N
Torque threshold	0.2N
Speed threshold	$\lvert Pos_{error} \rvert = 1.0mm$

Figure 5.8: General admittance control loop improvements related to noise filtering and safety thresholds.

Multiple modes

Different modes were developed for multi-modal admittance control. These modes have distinct functionalities adapted to optimise the system's performance. 5.3 gives an overview of the different modes, their description and the control gains.

Table 5.3: Overview of the different control modes, their description and control gains. With K_t and K_r the translational gain and the rotational gain, respectively and K_{setp} and K_{seth} the positional and heading set point gains, respectively.

Mode	Description	K_t	K_h	K_{setp}	K_{seth}
Translational	Operator can move end-effector in [x y z]	0.0-4.0	0.0	0.0	0.0
Rotational	Operator can change heading angle of end-effector	0.0	0.7	0.0	0.0
Fixed	Robot is not moved by either operator or itself	0.0	0.0	0.0	0.0
Safety	Robot moves slowly towards set position and heading	0.0	0.0	0.2	0.2
Registration	Robot is not and end-effector coordinates are saved	0.0	0.0	0.0	0.0

The modes are triggered through a foot-operated interface with several buttons, each dedicated to a specific function, adhering to the conditions:

if Left pedal is pressed **then**
  Translational mode
else if Right button is pressed **then**
  Rotational mode
else if Left button is pressed **then**
  Registration mode
else if Safety timer < 7s **then**
  Fixed mode
else
  Safety mode
end if

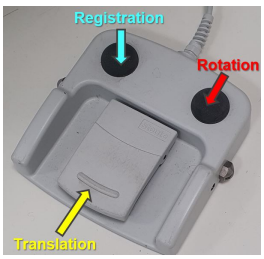


Figure 5.9: The foot pedal that the operator uses for switching to different modes of control according to the conditional statement. The pedal, as shown with the yellow arrow, is used to switch to the translational mode and adapt the speed of the robot arm accordingly. The left button, indicated with the light green arrow, is used to switch to the registration mode. The right button, specified with the orange arrow, switches to the rotational mode. The fixed mode activates through foot pedal inactivity. The safety mode starts when the inactivity endures for more than seven seconds.

Translational mode

In this mode, the human can control the robot in the x-, y- and z-plane without altering the heading angle of the robot end-effector. The gains K_r, K_{setp} and K_{seth} are set to 0.0. Translational control requires the endpoint velocity related to translation. The endpoint velocity is calculated with the formula

$$\dot{x} = K_t(x_{ref} - \hat{x}) \tag{5.1}$$

with x_{ref} the reference position and $\hat{x}$ the current estimated position. The reference position, x_{ref}, utilises the operator-applied force vector. Accordingly the formula is deducted to

$$\dot{x} = K_t(f_{sens} - \hat{x}) \tag{5.2}$$

to calculate the joint velocities. The translational gain K_t calculates the endpoint velocity with respect to translation and increases proportionally to the pedal input P as depicted in Figure 5.10. The gain is specified with the equation

$$\begin{aligned} K_t &= \frac{K_{t_{\text{max}}} - K_{t_{\text{min}}}}{P_{\text{max}} - P_{\text{min}}} * P - \frac{K_{t_{\text{max}}} - K_{t_{\text{min}}}}{P_{\text{max}} - P_{\text{min}}} * P_{\text{min}} \\ &= \frac{4.0 - 0.0}{30000 - 200} * P - \frac{4.0 - 0.0}{30000 - 200} * 200 \\ &= 1.3423 * 10^{-4} * P - 0.0268 \end{aligned} \tag{5.3}$$

with $K_{t_{\text{max}}}$ the maximum translational gain, $K_{t_{\text{min}}}$ the minimal translational gain, P_{max} the maximum pedal value, P_{min} the minimum pedal value and P the current pedal value. These values are summarised in Table 5.4. System stability in its contact dynamics establishes the maximum gain.

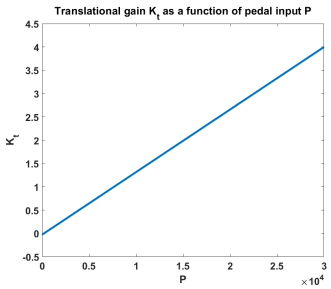


Figure 5.10: The translational gain K_t as a function of the pedal input P from the operator. Both parameters are unitless. The gain K_t increases linearly, following Equation 5.3 with a minimal value of 0.0 and a maximum value of 4.0.

The translation gain values are shown in the first row of Table 5.3, which permits the operator to move the robot end-effector in x-, y-, and z-plane with different gains related to the pedal value that the operator employs.

Table 5.4: Variables in determining the translational gain K_t according to Formula 5.3

Variable	Meaning	Value
$K_{t_{\max}}$	Maximum translational gain	4.0
$K_{t_{\min}}$	Minimal translational gain	0.0
$P_{\max}$	Maximum pedal value	30000
$P_{\min}$	Minimal pedal value	200

Rotational mode
In this mode, the operator can regulate the heading angle of the robot end-effector while holding the same position in the x-, y- and z- planes. The gains K_t, K_{setp} and K_{seth} are 0.0. The rotational gain K_r is 0.7. The endpoint velocities must be converted to joint velocities for the control loop. The endpoint velocity for rotation can be found with the cross-product of the current vector and the measured torque as described by Zhao (2016), and in the formula,

$$\vec{t_{vec}} = \vec{h_{cur}} \times \vec{t_{sens}}. \tag{5.4}$$

With $\vec{h_{cur}}$ the estimated heading angle as is and $\vec{t_{sens}}$ the torque vector measured by the sensor. $\vec{t_{vec}}$ is the rotational error that leads to the angular velocity ω with the formula

$$\omega = K_r(\vec{t_{vec}}) \tag{5.5}$$

ω is used to calculate the joint velocities of the robot. The second row of Table 5.3 summarises the gains, which allows the operator to adjust the heading angle without moving the robot in the x-, y-, and z-direction.

Fixed mode
The fixed mode starts when the pedal input P is zero. All gains K_t, K_r, K_{setp} and K_{seth} are 0.0. Thus, the operator cannot move the robot end-effector in any position or orientation, and the robot will not initiate any movement. This mode is beneficial to maintain the end-effector position fixed for saving the position or when the operator requires a break. For instance, the patient or the OR needs attention for other tasks like disinfecting the robot arm due to contamination. Any activity of the operator on the foot pedal ends the fixed mode. The fixed mode gains are shown in the third row of Table 5.3. These avert any movement of the robot end-effector initiated by neither the robot nor the operator. This mode is essential for the final registration to hold position and as a safety measure.

Safety mode
The safety mode initiates after the fixed mode timer extends for seven seconds. Any action of the operator on the foot pedal will reset this timer. The operator gains K_t, K_r are 0.0, whereas the robot gains for its movement towards the set position and set heading K_{setp} and K_{seth} are both 0.2. This allows the robot to move away from the patient. This mode prevents the robot from causing harm when distracted by the operator. The fourth row of Table 5.3 summarises the gain values.

Registration mode
In this mode, the gains K_t, K_r, K_{setp} and K_{seth} are all 0.0, and no difference in robot control is present compared to the fixed mode. Registration mode can be activated by pressing the left button and releasing the pedal, first initiating the fixed mode. The safety mode provides the operator feedback, discussed in the multi-feedback interface section. It is considered a mode so the operator can inspect its workings and ensure the data points are saved. The fifth row of Table 5.3 summarises the gain values. The data is saved by continuously scanning the foot pedal signals through a timer that updates the GUI. The timer frequency is 2000Hz. Every update follows the condition

if Left pedal is pressed and was not pressed before **then**

Get current [x y z] position
Save current [x y z] position to the end of the matrix
Save timing of saving position
Provide information to the supervisor about saved data and the number of data-points
Set previous value to current value
end if

Parameter optimisation
The parameters are optimised accordingly. The most critical parameters are control gains, force sensor cut-off values, filter pole and damping, positional and heading reference and the safety thresholds. The results are summarised in Figure 5.11. More details are provided in Appendix L.

Parameter optimisation	
Control gains	
Translational gain	4.0
Rotational gain	0.7
Set point gains	0.2
Force sensor	
Cut-off values force	[-1,1]
Cut-off values torque	[-0.3, 0.3]
Filter pole	10Hz
Filter damping	0.7
Reference position	
Positional reference	[315 -120 50]
Heading angle reference	[0 0.8829 -0.4695]
Safety thresholds	
Force threshold	8.0N
Torque threshold	0.2N
Speed threshold	$\lvert Pos_{error} \rvert$ = 1.0 / 5.0mm

Figure 5.11: A summary of the parameter optimisation results in optimising the control gains, force sensor readings including cut-off values and filter parameters, reference position, and safety thresholds.

5.3.2. Multi-feedback interface

Multi-feedback in this project is defined as feedback given at multiple alternative sensory modalities (Hart and Staveland, 1988). Supplying feedback in three different ways, as summarised in Figure 5.12: (1) GUI, (2) Auditory feedback, and (3) Visual feedback. Although the GUI can be considered visual feedback, it is categorised differently since it mainly provides feedback to the supervisor. In contrast, the auditory and visual feedback is provided primarily to the operator. Additionally, the GUI provides feedback on different aspects of the workflow compared to the auditory and visual feedback that only occurs during localisation.

GUI
The GUI implementation was coded through Matlab AppDesigner. The GUI consists of the following steps, depicted in Figure 5.12: (1) Description, (2) Checklist before, (3) Localisation, (4) Results, (5) Checklist after, and (6) Validation. These steps and their functionalities are discussed in brief. Appendix M describes these steps and functionalities in more detail and provides the GUI workflow.

Auditory feedback
Auditory feedback pays a sound when registration mode is active, saving the end-effector positions. The auditory feedback is executed inside the continuous timer update function described in the registration mode. Updating the condition according to

if Left pedal is pressed and was not pressed before **then**
Get current [x y z] position
Save current [x y z] position to the end of the matrix
Save timing of saving position
Play sound 0.1 s

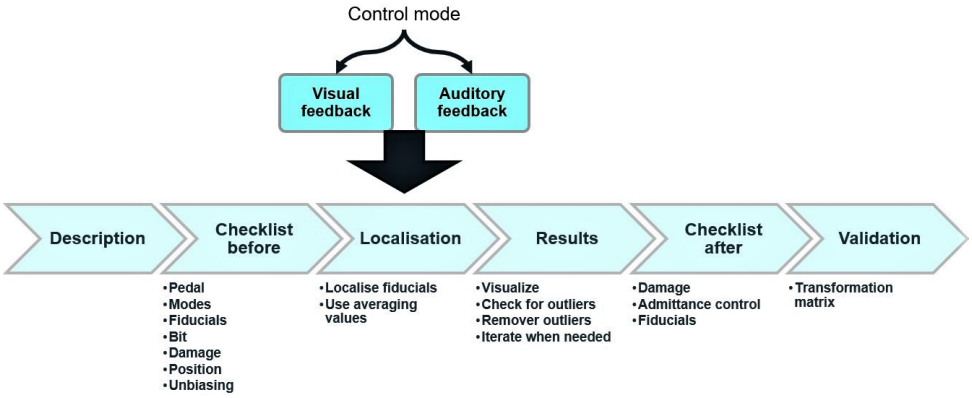


Figure 5.12: A schematic outline of the multi-feedback interface consisting of a step-by-step GUI, visual and auditory feedback. The GUI is divided into six steps: description, checklist before, localisation, results, checklist after and validation. The description step explains registration and its action steps. Checklist before consists of checks needed to ensure the safe and accurate use of robot localisation. Localisation is the process of performing fiducial localisation by the operator. This is where visual and auditory feedback is provided. Checklist after entails checks needed to prevent damages or changes from going neglected. Finally, validation contains the calculation of the transformation matrix.

Provide information to the supervisor about saved data and the number of data-points
Set previous value to current value
end if

This is done with an auditory sound of vector

$$y = 5 * sin(2 * \pi * 440 * t) \tag{5.6}$$

with t being a vector of sampling times for a frequency F_s of

$$t = 0 : \frac{1}{F_s} : 0.1 \tag{5.7}$$

The sound assures the operator that the position is saved and thus limits timing errors.

Visual feedback

An LED containing four different colours was implemented, which supplies feedback during localisation. Table 5.5 summarises the visual feedback.

Table 5.5: Visual feedback provided by a LED placed near the operator and the colour-depending information it displays

LED color	Meaning
Blue	Translational or Rotational mode within safety margins
Green	Fixed or safety mode within safety margins
Red	Force, torque or workspace exceeding safety margins
Multi	Registration mode

The force and torque boundaries equal their safety thresholds, being 8.0N and 0.2N, respectively. The workspace threshold, although not limited, is indicated by a red LED colour according to the condition

if $|Pos_{error}| > 20$ **then**
LED is red
end if

Due to the 0.2 value gain inside the $|Pos_{error}|$ calculation, a positional error of 100mm initiates a red LED. Figure 5.13 delivers an example of the LED placement. Information is provided to the operator regarding its performance, for instance, directly perceiving excessive force or torque. It also reveals force bias and provides the operator and supervisor more awareness concerning the force readings.

This allows the operator to correct where needed. Further, the operator must know the activated mode and the correct way to interact with the system.

Figure 5.13: An example of the LED placement and visual feedback to the operator. Here the colour is green, meaning the robot is in fixed or safety mode.

Training, roles and responsibilities
In order to restrain unsafe acts of operators, training and clearly knowing the role and responsibility of each human is important. It was concluded to include two human operators in the registration workflow: the operator and the supervisor. The operator performs registration by him or herself to prevent communication errors, while the supervisor keeps track of the process and prevent the large dependency on one human operator. As mentioned, these implementations are not defined as design functionalities, rather as additional documentation and processes to ensure correct usage of the design. Therefore, the details will be provided in Appendix C. For the training it is assumed that the operator and supervisor are only involved with the registration procedure and not with setting up the robot. However, the step by step procedure of setting up the robot can also be found in Appendix C. The training procedure is divided in multiple steps: (1) Explain procedure and general terminology, (2) Roles and responsibilities, (3) Walk through workflow, (4) Safety buttons, (5) Robot errors, and (6) Case-studies. For the old design the training, roles and responsibilities are described in Appendix C.

5.4. Conclusion

The design implementations were divided in three different categories: multi-modal admittance control; multi-feedback interface; and training, roles and responsibilities.

The multi-modal admittance enables the operator to move the robot through the sensed interaction force at the robot's end-effector with five modes of control: (1) translational, (2) rotational, (3) fixed, (4) safety and (5) registration mode. Each mode allows the operator to control the robot as needed to achieve high performance. To switch between modes a foot-operated interface with several buttons, each dedicated to a specific function is implemented.

The multi-feedback interface supplies visual and auditory feedback, and a step-by-step guide to the operator and supervisor without overloading one sensory modality. This was used to build operator awareness, control large errors introduced by operator input and prevent timing errors. The GUI contains safety checklists, workflow guidance and operator performance feedback and was minimises operator variations.

Lastly, the training, roles and responsibilities increases the knowledge of the operator and supervisor to prevent human-related errors and failures from occurring during registration.

Thus these three implementation categories together provide the operator and supervisor with the ability to perform highly accurate and safe registration.

6. Testing: Results

This chapter shows the yielded results from the benchmarking quasi-experiment.

The results from the benchmarking quasi-experiment discussed in Chapter 2 are summarised in Figure 6.1. Each metric is presented in detail.

Accuracy			
Metric	Median	IQR	Max
FLE (mm)			
New design	0.0550	0.0635	0.7805
Old design	0.8387	0.1672	2.5449
Difference	**-93.4%**	**-62.0%**	**-69.3%**
TRE (mm)			
New design	0.1677	0.0575	0.8783
Old design	0.9544	0.2256	2.3131
Difference	**-82.4%**	**-74.5%**	**-62.0%**

Workload			
Metric	Median	IQR	Max
NASA TLX 0-100			
New design	35.33	25.83	65.33
Old design	35.15	38.33	66
Difference	**0.5%**	**-32.6%**	**-1.0%**
Avg Force (N)			
New design	5.19E+03	1.65E+03	7.62E+03
Old design	1.21E+04	3.38E+03	1.51E+04
Difference	**-57.2%**	**-51.0%**	**-49.3%**

Usability			
Metric	Median	IQR	Max
SUS 0-100			
New design	85.00	6.25	95.00
Old design	75.00	23.75	92.50
Difference	**13.3%**	**-73.7%**	**2.7%**
CSUQ 0-7			
New design	6.20	0.97	7.00
Old design	5.30	1.15	7.00
Difference	**17.0%**	**-15.3%**	**0.0%**

Trust			
Metric	Median	IQR	Max
Trust 0-100			
New design	66.50	20.31	95.00
Old design	56.50	18.69	92.50
Difference	**17.7%**	**8.7%**	**2.7%**

Figure 6.1: Summary of the key results uncovered from the benchmark tests. Each metric shows its median, inter-quantile range (IQR) and maximum value for both the new and old design. The difference provides a percentage change between the new and old design, from new-old/old*100%

6.1. Accuracy

The human localisation performance is measured with FLE and TRE, yielding the result in Figure 6.2. Additionally, general system accuracy is provided by measuring the difference in FLE for different heading angles, repeated localisation by a skilled operator and a hold test.

Accuracy			
Metric	Median	IQR	Max
FLE (mm)			
New design	0.0550	0.0635	0.7805
Old design	0.8387	0.1672	2.5449
Difference	**-93.4%**	**-62.0%**	**-69.3%**
TRE (mm)			
New design	0.1677	0.0575	0.8783
Old design	0.9544	0.2256	2.3131
Difference	**-82.4%**	**-74.5%**	**-62.0%**

Figure 6.2: Summary of the accuracy results, FLE and TRE, uncovered from the benchmark tests. Each metric shows its median, inter-quantile range (IQR) and maximum value for both the new and old design. The difference provides a percentage change between the new and old design, from new-old/old*100%

6.1.1. FLE

FLE was calculated by clustering their data with kmeans clustering for four clusters and using Equations 2.1, resulting in the boxplot of Figure 6.3 and Table 6.1. These depict a decrease in median FLE of 93.4%, FLE IQR of 62.0% and Max FLE of 69.3%. A Kolmogorov-Smirnov test concludes no normal distributed data, as shown in Figure 6.4. Performing the one-sided Wilcoxon rank sum test was conducted and resulted in a p-value of 0, thus concluding statistical significance.

It also looks at the difference in FLE value per fiducial and in different directions with respect to the block or robot base, resulting in Figure 6.5. These differences per direction are further separated per fiducial, shown in Figure 6.6 and Table 6.2.

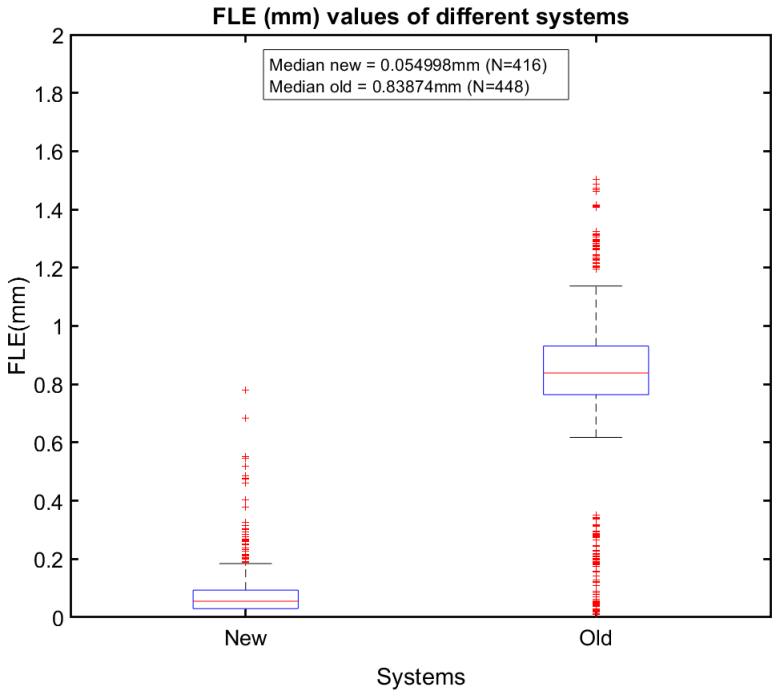


Figure 6.3: Two boxplots showing the distribution of FLE values in mm over all participants for the new and old design, calculated from Equations 2.1 and 2.2. The x-axis shows the different systems. The y-axis provides the FLE value, ranging from 0 to 2.0mm. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers. The medians are calculated as 0.0550mm (N=416) and 0.839mm (N=448) for the new and old systems, respectively, with $P_1 = 13$ and $P_2 = 14$ the number of participants for the new and old system.

Table 6.1: The FLE value compared between new and old systems. The FLE is compared based on its maximum (Max), median (Med), minimum (Min), IQR, the number of outliers, the number of data points (Data) and participants (P).

System	Max (mm)	Med (mm)	Min (mm)	IQR (mm)	Outliers	Data	P
New	0.7805	0.0550	0.0028	0.0635	40	416	13
Old	2.5449	0.8387	0.0097	0.1672	149	448	14

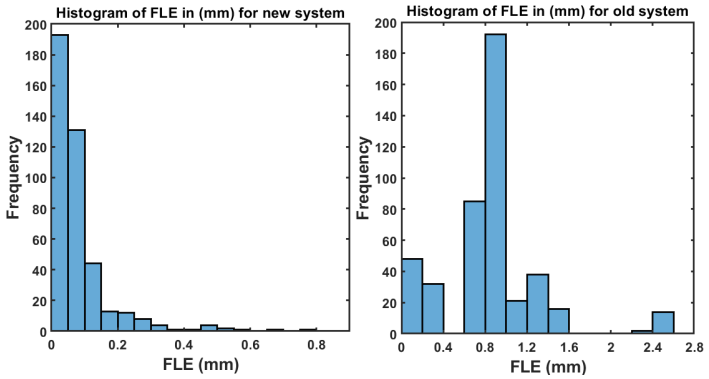


Figure 6.4: A histogram of FLE benchmark values, representing the frequency distribution of the FLE in mm for each group. The x-axis shows the FLE value in mm. The x-axis ranges from 0 to 1, with bins of 0.05 for the new system and from 0 to 2.8, with bins of 0.2 for the old system. The y-axis shows the related frequency, ranging from 0 to 200. The new and old systems consist of 13 and 14 participants, respectively.

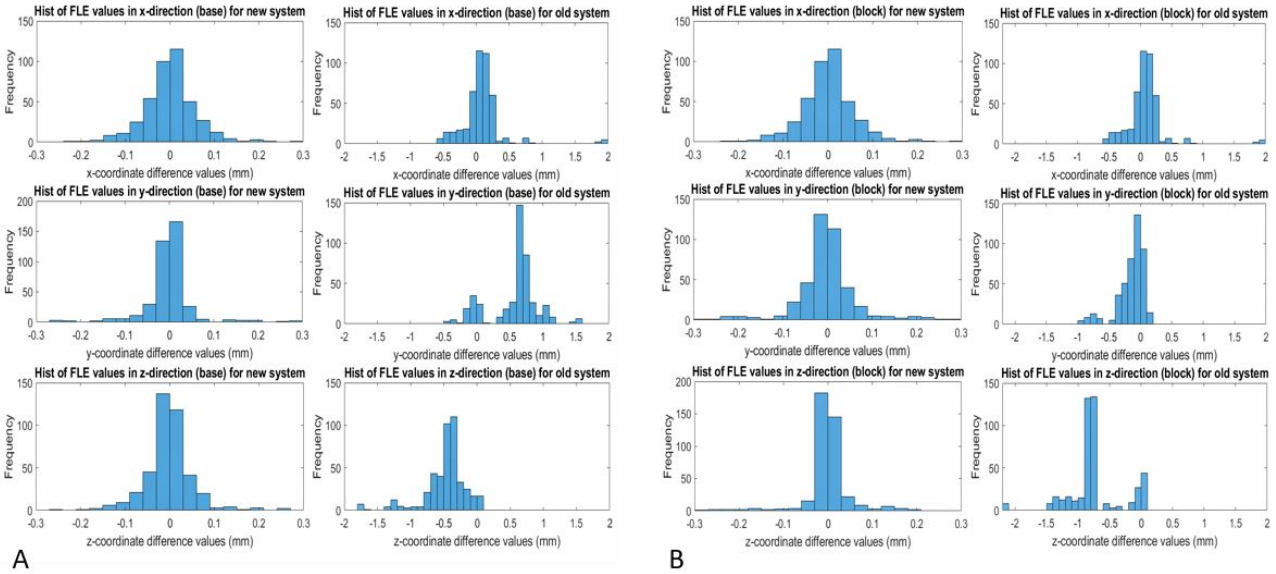


Figure 6.5: A) shows a histogram of FLE benchmark values split in x-, y- or z-direction with respect to the robot base, representing the frequency distribution of the FLE in mm for each group. The left side depicts the new system, and the right side the old system, where each row relates to the different coordinate directions x-, y- or z-coordinate, respectively. The x-axis shows the FLE value in mm along a certain direction. The x-axis ranges from -0.3 to 0.3 for the new system with bins of 0.3 and -2.0 to 2.0 for the old system with bins of 0.1. The y-axis shows the related frequency, ranging mostly from 0 to 150. The y-axis for the second row of the new system ranges from 0 to 200. B) shows a histogram of FLE benchmark values split in x-, y- or z-direction with respect to the block, representing the frequency distribution of the FLE in mm for each group. The left side depicts the new system, and the right side the old system, where each row relates to the different coordinate directions x-, y- or z-coordinate, respectively. The x-axis and y-axis contain the same information and ranges as A.

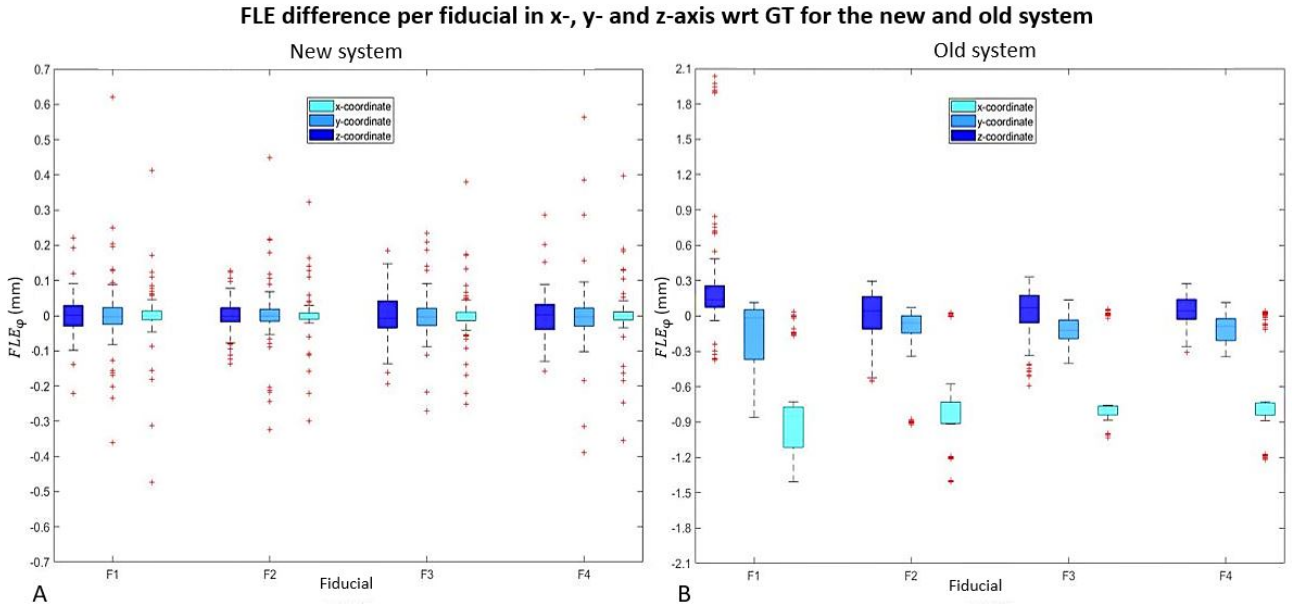


Figure 6.6: Multiple boxplots showing the distribution of FLE values per fiducial in each direction, either x, y or z, in mm over all participants for the new (A) and old (B) design. The x-axis shows the different fiducials, colour-coded per direction x,y or z from light to dark blue. The y-axis provides the FLE value in a certain direction indicated by ϕ, ranging from -0.7 to 0.7mm for the new system and -2.1 to 2.1 for the old system. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers.

Table 6.2: Difference in various FLE values of specific fiducials compared to the average value from all fiducials. The FLE is compared based on its maximum (Max), median (Med), minimum (Min), IQR, and the number of outliers. For example, the max FLE value for Fiducial 1 of the new design, 0.157mm, indicates that Fiducial 1 scored 0.157 higher than the average max FLE value of the four fiducials for the new system.

Differences	Max (mm)	Med (mm)	Min (mm)	IQR (mm)	Outliers
F1 new	0.157	-0.002	-0.001	-0.006	4.25
F1 old	0.984	0.049	0.009	0.203	8.25
F2 new	-0.078	-0.016	-0.003	0.001	1.25
F2 old	-0.144	-0.025	-0.007	0.169	-7.75
F3 new	-0.139	0.008	0.003	-0.002	-1.75
F3 old	-0.509	-0.011	-0.006	-0.185	8.25
F4 new	0.060	0.011	0.000	0.007	-3.75
F4 old	-0.331	-0.013	0.003	-0.187	-8.75

6.1.2. TRE

TRE was calculated according to the steps in Figure 2.14, resulting in the boxplot of Figure 6.7 and Table 6.3. These depict a decrease in median TRE of 82.4%, TRE IQR of 74.5% and Max TRE of 62.0%. A Kolmogorov-Smirnov test concludes no normal distributed data, as shown in Figure 6.8. Performing the two-sided Wilcoxon rank sum test was conducted and resulted in a p-value of 0, thus concluding statistical significance.

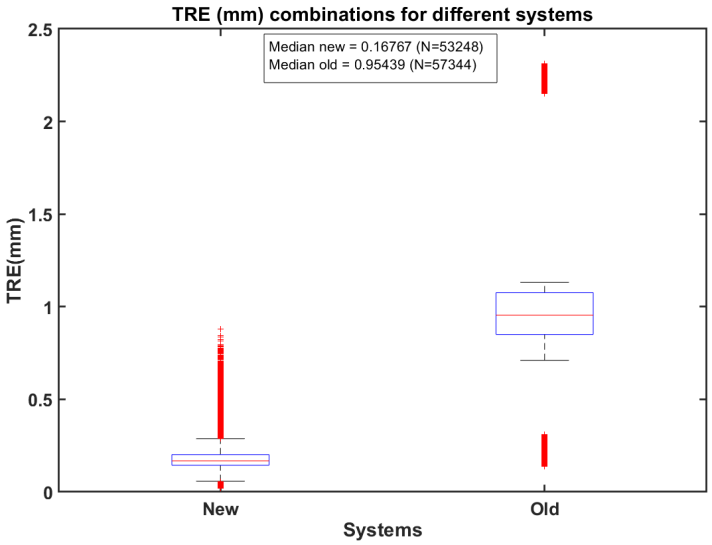


Figure 6.7: Two boxplots showing the distribution of TRE values in mm over all participants for the new and old design, calculated from the steps in Figure 2.14 based on 8 to the power of 4 different combinations. The x-axis shows the different systems. The y-axis provides the TRE value, ranging from 0 to 2.5mm. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers. The medians are calculated as 0.168mm (N=53248) and 0.0.954mm (N=57344) for the new and old systems, respectively, with $P_1 = 13$ and $P_2 = 14$ the number of participants for the new and old system.

Table 6.3: The TRE value compared between new and old systems. The TRE is compared based on its maximum (Max), median (Med), minimum (Min), IQR, the number of outliers, the number of data points (Data) and participants (P).

System	Max (mm)	Med (mm)	Min (mm)	IQR (mm)	Outliers	Data	P
New	0.8783	0.1677	0.0203	0.0575	4672	53248	13
Old	2.3131	0.9544	0.1385	0.2256	12288	57344	14

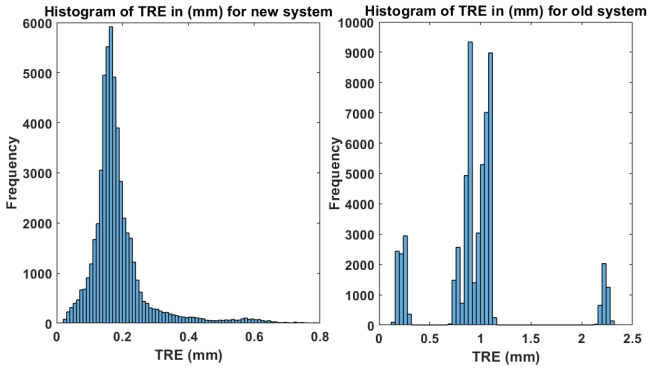


Figure 6.8: A histogram of TRE benchmark values, representing the frequency distribution of the TRE in mm for each group. The x-axis shows the TRE value in mm. The x-axis ranges from 0 to 0.8, with bins of 0.001 for the new system and from 0 to 2.5, with bins of 0.04 for the old system. The y-axis shows the related frequency, ranging from 0 to 6000 for the new system and 0 to 10000 for the old system. The new and old systems consist of 13 and 14 participants, respectively.

6.1.3. System accuracy

Additional research was conducted on the influence of the heading angle on the FLE, repeated localisation by a skilled operator, and holding the end-effector at the same position for a prolonged time.

Angle influence

The influence of the heading angle on the FLE with respect to the GT was tested by rotating the standard heading angle [0 0.8829 -0.4695] with -5, 5, -10 or 10 degrees around either the x-, y- or z-axis of the robot-frame. Each combination was measured 40 times per fiducial in one day; thus, 160 total data points and compared to the 56 measurements per fiducial for the standard angle. These results are summarised in Figure 6.10 and 6.9. More details for this are presented in Appendix N.

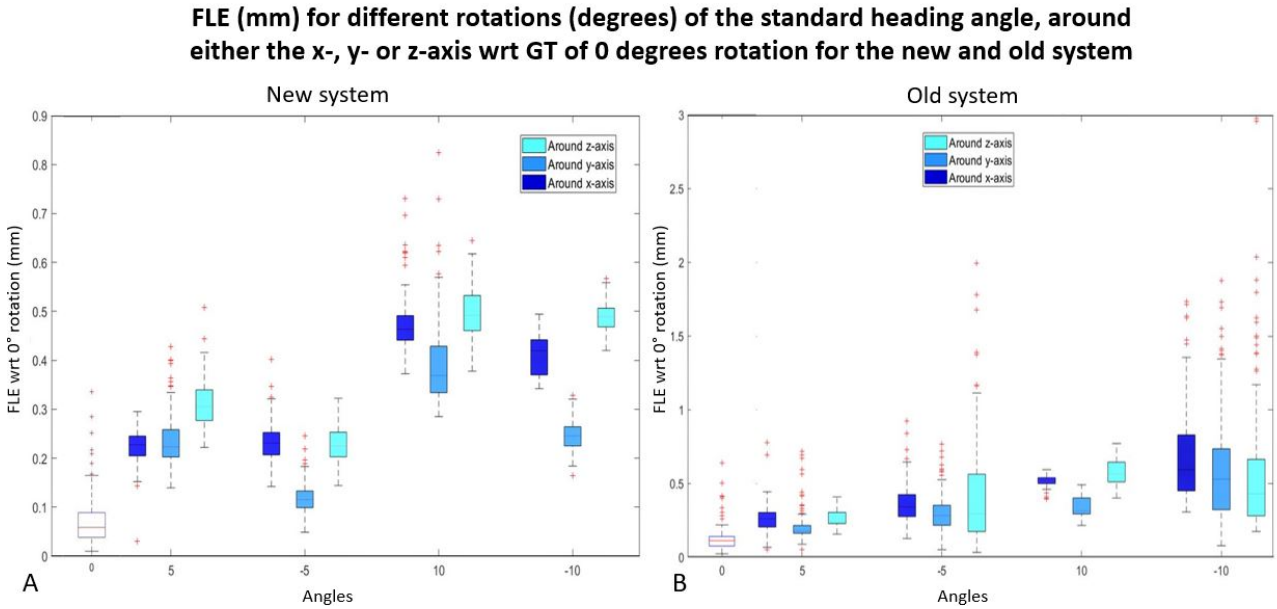


Figure 6.9: Multiple boxplots showing the distribution of FLE values per rotation angle, 5, -5, 10 or -10, of the standard heading angle around either the x-, y- or z-axis with respect to the GT 0 degrees fiducial positions for the new (A) and old (B) design. The x-axis shows the different rotation angles, colour-coded per direction x,y or z from dark to light blue. The y-axis provides the FLE vector value, ranging from 0 to 0.9mm for the new system and 0 to 3.0 for the old system. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers.

Average change due to rotation					
|Angle|	Max	Median	Min	IQR	Outliers
New design	mm	mm	mm	mm	#
5	0.0435	0.1519	0.1118	-0.0025	-4.1667
10	0.2622	0.3549	0.2975	0.0104	-4.1667
Old design	mm	mm	mm	mm	#
5	0.4995	0.1265	0.0549	0.0864	3.0000
10	0.7711	0.3813	0.2165	0.1791	1.8333
Difference 5	**-91.3%**	**20.0%**	**103.6%**	**-102.9%**	**-238.9%**
Difference 10	**-66.0%**	**-6.9%**	**37.4%**	**-94.2%**	**-327.3%**

Figure 6.10: Summary of the key results uncovered from the angle tests where different rotations of the standard heading angle, either -5, 5, -10, or 10 degrees around the x-, y- or z-axis for the old and new design. This Table shows the difference in average value with respect to the 0 degrees rotation GT results. The positive and negative rotations around the different axis are summed for 5, and 10 degrees and their average value is shown. For example, the average max FLE value of rotations rotating either -5 or 5 degrees around the x-, y- and z-axis are summed and show on average that these are 0.0435mm higher compared to the max FLE from zero rotation. The difference provides a percentage change between the new and old design, from new-old/old*100%

Repeated localisation skilled operator

FLE was calculated for a skilled operator over multiple days, resulting in the boxplot of Figure 6.11 and Table 6.4. Performing the one-sided Wilcoxon rank sum test was conducted and resulted in a p-value of 0, thus concluding statistical significance.

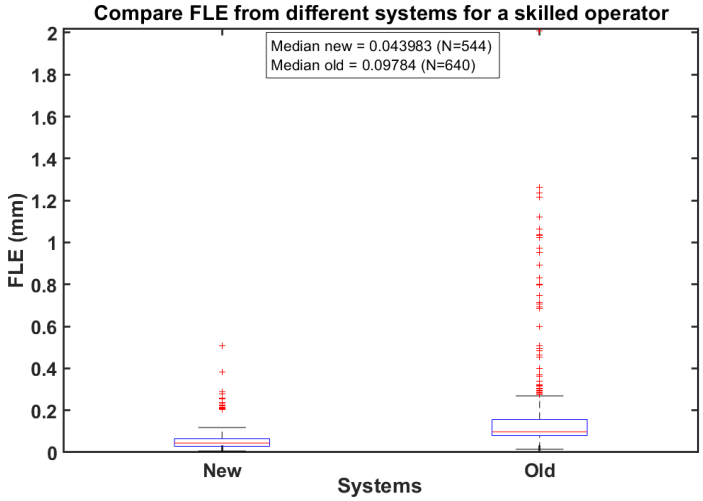


Figure 6.11: Two boxplots showing the distribution of FLE values in mm for a skilled operator for the new and old design, calculated from from Equations 2.1 and 2.2. The x-axis shows the different systems. The y-axis provides the FLE value, ranging from 0 to 2.0mm. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers. The medians are calculated as 0.0440mm (N=544) and 0.0978mm (N=640) for the new and old systems, respectively.

Table 6.4: The repeated skilled operator FLE value compared between new and old systems. The FLE is compared based on its maximum (Max), median (Med), minimum (Min), IQR, the number of outliers, and the number of data points (Data).

System	Max (mm)	Med (mm)	Min (mm)	IQR (mm)	Outliers	Data
New	0.5059	0.044	0.0063	0.036	19	544
Old	2.0091	0.0978	0.0148	0.0775	73	640
Difference	-74.82%	-55.01%	-57.43%	-53.55%	-73.97%	-15.00%

6.2. Human Workload

The human workload, measured with NASA TLX and average force, yields the results in Figure 6.12.

Workload			
Metric	Median	IQR	Max
NASA TLX 0-100			
New design	35.33	25.83	65.33
Old design	35.15	38.33	66
Improvement	**0.5%**	**-32.6%**	**-1.0%**
Avg Force (N)			
New design	5.19E+03	1.65E+03	7.62E+03
Old design	1.21E+04	3.38E+03	1.51E+04
Improvement	**-57.2%**	**-51.0%**	**-49.3%**

Figure 6.12: Summary of the workload results, NASA TLX and average force, uncovered from the benchmark tests. Each metric shows its median, inter-quantile range (IQR) and maximum value for both the new and old design. The difference provides a percentage change between the new and old design, from new-old/old*100%

6.2.1. Qualitative measure: NASA TLX

NASA TLX was calculated according to Equation 2.3 with their weights, resulting in the boxplot of Figure 6.13 and Table 6.5. These depict an increase in median NASA TLX of 0.5% and a decrease in NASA TLX IQR of 32.6% and Max NASA TLX of 1.0%. Using a Kolmogorov-Smirnov test concludes no normal distributed data as shown in Figure 6.14. Performing the one-sided Wilcoxon rank sum test was conducted and resulted in a p-value of 0.33623, thus concluding no statistical significance.

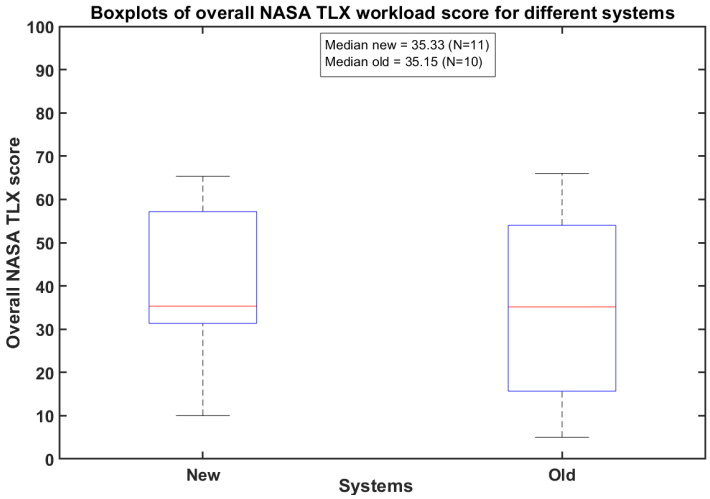


Figure 6.13: Two boxplots showing the distribution of NASA TLX values over all participants for the new and old design, calculated from Equation 2.3 with their weights. The x-axis shows the different systems. The y-axis provides the NASA TLX value, ranging from 0 to 100. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers. The medians are 35.33 (N=11) and 35.15 (N=10) for the new and old systems, respectively, with P_1 = 13 and P_2 = 14 the number of participants for the new and old system.

Table 6.5: The median NASA TLX values for underlying workload types for the new and old systems. MD = Mental Demand, PD = Physical Demand, TD = Temporal Demand, P = Performance, E = Effort, F = Frustration, Num p = Number of participants.

System	Med(MD)	Med(PD)	Med(TD)	Med(P)	Med(E)	Med(F)	Num p
New	40	30	40	15	30	15	11
Old	27.5	40	22.5	27.5	27.5	22.5	10

Table 6.6: The NASA TLX value compared between new and old systems. The NASA TLX is compared based on its maximum (Max), median (Med), minimum (Min), IQR, the number of outliers, the number of data points (Data) and participants (P).

System	Max	Med	Min	IQR	Outliers	Data	P
New	65.33	35.33	10	25.83	0	11	13
Old	66	35.15	5	38.33	0	10	14

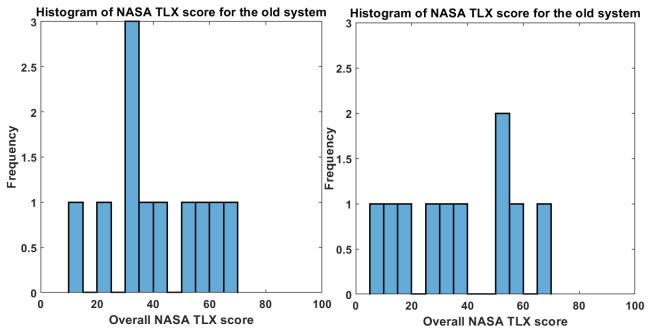


Figure 6.14: A histogram of NASA TLX benchmark values, representing the frequency distribution of the NASA TLX for each group. The x-axis shows the NASA TLX value in mm. The x-axis ranges from 0 to 100, with bins of 0.5 for both the new and old systems. The y-axis shows the related frequency, ranging from 0 to 3. The new and old systems consist of 13 and 14 participants, respectively.

6.2.2. Quantitative measure: Average force

The average force was calculated according to Equation 2.4, resulting in the boxplot of Figure 6.15 and Table 6.7. These depict a decrease in the median average force of 57.2%, the IQR average force of 51.0% and the max average force of 49.3%. A Kolmogorov-Smirnov test concludes no normal distributed data. Performing the one-sided Wilcoxon rank sum test was conducted and resulted in a p-value of 5.6255e-06, thus concluding statistical significance. Other metrics of workload: work (Total force*total time), Total force, time, and Force/time are summarised in Table 6.8.

Table 6.7: The average force value compared between new and old systems. The average force is compared based on its maximum (Max), median (Med), minimum (Min), IQR, the number of outliers, the number of data points (Data) and participants (P).

System	Max	Med	Min	IQR	Outliers	Data	P
New	7.6231e+03	5.1902e+03	3.8438e+03	1.6548e+03	0	13	13
Old	1.5050e+04	1.2135e+03	8.0476e+03	3.3755e+03	0	14	14

Table 6.8: Median values of different workload metrics: work (total force*total time), total force, time, and average force for different systems.

System	Work	Total force	Time	Avg Force
New	5.470e+08 Ns	1.587e+06 (N)	333.1 (s)	5.190e+03 (N/s)
Old	5.395e+08 Ns	2.605e+06 (N)	200.8 (s)	1.214e+04 (N/s)

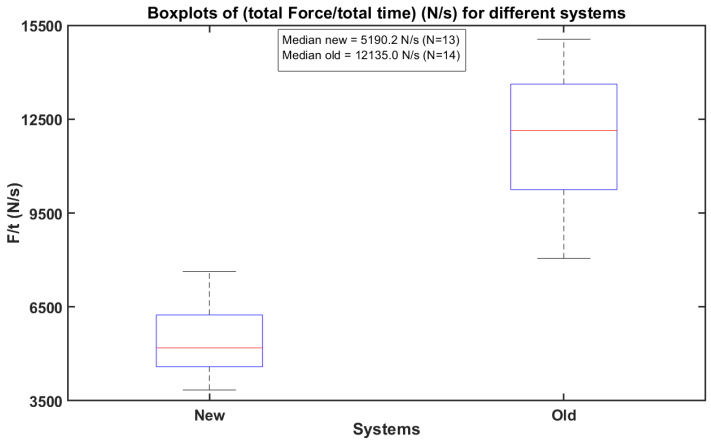


Figure 6.15: Two boxplots showing the distribution of average force values in N/s over all participants for the new and old design, calculated from Equation 2.4. The x-axis shows the different systems. The y-axis provides the average force value, ranging from 0 to 15500N/s. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers. The medians are 5190.2 (N/s) (N=13) and 12135.0 (N/s) (N=14) for the new and old systems, respectively.

6.3. Usability

The usability is measured with SUS and CSUQ, yielding the results in Figure 6.16.

Usability			
Metric	Median	IQR	Max
SUS 0-100			
New design	85.00	6.25	95.00
Old design	75.00	23.75	92.50
Difference	**13.3%**	**-73.7%**	**2.7%**
CSUQ 0-7			
New design	6.20	0.97	7.00
Old design	5.30	1.15	7.00
Difference	**17.0%**	**-15.3%**	**0.0%**

Figure 6.16: Summary of the usability results, SUS and CSUQ, uncovered from the benchmark tests. Each metric shows its median, inter-quantile range (IQR) and maximum value for both the new and old design. The difference provides a percentage change between the new and old design, from new-old/old*100%

6.3.1. SUS

SUS was calculated according to Equation 2.5, resulting in the boxplot of Figure 6.17 and Table 6.9. These depict an increase in median SUS of 13.3% and max SUS of 2.7%, and a decrease in SUS IQR of 73.7%. A Kolmogorov-Smirnov test concludes no normal distributed data, as shown in Figure 6.18. Performing the one-sided Wilcoxon rank sum test was conducted and resulted in a p-value of 0.0417, thus concluding statistical significance.

Table 6.9: The SUS value compared between new and old systems. The SUS is compared based on its maximum (Max), median (Med), minimum (Min), IQR, the number of outliers, the number of data points (Data) and participants (P).

System	Max	Med	Min	IQR	Outliers	Data	P
New	95	85	70	6.25	0	13	13
Old	92.50	75	47.50	23.75	0	13	14

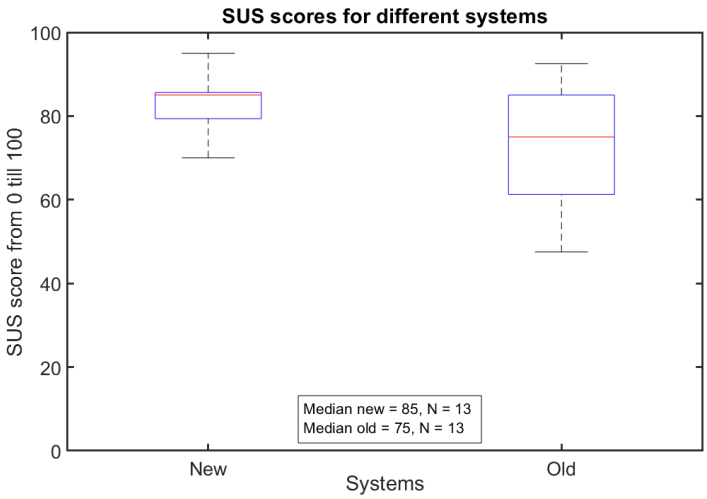


Figure 6.17: Two boxplots showing the distribution of SUS values over all participants for the new and old design, calculated from Equation 2.5. The x-axis shows the different systems. The y-axis provides the SUS value, ranging from 0 to 100. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers. The medians are 85 (N=13) and 75 (N=13) for the new and old systems, respectively, with P_1 = 13 and P_2 = 14 the number of participants for the new and old system.

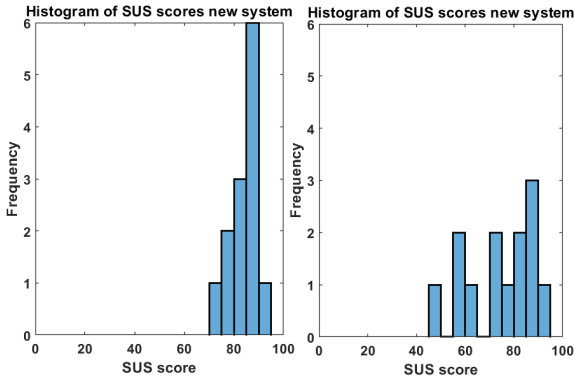


Figure 6.18: A histogram of SUS benchmark values representing the frequency distribution of the SUS for each group. The x-axis shows the SUS value. The x-axis ranges from 0 to 100, with bins of 5 for both systems. The y-axis shows the related frequency, ranging from 0 to 6 for both systems. The new and old systems consist of 13 and 14 participants, respectively.

6.3.2. CSUQ

CSUQ was calculated according to Equation 2.3, resulting in the boxplot of Figure 6.19 and Table 6.10. These depict an increase in median CSUQ of 17.0%, a decrease in CSUQ IQR of 15.3% and no difference in Max CSUQ. A Kolmogorov-Smirnov test concludes no normal distributed data, as shown in Figure 6.20. Performing the one-sided Wilcoxon rank sum test was conducted and resulted in a p-value of 0.0015747, thus concluding statistical significance.

Table 6.10: The CSUQ value compared between new and old systems. The CSUQ compared based on its maximum (Max), median (Med), minimum (Min), IQR, the number of outliers, the number of data points (Data) and participants (P).

System	Max	Med	Min	IQR	Outliers	Data	P
New	7.0	6.2	5.6	0.972	0	13	13
Old	7.0	5.3	4.0	1.147	0	13	14

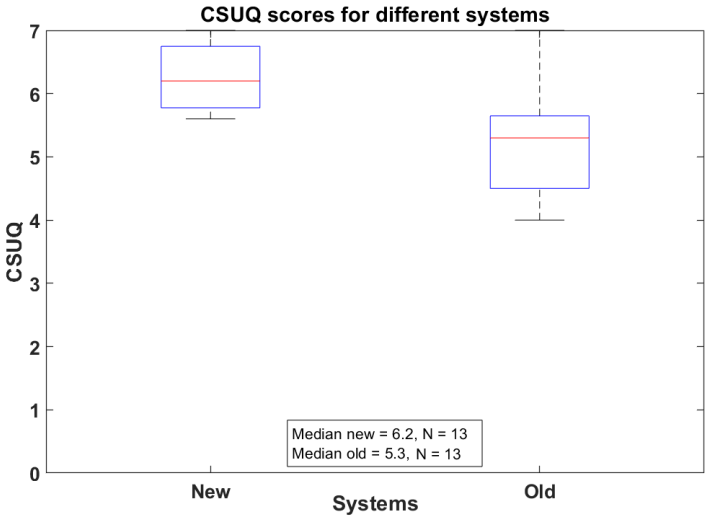


Figure 6.19: Two boxplots showing the distribution of CSUQ values over all participants for the new and old design, calculated from Equation 2.3,. The x-axis shows the different systems. The y-axis provides the CSUQ value, ranging from 0 to 7. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers. The medians are 6.2 (N=13) and 5.3 (N=14) for the new and old systems, respectively.

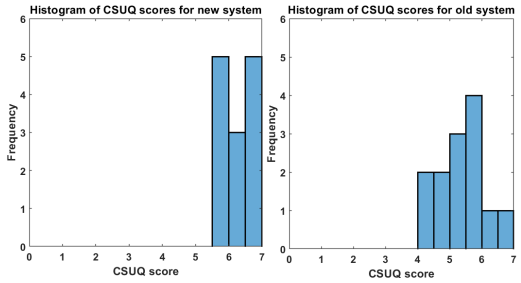


Figure 6.20: A histogram of CSUQ benchmark values, representing the frequency distribution of the CSUQ for each group. The x-axis shows the CSUQ value. The x-axis ranges from 0 to 7, with bins of 0.5 for both systems. The y-axis shows the related frequency, ranging from 0 to 6 for both systems. The new and old systems consist of 13 and 14 participants, respectively.

6.4. Trust

Trust was calculated with the Trust Perception Scale-HRI according to Equation 2.3, resulting in the boxplot of Figure 6.22 and Table 6.11. These depict an increase in median Trust of 17.7%, Trust IQR of 8.7% and Max Trust of 2.7%, summarised in Figure 6.21. A Kolmogorov-Smirnov test concludes no normal distributed data, as shown in Figure 6.23. Performing the one-sided Wilcoxon rank sum test was conducted and resulted in a p-value of 0.068484, thus not concluding statistical significance.

Table 6.11: The Trust Perception Scale-HRI value compared between new and old systems. The Trust Perception Scale-HRI is compared based on its maximum (Max), median (Med), minimum (Min), IQR, the number of outliers, the number of data points (Data) and participants (P).

System	Max	Med	Min	IQR	Outliers	Data	P
New	92.25	66.50	46.00	20.31	0	13	13
Old	90.50	56.50	37.50	18.69	0	13	14

Trust			
Metric	Median	IQR	Max
Trust 0-100			
New design	66.50	20.31	95.00
Old design	56.50	18.69	92.50
Improvement	**17.7%**	**8.7%**	**2.7%**

Figure 6.21: Summary of the Trust Perception Scale-HRI results, uncovered from the benchmark tests. Each metric shows its median, inter-quantile range (IQR) and maximum value for both the new and old design. The difference provides a percentage change between the new and old design, from new-old/old*100%

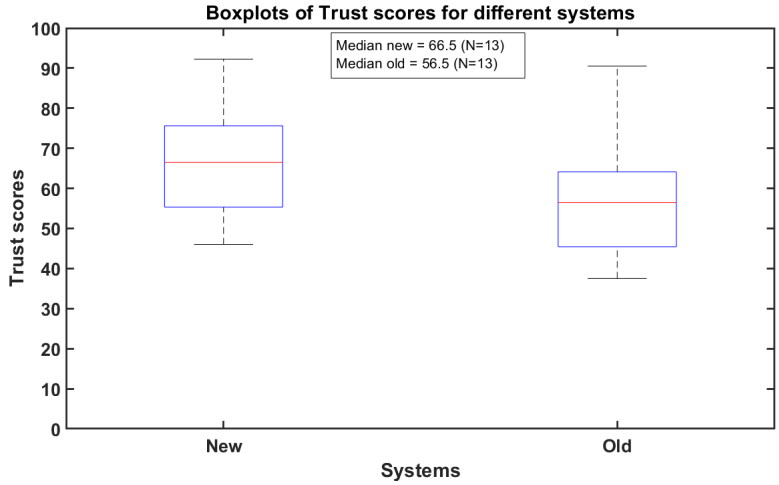


Figure 6.22: Two boxplots showing the distribution of Trust Perception Scale-HRI values in % over all participants for the new and old design, calculated from Equation 2.3. The x-axis shows the different systems. The y-axis provides the Trust Perception Scale-HRI value, ranging from 0 to 100. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers. The medians are 66.5 (N=13) and 56.5 (N=13) for the new and old systems, respectively.

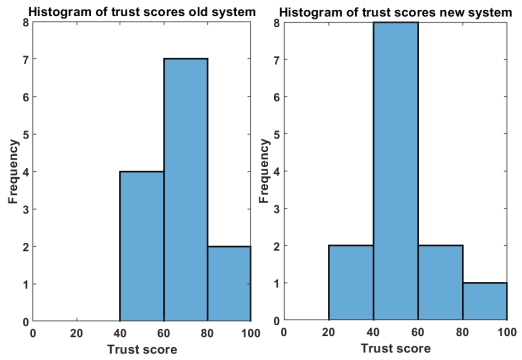


Figure 6.23: A histogram of Trust Perception Scale-HRI benchmark values, representing the frequency distribution of the Trust Perception Scale-HRI for each group. The x-axis shows the Trust Perception Scale-HRI value. The x-axis ranges from 0 to 100, with bins of 20 for both systems. The y-axis shows the related frequency, ranging from 0 to 8 for both systems. Both systems consist of 13 participants, respectively.

7. Discussion

This chapter reflects on the yielded results and design and provides additional recommendations regarding the results, design, CT localisation and CI surgery.

7.1. Results

This chapter first analyses the results found in 6 and examines striking outcomes. Analysing Figure 6.1 shows a clear negative difference in results for accuracy, meaning an increase in accuracy. The workload results indicate a clear negative difference for the average force and the IQR of NASA TLX and no significant results for the NASA TLX median and max values. The usability results reveal a clear negative difference for the SUS and CSUQ median and IQR values and no significant difference for their max values. The trust values show a positive difference for the median value and IQR and an insignificant difference in the maximum value. Thus answering Q2, the hypothesis H_0 = **Multi-modal multi-feedback admittance control will improve the human-robot (system) performance measured by accuracy, workload, usability and trust** is not accepted. However, H_a = **Multi-modal multi-feedback admittance control will improve the human-robot (system) performance measured by accuracy, workload and usability** is accepted. Figure 7.1 summarises the other hypothesis and results. The results show all results to be non-normally distributed. This is expected due to maximum, minimum and limiting scales.

Hypothesis						
Metric	H_1	H_2	H_3	H_4	H_0	H_a
Accepted/rejected						

Figure 7.1: Summary of the different hypotheses. Green means they are accepted, and red equals rejection.

Other topics of the results are summarised in Figure 7.2 and given in more detail below.

Discussion	
Results	Meaning
General	
Non clinical setting	Metrics can differ significantly in clinical settings and even in phantom studies. Therefore it is suggested to repeat the study in a clinical environment
Small and bias sample	The study entails a relatively small and biased participant sample. Groups are compared thus difference still holds. It is recommended to include medical related participant.
Situation specific	All data is situation specific and dependant on the configuration and experimental set-up. The results will differ for a slightly different set-up and for a different TP.
Acceptance	As many different factors influence acceptance, no conclusions can be drawn
Non normal distributed data	The results show all results to be non-normally distributed. This is expected due to maximum, minimum and limiting scales
HRI metrics connection	Since HRI metrics are highly dependant on each other, it is difficult to state conclusions. More research on their dependancies is
Accuracy	
Unskilled operators	Good results new system, despite highly skilled operator. Training has less influence compared to old system. The worst-case scenario does decrease for both systems, thus training is still important
Controllable variables	The non-kinematic errors are controlled well as results are gathered over multiple days and no relationship is found between different days and larger errors. Moreover, adverse environmental factors are not tested, which is recommended
Repeatability and reproducibility	The significant decrease in IQR indicates higher repeatability and reproducibility for the new system this also relates to less training needed for the new system
Gap data points old system	This result occurs most likely due to the spring-like behaviour of the design, which requires operators to fight against the robot's movement.
Directionality dependance	The FLE value differs most in the x-direction, than y-direction and least in z-direction. Many operators use the same localisation technique, using the bottom of the screw edge and the z-direction through haptic feedback to perform localisation
Angle influence	The rotation of the heading angle largely influences the median FLE value
Systematic errors	The FLE is calculated based on one operator GT value. Nevertheless, if one participant performs localisation outside the fiducial position, this will not be noted directly in this project
CT localisation	The TRE values do not include the imaging localisation results, and are therefore significantly smaller
Validation	One recommended validation method is using multiple known targets and evaluating if their distances compare with the known distances in the CT image
Workload	
Not usable data	Some participants did not fill out the NASA TLX survey correctly, and too much time was passed to re-fill the scores after discovery.
More similarity	The IQR decreased significantly for the new system, indicating the participants have more similar experiences, likely due to the feedback provided to the operators
Time increase	The auditory feedback delay and checklists caused an increase in time to perform registration
Mental and temporal demand	The new system required much more training and information than the old one and provided more feedback that the operators had to pay attention to and respond to. This explains the increase in mental and temporal demand scores, which are expected to decrease when more training is provided
NASA TLX	Difficult to conclude whether the workload is high or low, but in comparison, it can be concluded that one system requires either a higher or lower workload
Usability	
Participants	One participant did not fill in the survey, providing the data for 13 participants in each group
More similarity	Lower IQR is due to the provided feedback that results in a more uniform experience for all participants operating on the new system
Participant pool	Previous experience with robots and computers can impact the scores
Trust	
Incorrect questionnaire	The standardised questionnaire related to HRI, including social and intelligent robots. Statements that relate to social and intelligence metrics,led to confusion for the participants
Individual characteristics	Trust depends on individual characteristics such as propensity to trust, self-confidence, individual differences and culture
Participant pool	Participant pool is biased
More research	More research is needed to understand the human trust in robots for human operators outside this project's scope and the relationships between trust and performance in HRI, especially over time
Medical staff	It is mostly recommended to include medical staff in this research.

Figure 7.2: Summary of the discussion on the thesis results

The most important notice is that none of the metrics described was measured in a non-clinical setting. Metrics can differ significantly in clinical settings and even in phantom studies. Therefore it is suggested to repeat the study in a clinical environment.

One general study limitation includes the relatively small participant sample. Due to time and other limitations, no other participants could participate in this study. Therefore, the significance of this experiment can be debated. Zheng et al. (2018) reports that a sample of five users can identify most usability problems, thus providing the assumption of a sufficiently large participant sample for usability. The number of data points is enlarged by repeated measurements, sufficing the data sample for accuracy. Moreover, the participant pool is highly biased towards employees or robotics students. This can influence the results, especially related to usability and trust. However, since it compared two groups inside the sample and the groups are consciously divided, it is expected to average out. It is recommended to perform more testing with participants outside this pool, preferably related to the medical field.

Acceptance is an essential variable for the success of HRI design; it was assumed that acceptance is directly related to system performance, human workload, trust and usability. However, other factors like innovative disruption influence acceptance; therefore, no conclusions can be drawn about this project's design acceptance. A radical technology is likely less readily accepted than more incremental automation surgery designs. Generally, HRI metrics are connected in a large network and conclusions on these individual metrics undermine their relations. Additional research is needed to understand better the complex relations between these HRI metrics and how to score them both subjectively and objectively.

Lastly, all data is situation-specific and dependent on the configuration and experimental setup. The

results will differ for a slightly different setup and a different TP.

7.1.1. Accuracy

According to Figures 6.2, 6.3, 6.7 and Tables 6.1 and 6.3, the FLE and TRE show significant negative differences between the new and old systems. Concluding a higher accuracy for the new system and accepting hypothesis H_1 = **Multi-modal multi-feedback admittance control will improve accuracy by decreasing the FLE and thus TRE**. Comparing the achieved results to STOTA 3.21 shows the new system outperforms most methods, except Gerber et al. (2013), in accuracy despite using unskilled operators and including the tracking error. Therefore, repeating the tests with a skilled operator, as mentioned in Section 6.1.3, yields better FLE accuracy results than any STOTA method. It also indicates that the variables to control are controlled well since these tests were executed over multiple days. Additionally, the worst-case scenario decreases significantly for a skilled operator for both systems. The skilled operator also indicates that training does influence the results and can improve the operator's performance. This training has significantly less influence on the new system than the old one. Moreover, the number of data points for the old system is substantially larger than the new one, which could lead to larger outliers. However, the ratio between outliers and data points for the old system indicates that the number of outliers is not solely due to the larger data.

Also related to the variables to be controlled is the experimental setup. The wooden block can be susceptible to non-kinematic errors such as temperature, dry air, and force exerted on it. However, the repeated measurements' results indicate that the difference between different days is most likely still sub-millimetric. The new and old systems' data were accumulated over five days. Even though the conditions were generally equal, some warm days (30+ degrees Celcius) were encountered during the testing phase, and this did not relate to large additional errors.

Also related to the environmental factors, this experiment does not consider adverse environmental factors such as poor sensing due to bad lighting, which can lead to incorrect responses (Vasic and Billard, 2013). However, in the OR, most likely, these environmental factors will not occur. Nevertheless, factors that could occur outside this project's scope are human factors such as fatigue, which could influence both the accuracy and the workload. The repeated localisation task does reflect slightly on boredom and fatigue and does provide some measure of contained focus over a longer time, but this still reflects approximately 2-3 minutes only. Thus it is recommended to research the influence of fatigue and boredom on localisation accuracy.

Besides, Table 6.3 shows that the maximum TRE still yields sub-millimetric accuracy for the new system. The significant decrease in IQR indicates higher repeatability and reproducibility for the new system. Additionally, the number of outliers for the new system is significantly lower, also indicating higher repeatability and reproducibility.

A gap is visible between the IQR and outliers for the old system. This result occurs most likely due to the spring-like behaviour of the design, which requires operators to fight against the robot's movement. A small resistance force towards the reference position can yield FLE and TRE values more extensive than the required sub-millimetric accuracy. Hence, acquiring high-accuracy results with this system is unlikely, and most participants will yield millimetric accuracy results, whereas timing and communication errors occur regularly, resulting in significant errors.

The non-normal distribution of the FLE and TRE for the new system, illustrated in the right-skewed data of Figures 6.4 and 6.8, is expected due to the data being close to the minimal limit. It indicates that the FLE values are close to the minimum value for the new system. The old system, in contrast, shows no clear shape with a tendency towards bell shape and multi-modal, indicating not reaching a

minimal limit.

The distribution of the FLE in x-, y- and z-direction with respect to the robot base plate or the fiducial block, as shown in Figures 6.5 and 6.6, show a symmetrical bell shape around 0.0 for the new system in all directions. For the fiducial block reference, the z-direction is less spread than the x- and y-direction, as expected, since operators can feel the z-direction limits through haptic feedback. The old system shows more randomly distributed data and does not always revolve around 0.0. The x-direction representing the top-to-bottom movement of the tip by the operator shows a random spread in old system data. For the y-direction representing the left-to-right movement, the old system data has a bias towards positive with respect to the robot base and negative with respect to the fiducial block. This result indicates that many operators use the same localisation technique, using the bottom of the screw edge to perform localisation.

The difference per fiducial in x-, y- and z-direction indicates that Fiducial 1 holds the largest FLE value and contributes most to the number of outliers for both systems; however, for the new system, it did not yield a higher median value. Fiducial 2 represents a lower maximum, median and minimal FLE value than the average, indicating an easy-to-locate fiducial. For fiducials 3 and 4, no significant relation between fiducials and results was found. Generally, it is concluded that the FLE distribution is biased and heterogeneous for the old system and unbiased and heterogeneous for the new system.

It is seen in Figures 6.9 and 6.10 that the rotation of the heading angle largely influences the median FLE value. This indicates a lack of a correct kinematic model for rotation. The influence of the angle on the new design is larger for the median value of 5 degrees and minimal FLE for 5 and 10 degrees. However, the new design limits the influence of the angle rotation on the max, IQR and number of outliers compared to the old design.

FLE and TRE are no measure of the total system accuracy as they exclude kinematics errors and only measure a relative error. This study, therefore, does not analyse systematic errors. Calculating the total system accuracy would require a high-accuracy tracking device.

In addition, the FLE is calculated based on one operator GT value. Nevertheless, if one participant performs localisation outside the fiducial position, this will not be noted directly in this project. Due to the design and researcher observing the task, such bias only has limited influence. It would be recommended to repeat the testing phase, only measuring the localisation accuracy and measuring multiple participants for the same system on the same day in sequence. Then their localisation accuracy can be compared to an average GT over all participants that day.

The TRE values are a simplification since they do not include an error for the image coordinate system, assuming that the microscope measurements of the base plate were exact. However, Perwög et al. (2018) states that using a microscope for localisation is likely to introduce more errors. Optical tracking systems tend to have more uncertainty along the axis perpendicular to the tracking system sensors. Thus these measurements are likely not exact, but their influence is disregarded in this thesis. Furthermore, the chosen PT influences TRE values, and others will give different results. It is recommended to repeat the tests for other important target positions to gain insight into the position relatedness.

One known problem for registration errors is the difficulty of validation, which is also the case in this project, where the PT is only an estimate, and the TRE assumes no error in the script that calculates the transformation matrix. J. Liu et al. (2019) states that validation primarily depends on clinician judgements and is not related to well-defined standard validations. Many error values given by dif-

ferent registration systems are not related to any actual target accuracy (Mongen and Willems, 2019). One recommended validation method is using multiple known targets and evaluating if their distances compare with the known distances in the CT image. However, the given accuracy is still considered an estimate, especially for deeper structures.

7.1.2. Workload

According to Figures 6.12, 6.13, 6.15 and Tables 6.6 and 6.7, the average force shows a significant negative difference between the new and old systems. In contrast, the NASA TLX score is only decreased in IQR, and no difference is found in median and maximum values. Concluding a lower workload for the new system and accepting hypothesis H_2 = **Multi-modal multi-feedback admittance control will decrease the human workload measured by the average force and NASA TLX survey.**

Firstly, it stands out that the number of participants is lower as compared to the number of participants that were used in the testing. This is because some participants did not fill out the NASA TLX survey correctly, and too much time was passed to re-fill the scores after discovery. Strikingly, the overall median NASA TLX score for both systems is relatively low, indicating a low workload.

For the NASA TLX, the difference between the new and old system for the median value is insignificant, thus not improving the workload. However, the IQR decreases significantly, indicating the participants have more similar experiences, likely due to the feedback provided to the operators. According to Table 6.5, the new system mainly yields an increase in perceived mental and temporal demand and a significant decrease in physical demand, performance and frustration. Therefore, it is logical that the results in Figure 6.15 show a significant decrease in average force for the new system. This results from the provided feedback, whereas the old system does not provide any such feedback; thus, the participants are unaware of high forces and torques and will not adjust their behaviour accordingly. The new system required much more training and information than the old one and provided more feedback that the operators had to pay attention to and respond to. This explains the increase in mental and temporal demand scores, which are expected to decrease over time when subjected to more extended training. It is concluded that the workload has decreased for the new system compared to the old system

Other metrics related to workload are given in Table 6.8 and shows that the new system decreases the work, total force and average force but yields an increase in time. This result is most likely due to the auditory feedback delay as well as the checklists, which causes operators to wait a long time before moving to the next fiducial. Decreasing this delay will significantly decrease the time.

Concerning workload metrics, Hart (2006) noted that a limitation of the NASA TLX is the interpretation of scores. Therefore the analysis is mainly limited to performing the same test (Grier, 2015). In itself, it is difficult to conclude whether the workload is high or low, but in comparison, it can be concluded that one system requires either a higher or lower workload. Thus the application of NASA TLX in this project is considered interpretable, but the limitation could be the cause of the widespread NASA TLX scores.

7.1.3. Usability

According to Figures 6.16, 6.17, 6.19 and Tables 6.9 and 6.10, the median SUS and CSUQ scores show a significant increase for the new system. The IQR for SUS and CSUQ shows a significant decrease, and the maximum SUS and CSUQ show no significant difference. Concluding higher usability for the new system and accepting hypothesis H_3 = **Multi-modal multi-feedback admittance control will improve the human-robot (system) usability, measured by the CSUQ and SUS scores**. For SUS and CSUQ, one participant did not fill in the survey, providing the data for 13 participants in each

group.

The significantly lower IQR is due to the provided feedback that results in a more uniform experience for all participants operating on the new system. This is perceived in both SUS and CSUQ scores. The SUS and CSUQ scores yield similar results except for the relative data spread. This is mostly due to the different scales, questions, and calculations of CSUQ and SUS. The non-normal distribution of data is expected due to the nature of scores ranging between a minimum and maximum value.

For the usability scores, previous experience with robots and computers can impact the scores. This project considered this by dividing the groups as equally as possible based on self-reflecting experience. It could be debated, however, how equally this was divided and how much self-reflecting experience resembles actual experience. Moreover, the participant pool is highly biased towards employees or robotics students. Therefore it is recommended to perform more testing with participants outside this pool and preferably related to the medical field.

7.1.4. Trust

According to Figures 6.21, 6.22 and Table 6.11, the median Trust Perception Scale-HRI for the new system yields an increase of 17.7% over the old system. However, the IQR also increases for the new system, whereas the maximum value does not provide a significant difference. This thesis concludes with insignificant results for the trust scores, thus rejecting the hypothesis H_4 = **Multi-modal multi-feedback admittance control will improve the human trust in the system, measured by the Trust Perception Scale-HRI.**

Both systems yield a large spread of data due to the standardised questionnaire related to HRI, including social and intelligent robots. Therefore statements that relate to social and intelligence metrics, such as incompetency, friendliness, and pleasantly, led to confusion for the participants. This is clearly stated by both groups of participants and shown in the data. In conversations, it turned out that participants operating on the old system asked more questions compared to the new system. This is likely visible in Figure 6.23, where the old system covers a larger data spread. However, adjusting the Trust Perception Scale-HRI would have been subjective and biased to the researcher. A better survey is recommended for measuring trust in the future that is more adaptable to the situation at hand.

For the trust, it is essential to consider its dependence on individual characteristics such as propensity to trust, self-confidence, individual differences and culture. Since the pool of participants was limited to mostly company employees or other robot-related students, the trust scores are likely positively biased. Another problem is that trust is only measured subjectively, and the discrepancies between an operator's subjective and objective measures have been a known issue (Hancock et al., 2011). This thesis concludes that more research, including both subjective and objective measurements, is needed to understand the human trust in robots for human operators outside this project's scope and the relationships between trust and performance in HRI, especially over time (M. Lewis et al., 2018). It is recommended to include medical staff in this research.

7.2. Concept and Design

The discussion and recommendations of the concept and design are summarised in Figure 7.3 and discussed in more detail below.

Discussion	
Concept and design	Meaning
General	
Criteria list	Stakeholders all focused on different criteria; therefore, this criteria list might have to be re-evaluated when more consensus has been reached. Changing the criteria list will significantly impact the final concept design
Less invasive option	For a less invasive option, another concept that can be considered is a dental guard with the addition of bone-anchored fiducials
Situation and configuration specific	The chosen fiducial configuration, block dimensions, number of fiducials do influence the results and design. Optimisation of these decisions was outside the scope of this thesis. It is recommended to place the most possible fiducials as far away from the surgical area as possible while remaining visible in the CT image
Functional implementations	Some chosen functional implementations were not included due to time limits or were considered unnecessary
Multi-modal admittance	
Vibrations	Vibrations were visible in surface contact, especially when the operator applied excessive force, adjustment of control gains is needed
Foot pedal haptics	Car pedals have increased resistance when pushing the pedal; this was lacking in the design, causing some overshooting. It also had some delays. A newer foot pedal with buttons that respond more ad-hoc would be recommended
Limit admittance control	Limiting the allowance of admittance control when a registration bit is attached is recommended, as a drill can cause significant accidents
Limit rotational control	The robot's kinematic model must be updated so that the rotation does not influence the TCP position. Consequently, the rotational control should be limited to an angle within x degrees, ensuring no increase in kinematic errors
Multi-feedback interface	
Information spiral	The more feedback provided, the more the participants consider what other information they are missing or would like to receive.
Similar experience	The feedback created a more similar experience between operators with the new system compared to the old system
Training and roles	
Increase accuracy	Training can significantly increase the accuracy of localisation tasks.
Less influence	Training had less influence on the new system, and the difference in skill level is less crucial
Communication	Communication does introduce significant outliers in the old system

Figure 7.3: Summary of the discussion on the thesis design

The final concept is chosen based on the drafted criteria list. During this project, it became clear that the stakeholders all focused on different criteria; therefore, this criteria list might have to be re-evaluated when more consensus has been reached. Changing the criteria list will significantly impact the final concept design. For a less invasive option, another concept that can be considered is a dental guard with the addition of bone-anchored fiducials; however, this was not considered due to lack of time.

For the chosen concept, the workflow adjustments in the hospital were only considered briefly since no clinical tests were performed. This registration type comes with additional pre-operative efforts, such as pre-operative imaging, fiducial localisation in the coordinate image frame, and fiducial placement. The fact that the chosen concept counteracts the advantages of IGS and CIS is an important notice that can limit the success of the design and its adaptation. The higher the accuracy of the method, the more complex the workflow.

Some chosen functional implementations were not included due to time limits or were considered unnecessary. These include the visible timer, blurring of positional control GUI, the process bar, automatic removal of double localisation positions, stopping of robot movement when workspace limit is exceeded, cancellation of the fixed mode, and sound when touching. The timer was replaced by time tracking, saving this data to be analysed afterwards. Blurring the positional control GUI and cancellation of the fixed mode were considered unimportant since it is far from the registration GUI and the safety mode is used. The process bar, automatic removal of double localisation positions, stopping robot movement when workspace limit is exceeded, and sound when touching was not implemented due to time limits.

Lastly, the number and placement of fiducials are essential factors in the FLE and TRE values, but the optimisation, in this case, was considered outside this project's scope. The chosen fiducial configuration resembled the shape of the ear and previously used configuration so that the PT could also be determined. However, changing the fiducial configuration and amount on the block could have changed the found FLE and TRE values, as they are configuration specific. It is recommended to place a large number of fiducials as far away from the surgical area as possible while remaining visible in the CT image.

7.2.1. Multi-modal admittance control

Participants intuitively understood the multi-modal admittance control and executed their tasks easily. During the localisation task, vibrations were visible in surface contact, especially when the operator applied excessive force. This must be avoided and requires adjustments to the admittance control loop

parameters. However, in the new system, the vibration does not increase the FLE since the operator has to let go of the registration bit to save its position.

Another remark is the lack of haptic feedback from the foot pedal. Car pedals have increased resistance when pushing the pedal; this was lacking in the design, causing some overshooting. Moreover, the foot pedal buttons sometimes responded after some time and required more effort. A newer foot pedal with buttons that respond more ad-hoc would be recommended.
One implementation that was not considered but is vital according to Figures 6.9 and 6.10 is a limitation to the rotational control. Only two participants used rotational control, increasing the FLE and TRE values. Still, it could be convenient in other situations and prevent human errors. The robot's kinematic model must be updated so that the rotation does not influence the TCP position. Consequently, the rotational control should be limited to an angle within x degrees, ensuring no increase in kinematic errors. Additionally, limiting the allowance of admittance control when a registration bit is attached is recommended, as a drill can cause significant accidents.

7.2.2. Multi-feedback interface

Participants were satisfied with the feedback interface. The more feedback provided, the more the participants consider what other information they are missing or would like to receive. This was found as the participant operating on the new system requested more additions and changes compared to the old system, such as the addition of a playing sound when the robot is moving by itself during safety mode and clear feedback when the registration mode is started.

The feedback created a more similar experience between operators with the new system compared to the old system. This is reflected in the IQR decrease in most metrics.

7.2.3. Training and roles

Figure 6.11 concludes that training can significantly increase the accuracy of localisation tasks. This has less influence on the new system, and the difference in skill level is less crucial. It can not be said that removing the additional operator directly leads to higher accuracy. However, from the perceived tests, communication does introduce significant outliers in the old system.

7.3. Other

Other topics include CT localisation and CI surgery, summarised in Figure 7.4. Although these topics are not chosen or outside this project, some recommendations can be made.

Discussion	
Other	Meaning
CT localisation	
Changes influence each other	Changes to the design for real-world localisation fundamentally change the FLE errors introduced for CT localisation. Hence designs should be considered for both tasks simultaneously.
Automate	For CT localisation, automated methods have been used to minimise the FLE, which is recommended
CI surgery	
Facial recess threshold	Limiting the patients undergoing surgery to certain thresholds related to their facial recess
Added value	The added value of the robot performing the surgery compared to the surgeon is limited and has to be increased
HITL required	Hence, a HITL design is recommended that increases autonomy with separate functions that can be added or removed to provide a synergy that builds trust and removes problems with responsibility, accountability, and adjustment to sudden situations

Figure 7.4: Summary of the discussion on other parts of this thesis

7.3.1. CT localisation

Although outside this project's scope, some recommendations are considered for CT localisation. The use of surface-based registration was excluded, but it is mentioned in the literature that SBR, in com-

bination with BFR, can improve the TRE due to its considerable improvement in FLE for the CT localisation (Kim and Kazanzides, 2017). Changes to the design for real-world localisation fundamentally change the FLE errors introduced for CT localisation. Hence designs should be considered for both tasks simultaneously. For CT localisation, automated methods have been used to minimise the FLE, which is recommended (Regodic et al., 2020).

7.3.2. CI surgery

The choice of robotic CI surgery is logical due to the static structures, the trend in otology towards IGS and CIS, its high-level precision, and its repetitiveness of tiring tasks (Panara et al., 2021). However, it requires a total system with sub-millimetric accuracy, which is a substantial challenge. One workaround of other robot surgical methods is limiting the patients undergoing surgery to certain thresholds related to their facial recess. Another interesting method developed by Labadie et al. (2004) is a different type of CI surgery that uses patient-specific frames that guide the grill along a surgical path and avoids important anatomical structures. In using robots for CI surgery, standard surgery methods must be reconsidered since robots do not have to apply the same rules as humans.

Even if the registration performs with high accuracy and sub-millimetric surgery is possible, the added value of the robot performing the surgery compared to the surgeon is limited and has to be increased. The time needed to perform the surgery and added pre-operative and intraoperative times seem larger compared to traditional surgeries. In addition, a robot can not respond to sudden accidents and poses challenges concerning responsibility, and legal liability in case of a calamity (Schaerer et al., 2009). Humans feel less responsible and accountable with increasing LORA (Beer et al., 2014; Datteri, 2013). Besides, surgeons are better than robots at weighing their past experiences to make complex choices and adapt to changing environments.

Hence, a HITL design that increases autonomy with separate functions that can be added or removed, such as lane-keeping and cruise control in self-driving vehicles is recommended. This provides a human-robot synergy that will build trust and removes problems with responsibility, accountability, and adjustment to sudden situations. The robot works alongside the human, utilising both abilities to their best yielding clinical relevance.

7.4. Conclusion

Benchmarking the design to a baseline non-human-robot interaction-focused design shows a significant increase in accuracy and usability and a significant decrease in workload. The study is inconclusive for trust, and additional research into trust-influencing factors specified to the specific application must be performed to draw any conclusion. The multi-modal multi-feedback design generally provided higher repeatability and reproducibility through feedback and better localisation control.

The results are highly dependent on the specific context, situation, configuration and experimental setup and clinical testing, as well as the inclusion of imaging localisation, is required. Besides, the participant pool is highly biased and limited, but the results still hold due to the comparison study and dividing of the participants based on the pre-questionnaire. Reproducing the tests in a clinical environment with medical staff is crucial and will provide new insights.

Performing CI robotic surgery autonomously seems out of reach as the added value is limited, and the robot's accuracy needs to be improved. HITL is recommended to increase autonomy with separate functions that can be added or removed. Human factors must be addressed as long as surgeons and robots work alongside performing surgery on human patients in a hospital workflow.

8. Conclusion

Including technical and human-robot interaction-related criteria in the double-diamond design process provides a successful high-accuracy, sub-millimetric registration method for an autonomous robot equipped to drill bone for cochlear implant surgery. Its performance, and thus success, is measured in accuracy, workload, usability and trust.

This thesis concludes that the inclusion of human-robot interaction-related criteria is essential for the success of any registration method for surgical robots due to their inevitable involvement with human teams in hospital workflows.

The research concludes that a multi-modal, multi-feedback admittance control, touch-based pair-point bone-anchored fiducial registration design is the best for cochlear implant surgery registration. Benchmarking this design to a baseline non-human-robot interaction-focused design shows a significant increase in accuracy and usability and a significant decrease in workload. The study is inconclusive for trust, and additional research into trust-influencing factors specified to the specific application must be performed to draw any conclusion.

This thesis addresses the literature gap concerning human influence on registration performance and contributes by showing its significance. It also contributes to the registration method literature by providing new theoretical knowledge on all factors influencing registration accuracy. Besides, it contributes to human-factor analysis research by providing new theoretical knowledge on human factors influencing robots in the medical field.

The unique symbiotic design significantly outperforms most state-of-the-art registration methods in accuracy despite using unskilled operators. Hence, it can provide a gold standard for registration accuracy.

"*Humans and robots have to engage in a reciprocal relationship that leads to mutual understanding and cooperation so that they can learn from and support each other. Not either humans or robots, but humans and robots – together in a symbiotic relationship.*" (Abbink, 2019).

References

Abbink, D. (2019). Tu delft robotics institute - human-robot symbiosis improves interaction. https://tudelftroboticsinstitute.nl/news/human-robot-symbiosis-improves-interaction

Abdelaal, A. E., Mathur, P., & Salcudean, S. E. (2019). Annual review of control, robotics, and autonomous systems robotics in vivo: A perspective on human-robot interaction in surgical robotics. https://doi.org/10.1146/annurev-control-091219

Advanced searching in web of science - advanced searching techniques - libguides at tufts university. (2022). https://researchguides.library.tufts.edu/c.php?g=249270&p=1660077

Alam, F., & Rahman, S. U. (2016). Intrinsic registration techniques for medical images: A state-of-the-art review optimal catheter selection during coronary angiography view project robbery detection using cctv video stream view project intrinsic registration techniques for medical images: A state-of-the-art review. https://www.researchgate.net/publication/302975396

Alam, F., Rahman, S. U., Ullah, S., & Khalil, A. (2016). A review on extrinsic registration methods for medical images. https://www.researchgate.net/publication/313160195

Albritton, F. D., Kingdom, T. T., & DelGaudio, J. M. (2001). Malleable registration mask: Application of a novel registration method in image guided sinus surgery. *American Journal of Rhinology*, *15*(4), 219–224.

Automaticaddison, A. . (2020). Find Homogeneous Transformation Matrices for a Robotic Arm – Automatic Addison. https://automaticaddison.com/find-homogeneous-transformation-matrices-for-a-robotic-arm/

Azagury, D. E., Dua, M. M., Barrese, J. C., Henderson, J. M., Buchs, N. C., Ris, F., Cloyd, J. M., Martinie, J. B., Razzaque, S., Nicolau, S., et al. (2015). Image-guided surgery. *Current problems in surgery*, *52*(12), 476–520.

Bale, R. J., Burtscher, J., Eisner, W., Obwegeser, A. A., Rieger, M., Sweeney, R. A., Dessl, A., Giacomuzzi, S. M., Twerdy, K., & Jaschke, W. (2000). Computer-assisted neurosurgery by using a noninvasive vacuum-affixed dental cast that acts as a reference base: Another step toward a unified approach in the treatment of brain tumors. *Journal of neurosurgery*, *93*(2), 208–213.

Ballesteros-Zebadúa, P., García-Garduño, O., Galván de La Cruz, O., Arellano-Reynoso, A., Lárraga-Gutiérrez, J., & Celis, M. (2016). Assessment of an image-guided neurosurgery system using a head phantom. *British journal of neurosurgery*, *30*(6), 606–610.

Bao, N., Chen, Y., Yue, Y., Li, H., Cui, Z., Zhuang, J., Tian, S., & Kang, Y. (2014). Fiducial markers configuration optimization in image-guided surgery. *Bio-medical materials and engineering*, *24*(6), 3361–3371.

Beatty, S. (2022). 6 simple search tips: Lessons learned from the scopus webinar | elsevier scopus blog. https://blog.scopus.com/posts/6-simple-search-tips-lessons-learned-from-the-scopus-webinar

Beer, J. M., Fisk, A. D., & Rogers, W. A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of human-robot interaction*, *3*(2), 74.

Beer, J. M., Prakash, A., Mitzner, T. L., & Rogers, W. A. (2011). *Understanding robot acceptance* (tech. rep.). Georgia Institute of Technology.

Boas, F. E., Fleischmann, D. et al. (2012). Ct artifacts: Causes and reduction techniques. *Imaging Med*, *4*(2), 229–240.

Brooke, J. et al. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, *189*(194), 4–7.

Bruckner, S., Seemann, R., & Elmenreich, W. (2002). *Applying a real-time interface to an optical tracking system*. na.

Cao, L., Liu, H., Zhou, Y., & Han, J. (2016). An image registration method for surgical robots based on human-robot cooperation. *2016 IEEE International Conference on Mechatronics and Automation, IEEE ICMA 2016*, 1107–1112. https://doi.org/10.1109/ICMA.2016.7558717

Cardinale, F., Rizzi, M., D'Orio, P., Casaceli, G., Arnulfo, G., Narizzano, M., Scorza, D., Momi, E. D., Nichelatti, M., Redaelli, D., Sberna, M., Moscato, A., & Castana, L. (2017). A new tool for touch-free patient registration for robot-assisted intracranial surgery: Application accuracy from a phantom study and a retrospective surgical series. *Neurosurgical focus*, *42*. https://doi.org/10.3171/2017.2.FOCUS16539

Carminucci, A., Nie, K., Weiner, J., Hargreaves, E., & Danish, S. F. (2018). Assessment of motion error for frame-based and noninvasive mask-based fixation using the leksell gamma knife icon radiosurgery system. *Journal of neurosurgery*, *129*(Suppl1), 133–139.

Carter, C. (n.d.). Web of science search tips - how to search in biomedical databases - neomed library at northeast ohio medical university (neomed). https://libraryguides.neomed.edu/library-tutorials/web-of-science-search-tips

Caversaccio, M., Gavaghan, K., Wimmer, W., Williamson, T., Ansò, J., Mantokoudis, G., Gerber, N., Rathgeb, C., Feldmann, A., Wagner, F., et al. (2017). Robotic cochlear implantation: Surgical procedure and first clinical experience. *Acta oto-laryngologica*, *137*(4), 447–454.

Chacón, A., Ponsa, P., & Angulo, C. (2021). Usability study through a human-robot collaborative workspace experience. *Designs*, *5*(2), 35.

Charlton, S. G., & O'Brien, T. G. (2019). *Handbook of human factors testing and evaluation*. CRC Press.

Chowdhury, A., Ahtinen, A., Pieters, R., & Vaananen, K. (2020). User experience goals for designing industrial human-cobot collaboration: A case study of franka panda robot. *Proceedings of the 11th nordic conference on human-computer interaction: Shaping experiences, shaping society*, 1–13.

Clancy, S. (2014). Ten ways to improve your pubmed search. http://www.nlm.nih.gov/pubs/techbull/jf11/jf11_skill_kit_pm_phrase_searching.html

Clinic, M. (n.d.). Cochlear implants. https://www.mayoclinic.org/tests-procedures/cochlear-implants/about/pac-20385021

Cohen, E. A., Kim, D., & Ober, R. J. (2015). Cramer-rao lower bound for point based image registration with heteroscedastic error model for application in single molecule microscopy. *IEEE Transactions on Medical Imaging*, *34*, 2632–2644. https://doi.org/10.1109/TMI.2015.2451513

Crouser, R. J., Ottley, A., & Chang, R. (2013). Balancing human and machine contributions in human computation systems. *Handbook of human computation* (pp. 615–623). Springer.

Datteri, E. (2013). Predicting the long-term effects of human-robot interaction: A reflection on responsibility in medical robotics. *Science and engineering ethics*, *19*(1), 139–160.

De Santis, A., Siciliano, B., De Luca, A., & Bicchi, A. (2008). An atlas of physical human–robot interaction. *Mechanism and Machine Theory*, *43*(3), 253–270.

Design Council. (2019). Framework for Innovation: Design Council's evolved Double Diamond. https://www.designcouncil.org.uk/our-work/skills-learning/tools-frameworks/framework-for-innovation-design-councils-evolved-double-diamond/

Eggers, G., Mühling, J., & Marmulla, R. (2006). Image-to-patient registration techniques in head surgery. *International Journal of Oral and Maxillofacial Surgery*, *35*, 1081–1095. https://doi.org/10.1016/J.IJOM.2006.09.015

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of pubmed, scopus, web of science, and google scholar: Strengths and weaknesses. *The FASEB Journal*, *22*, 338–342. https://doi.org/10.1096/FJ.07-9492LSF

Fan, Y., Yao, X., & Xu, X. (2020). A robust automated surface-matching registration method for neuronavigation. *Medical physics*, *47*(7), 2755–2767.

Feng, Y., & Max, L. (2014). Accuracy and precision of a custom camera-based system for 2-d and 3-d motion tracking during speech and nonspeech motor tasks. *Journal of speech, language, and hearing research*, *57*(2), 426–438.

Fitts, P. M. (1951). Human engineering for an effective air-navigation and traffic-control system.

Fitzpatrick, J. M., & West, J. B. (2001). The distribution of target registration error in rigid-body point-based registration. *IEEE Transactions on Medical Imaging*, *20*, 917–927. https://doi.org/10.1109/42.952729

Fitzpatrick, J. M. (2009). Fiducial registration error and target registration error are uncorrelated. *Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling*, *7261*, 21–32.

Fosch-Villaronga, E., Khanna, P., Drukarch, H., & Custers, B. H. (2021). A human in the loop in surgery automation. *Nature Machine Intelligence*, *3*(5), 368–369.

Franaszek, M., & Cheok, G. S. (2020). Using locally adjustable hand-eye calibrations to reduce robot localization error. *SN Applied Sciences*, *2*(5), 1–13.

Gerber, N., Gavaghan, K. A., Bell, B. J., Williamson, T. M., Weisstanner, C., Caversaccio, M. D., & Weber, S. (2013). High-accuracy patient-to-image registration for the facilitation of image-guided robotic microsurgery on the head. *IEEE transactions on bio-medical engineering*, *60*, 960–968. https://doi.org/10.1109/TBME.2013.2241063

Giovannitti, E., Nabavi, S., Squillero, G., & Tonda, A. (2022). A virtual sensor for backlash in robotic manipulators. *Journal of Intelligent Manufacturing*, 1–17.

Grauvogel, T. D., Engelskirchen, P., Semper-Hogg, W., Grauvogel, J., & Laszig, R. (2017). Navigation accuracy after automatic- and hybrid-surface registration in sinus and skull base surgery. *PLOS ONE*, *12*(7), 1–10. https://doi.org/10.1371/journal.pone.0180975

Greenwood, L., & Vallee, M. (2021). A step-by-step guide to image registration for spine surgery - brainlab. https://www.brainlab.com/journal/step-by-step-guide-to-image-registration-for-spine-surgery/

Grier, R. A. (2015). How high is high? a meta-analysis of nasa-tlx global workload scores. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 1727–1731.

Grimes, C. E., Bowman, K. G., Dodgion, C. M., & Lavy, C. B. (2011). Systematic review of barriers to surgical care in low-income and middle-income countries. *World journal of surgery*, *35*, 941–950. https://doi.org/10.1007/S00268-011-1010-1

Güler, Ö., Perwög, M., Kral, F., Schwarm, F., Bardosi, Z., Göbel, G., & Freysinger, W. (2013). Quantitative error analysis for computer assisted navigation: A feasibility study. *Medical physics*, *40*(2), 021910.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*, *53*(5), 517–527.

Harriott, C. E. (2015). *Workload and task performance in human-robot peer-based teams* (Doctoral dissertation). Vanderbilt University.

Hart, S. G. (2006). Nasa-task load index (nasa-tlx); 20 years later. *Proceedings of the human factors and ergonomics society annual meeting*, *50*(9), 904–908.

Hart, S. G., & Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology* (pp. 139–183). Elsevier.

Hawkins, J. (2020). Human ear | structure, function, & parts | britannica. https://www.britannica.com/science/ear#assembly=url~http%5C%3A%5C%2F%5C%2Fwww.britannica.com%5C%2FEBchecked%5C%2Ftopic-art%5C%2F175622%5C%2F530%5C%2FStructure-of-the-human-ear

Heineman, R. A. (1984). The logic and limits of trust. by bernard barber.(new brunswick, nj: Rutgers university press, 1983. pp. 190. 9.95, paper.) *American Political Science Review*, *78*(1), 209–210.

Horn, B. K., Hilden, H. M., & Negahdaripour, S. (1988). Closed-form solution of absolute orientation using orthonormal matrices. *JOSA A*, *5*(7), 1127–1135.

Image-guided technique in neurotology | ento key. (n.d.). https://entokey.com/image-guided-technique-in-neurotology/

ISO, I. (2011). Iso. *IEC*, *25010*, 2011.

Jiam, N. T., & Limb, C. J. (2016). The impact of round window vs cochleostomy surgical approaches on interscalar excursions in the cochlea: Preliminary results from a flat-panel computed tomography study. *World Journal of Otorhinolaryngology - Head and Neck Surgery*, *2*, 142–147. https://doi.org/10.1016/J.WJORL.2016.07.001

Ke, J., Zhang, S.-X., Hu, L., Li, C.-S., Zhu, Y.-F., Sun, S.-L., Wang, L.-F., & Ma, F.-R. (2016). Minimally invasive cochlear implantation assisted by bi-planar device: An exploratory feasibility study in vitro. *Chinese Medical Journal*, *129*(20), 2476–2483.

Keemink, A. Q., van der Kooij, H., & Stienen, A. H. (2018). Admittance control for physical human–robot interaction: *https://doi-org.tudelft.idm.oclc.org/10.1177/0278364918768950*, *37*, 1421–1444. https://doi.org/10.1177/0278364918768950

Kim, S., & Kazanzides, P. (2017). Fiducial-based registration with a touchable region model. *International journal of computer assisted radiology and surgery*, *12*(2), 277–289.

Labadie, R. F., Shah, R. J., Harris, S. S., Cetinkaya, E., Haynes, D. S., Fenlon, M. R., Juscyzk, A. S., Galloway, R. L., & Fitzpatrick, J. M. (2004). Submillimetric target-registration error using a novel, non-invasive fiducial system for image-guided otologic surgery. *Computer aided surgery : official journal of the International Society for Computer Aided Surgery*, *9*, 145–153. https://doi.org/10.3109/10929080500066922

Ledderose, G. J., Hagedorn, H., Spiegl, K., Leunig, A., & Stelter, K. (2012). Image guided surgery of the lateral skull base: Testing a new dental splint registration device. *Computer Aided Surgery*, *17*(1), 13–20.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, *46*(1), 50–80.

Lewis, J. R. (1995). Ibm computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, *7*(1), 57–78.

Lewis, M., Sycara, K., & Walker, P. (2018). The role of trust in human-robot interaction. *Foundations of trusted autonomy* (pp. 135–159). Springer, Cham.

Liao, R., Zhang, L., Sun, Y., Miao, S., & Chefd'Hotel, C. (2013). A review of recent advances in registration techniques applied to minimally invasive therapy. *IEEE transactions on multimedia*, *15*(5), 983–1000.

Liu, J., Singh, G., Al'Aref, S., Lee, B., Oleru, O., Min, J. K., Dunham, S., Sabuncu, M. R., & Mosadegh, B. (2019). Image registration in medical robotics and intelligent systems: Fundamentals and applications. *Advanced Intelligent Systems*, *1*, 1900048. https://doi.org/10.1002/AISY.201900048

Liu, W., Ding, H., Han, H., Xue, Q., Sun, Z., & Wang, G. (2009). The study of fiducial localization error of image in point-based registration. *Proceedings of the 31st Annual International Con-*

ference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009, 5088–5091. https://doi.org/10.1109/IEMBS.2009.5332731

Lubrano, E., & Clavel, R. (2010). Thermal calibration of a 3 dof ultra high-precision robot operating in industrial environment. *2010 IEEE international conference on robotics and automation*, 3692–3697.

Luebbers, H.-T., Messmer, P., Obwegeser, J. A., Zwahlen, R. A., Kikinis, R., Graetz, K. W., & Matthews, F. (2008). Comparison of different registration methods for surgical navigation in cranio-maxillofacial surgery. *Journal of Cranio-Maxillofacial Surgery*, *36*(2), 109–116.

Machetanz, K., Grimm, F., Wang, S., Bender, B., Tatagiba, M., Gharabaghi, A., & Naros, G. (2021). Patient-to-robot registration: The fate of robot-assisted stereotaxy. *The International Journal of Medical Robotics and Computer Assisted Surgery*, *17*, e2288. https://doi.org/10.1002/RCS.2288

Machetanz, K., Grimm, F., Wuttke, T. V., Kegele, J., Lerche, H., Tatagiba, M., Rona, S., Gharabaghi, A., Honegger, J., & Naros, G. (2021). Frame-based and robot-assisted insular stereo-electroencephalograph via an anterior or posterior oblique approach. *Journal of Neurosurgery*, *135*(5), 1477–1486.

Marmulla, R., Eggers, G., & Mühling, J. (2005). Laser surface registration for lateral skull base surgery. *min-Minimally Invasive Neurosurgery*, *48*(03), 181–185.

Matt. (2022). Absolute Orientation - Horn's method. https://nl.mathworks.com/matlabcentral/fileexchange/26186-absolute-orientation-horn-s-method

Medicine, S. (n.d.). Cochlear implant surgery – oto surgery atlas. https://otosurgeryatlas.stanford.edu/otologic-surgery-atlas/cochlear-implantation/cochlear-implant-surgery/#

Mitchell, M. B., & Labadie, R. F. (2020). Cost-effectiveness of intraoperative ct scanning in cochlear implantation in fee-for-service and bundled payment models. *Ear, nose, & throat journal*. https://doi.org/10.1177/0145561320952192

Mongen, M. A., & Willems, P. W. (2019). Current accuracy of surface matching compared to adhesive markers in patient-to-image registration. *Acta Neurochirurgica*, *161*(5), 865–870.

MRPT. (2022). Iterative Closest Point (ICP) and other registration algorithms — MRPT 2.5.3 documentation. https://docs.mrpt.org/reference/latest/tutorial-icp-alignment.html

Mudry, A., & Mills, M. (2013). The early history of the cochlear implant: A retrospective. *JAMA Otolaryngology–Head & Neck Surgery*, *139*, 446–453. https://doi.org/10.1001/JAMAOTO.2013.293

Murphy, M. J. (1999). The importance of computed tomography slice thickness in radiographic patient positioning for radiosurgery. *Medical Physics*, *26*(2), 171–175.

Murphy, R. R. (2019). *Introduction to ai robotics*. MIT press.

Murphy, R. R., Nomura, T., Billard, A., & Burke, J. L. (2010). Human–robot interaction. *IEEE robotics & automation magazine*, *17*(2), 85–89.

Nahavandi, S. (2017). Trusted autonomy between humans and robots: Toward human-on-the-loop in robotics and autonomous systems. *IEEE Systems, Man, and Cybernetics Magazine*, *3*(1), 10–17.

Naik, N. S., & Rube, P. (2014). Physiological tremor estimation methods: An overview. *Int. J. Innov. Sci. Eng. Technol*, *1*, 556–559.

NIDCD. (2021). What are cochlear implants for hearing? | nidcd. https://www.nidcd.nih.gov/health/cochlear-implants

Oka, M., Cho, B., Matsumoto, N., Hong, J., Jinnouchi, M., Ouchida, R., Komune, S., & Hashizume, M. (2014). A preregistered stamp method for image-guided temporal bone surgery. *International journal of computer assisted radiology and surgery*, *9*(1), 119–126.

Okamura, A. M., Matarić, M. J., & Christensen, H. I. (2010). Medical and health-care robotics. *IEEE Robotics & Automation Magazine*, *17*(3), 26–37.

Oliveira, F. P., & Tavares, J. M. R. (2014). Medical image registration: A review. *Computer methods in biomechanics and biomedical engineering*, *17*, 73–93. https://doi.org/10.1080/10255842.2012.670855

Omara, A. I., Wang, M., Fan, Y., & Song, Z. (2014). Anatomical landmarks for point-matching registration in image-guided neurosurgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, *10*(1), 55–64.

Opdenakker, Y., Swennen, G., & Abeloos, J. (2017). Application of a non-invasive reference headband and a surgical splint for intraoperative paediatric navigation. *International journal of oral and maxillofacial surgery*, *46*(3), 360–362.

Owen-Hill, A. (2016). Robotics research 101: Getting started with force control. https://blog.robotiq.com/robotics-research-101-getting-started-with-force-control

Panara, K., Shahal, D., Mittal, R., & Eshraghi, A. A. (2021). Robotics for cochlear implantation surgery: Challenges and opportunities. *Otology & neurotology : official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, *42*, e825–e835. https://doi.org/10.1097/MAO.0000000000003165

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human factors*, *39*(2), 230–253.

Park, S.-Y., & Subbarao, M. (2003). An accurate and fast point-to-plane registration technique. *Pattern Recognition Letters*, *24*(16), 2967–2976.

Payne, C. J., Vyas, K., Bautista-Salinas, D., Zhang, D., Marcus, H. J., & Yang, G.-Z. (2021). Shared-control robots. *Neurosurgical Robotics*, 63–79.

Perwög, M., Bardosi, Z., Diakov, G., Jeleff, O., Kral, F., & Freysinger, W. (2018). Probe versus microscope: A comparison of different methods for image-to-patient registration. *International Journal of Computer Assisted Radiology and Surgery*, *13*(10), 1539–1548.

Peternel, L., & Babič, J. (2019). Target of initial sub-movement in multi-component arm-reaching strategy. *Scientific reports*, *9*(1), 1–9.

Prewett, M. S., Johnson, R. C., Saboe, K. N., Elliott, L. R., & Coovert, M. D. (2010). Managing workload in human–robot interaction: A review of empirical studies. *Computers in Human Behavior*, *26*(5), 840–856.

Raja. (2019). Butterworth Filter: First Order and Second Order Low Pass Butterworth Filter. https://circuitdigest.com/tutorial/butterworth-filters-first-order-and-second-order-low-pass-butterworth-filters

Rajiv. (2022). Low Pass, High Pass and Band Pass Filters – Simple Explanation. https://www.rfpage.com/low-pass-high-pass-and-band-pass-filters-simple-explanation/

Regodic, M., Bardosi, Z., & Freysinger, W. (2020). Automatic fiducial marker detection and localization in ct images: A combined approach. *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*, *11315*, 507–514.

Robillard, A. (2019). Target registration error. https://slideplayer.com/slide/16750806/

Schaefer, K. E. (2016). Measuring trust in human robot interactions: Development of the "trust perception scale-hri". *Robust intelligence and trust in autonomous systems* (pp. 191–218). Springer.

Schaerer, E., Kelley, R., & Nicolescu, M. (2009). Robots as animals: A framework for liability and responsibility in human-robot interactions. *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*, 72–77.

Schneider, D., Hermann, J., Gerber, K. A., Ansó, J., Caversaccio, M. D., Weber, S., & Anschuetz, L. (2018). Noninvasive registration strategies and advanced image guidance technology for submillimeter surgical navigation accuracy in the lateral skull base. *Otology & neurotology*, *39*(10), 1326–1335.

Shah, M. (2014). Calibration and registration techniques for robotics. http://faculty.cooper.edu/mili/Calibration/index.html

Shah, N. (2014). Access to imaging technology in the developing world. *Radiology in global health* (pp. 13–17). Springer.

Shamir, R. R., Joskowicz, L., & Shoshan, Y. (2011). Fiducial optimization for minimal target registration error in image-guided neurosurgery. *IEEE transactions on medical imaging*, *31*(3), 725–737.

Shamir, R. R., Joskowicz, L., Spektor, S., & Shoshan, Y. (2009). Localization and registration accuracy in image guided neurosurgery: A clinical study. *International journal of computer assisted radiology and surgery*, *4*(1), 45–52.

Shanmugasundar, G., Sivaramakrishnan, R., Meganathan, S., & Balasubramani, S. (2019). Structural optimization of an five degrees of freedom (t-3r-t) robot manipultor using finite element analysis. *Materials Today: Proceedings*, *16*, 1325–1332.

Shappell, S. A., & Wiegmann, D. A. (2000). The human factors analysis and classification system–hfacs.

Sheridan, T. B. (2016). Human–robot interaction: Status and challenges. *Human factors*, *58*(4), 525–532.

Siciliano, B., Sciavicco, L., Villani, L., & Oriolo, G. (2009). *Force control*. Springer.

Siciliano, B., & Villani, L. (1999). *Robot force control*. Springer Science & Business Media.

SickKids. (n.d.). Mastoidectomy to treat cholesteatoma or ear infection. https://www.aboutkidshealth.ca/Article?contentid=1006&language=English

Skumsnes, B. H. (2012). Teleoperation of mobile robot manipulators.

Slamani, M., Nubiola, A., & Bonev, I. A. (2012). Modeling and assessment of the backlash error of an industrial robot. *Robotica*, *30*(7), 1167–1175.

Soteriou, E., Grauvogel, J., Laszig, R., & Grauvogel, T. D. (2016). Prospects and limitations of different registration modalities in electromagnetic ent navigation. *European Archives of Oto-Rhino-Laryngology*, *273*(11), 3979–3986.

Spinczyk, D., & Fabian, S. (2017). Target registration error minimization involving deformable organs using elastic body splines and particle swarm optimization approach. *Surgical Oncology*, *26*(4), 489–497.

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006). Common metrics for human-robot interaction. *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, 33–40.

Su, Y., Sun, Y., Hosny, M., Gao, W., & Fu, Y. (2022). Facial landmark-guided surface matching for image-to-patient registration with an rgb-d camera. *The International Journal of Medical Robotics and Computer Assisted Surgery*, *18*(3), e2373.

Šuligoj, F., Jerbić, B., Švaco, M., & Šekoranja, B. (2018). Fully automated point-based robotic neurosurgical patient registration procedure. *International Journal of Simulation Modelling*, *17*(3), 458–471.

Sun, Y., Luebbers, H.-T., Agbaje, J. O., Schepers, S., Vrielinck, L., Lambrichts, I., & Politis, C. (2013). Validation of anatomical landmarks-based registration for image-guided surgery: An in-vitro study. *Journal of Cranio-Maxillofacial Surgery*, *41*(6), 522–526.

Suwelack, S., Röhl, S., Bodenstedt, S., Reichard, D., Dillmann, R., dos Santos, T., Maier-Hein, L., Wagner, M., Wünscher, J., Kenngott, H., et al. (2014). Physics-based shape matching for intraoperative image guidance. *Medical physics*, *41*(11), 111901.

Taylor, R. H. (2006). A perspective on medical robotics. *Proceedings of the IEEE*, *94*, 1652–1664. https://doi.org/10.1109/JPROC.2006.880669

Teatini, A., de Frutos, J. P., Langø, T., Edwin, B., & Elle, O. (2018). Assessment and comparison of target registration accuracy in surgical instrument tracking technologies. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1845–1848.

Übelhör, T., Gesenhues, J., Ayoub, N., Modabber, A., & Abel, D. (2020). 3d camera-based markerless navigation system for robotic osteotomies. *At-Automatisierungstechnik*, *68*, 863–879. https://doi.org/10.1515/AUTO-2020-0032/MACHINEREADABLECITATION/RIS

Van Wyk, K., & Marvel, J. A. (2017). Strategies for improving and evaluating robot registration performance. *IEEE Transactions on Automation Science and Engineering*, *15*(1), 320–328.

Varma, T., & Eldridge, P. (2006). Use of the neuromate stereotactic robot in a frameless mode for functional neurosurgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, *2*(2), 107–113.

Vasconez, J. P., Kantor, G. A., & Cheein, F. A. A. (2019). Human–robot interaction in agriculture: A survey and current challenges. *Biosystems engineering*, *179*, 35–48.

Vasic, M., & Billard, A. (2013). Safety issues in human-robot interactions. https://ieeexplore-ieee-org.tudelft.idm.oclc.org/stamp/stamp.jsp?tp=&arnumber=6630576

Vocetka, M., Bobovskỳ, Z., Babjak, J., Suder, J., Grushko, S., Mlotek, J., Krys, V., & Hagara, M. (2021). Influence of drift on robot repeatability and its compensation. *Applied Sciences*, *11*(22), 10813.

Wang, A., & Gollakota, S. (2019). Millisonic: Pushing the limits of acoustic motion tracking. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–11.

Wang, J., Liu, H., Ke, J., Hu, L., Zhang, S., Yang, B., Sun, S., Guo, N., & Ma, F. (2020). Image-guided cochlear access by non-invasive registration: A cadaveric feasibility study. *Scientific Reports 2020 10:1*, *10*, 1–13. https://doi.org/10.1038/S41598-020-75530-7

Wellborn, P. S., Dillon, N. P., Russell, P. T., & Webster, R. J. (2017). Coffee: The key to safer image-guided surgery—a granular jamming cap for non-invasive, rigid fixation of fiducial markers to the patient. *International journal of computer assisted radiology and surgery*, *12*(6), 1069–1077.

West, J. B., Fitzpatrick, J. M., Toms, S. A., Maurer Jr, C. R., & Maciunas, R. J. (2001). Fiducial point placement and the accuracy of point-based, rigid body registration. *Neurosurgery*, *48*(4), 810–817.

Westebring–van der Putten, E. P., Goossens, R. H., Jakimowicz, J. J., & Dankelman, J. (2008). Haptics in minimally invasive surgery–a review. *Minimally Invasive Therapy & Allied Technologies*, *17*(1), 3–16.

WHO. (2021). Deafness and hearing loss. https://www.who.int/en/news-room/fact-sheets/detail/deafness-and-hearing-loss

Widmann, G., Stoffner, R., & Bale, R. (2009). Errors and error management in image-guided craniomaxillofacial surgery. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, *107*(5), 701–715.

Wilkinson, A., Gonzales, M., Hoey, P., Kontak, D., Wang, D., Torname, N., Laderoute, S., Han, Z., Allspaw, J., Platt, R., et al. (2021). Design guidelines for human–robot interaction with assistive robot manipulation systems. *Paladyn, Journal of Behavioral Robotics*, *12*(1), 392–401.

Woerdeman, P. A., Willems, P. W., Noordmans, H. J., Tulleken, C. A., & van der Sprenkel, J. W. B. (2007). Application accuracy in frameless image-guided neurosurgery: A comparison study of three patient-to-image registration methods. *Journal of neurosurgery*, *106*(6), 1012–1016.

Wolf, A., & Shoham, M. (2009). Medical automation and robotics. *Springer handbook of automation* (pp. 1397–1407). Springer.

Wong, R., Jivraj, J., & Yang, V. (2014). Current limitations and opportunities for surgical navigation. https://www.researchgate.net/publication/272176205_Current_Limitations_and_Opportunities_for_Surgical_Navigation

Woodside, M. R., Bristow, D. A., & Landers, R. G. (2020). A kinematic error observer for robot end effector estimation. *Procedia Manufacturing*, *48*, 1054–1063.

Wyawahare, M. V., Patil, P. M., & Abhyankar, H. K. (2009). Image registration techniques: An overview. *International Journal of Signal Processing*, *2*.

Yang, G.-Z., Cambias, J., Cleary, K., Daimler, E., Drake, J., Dupont, P. E., Hata, N., Kazanzides, P., Martel, S., Patel, R. V., et al. (2017). Medical robotics—regulatory, ethical, and legal considerations for increasing levels of autonomy.

Zanotto, V., Boscariol, P., Gasparetto, A., Lanzutti, A., Vidoni, R., Lorenzo, N. D., Gallina, P., Via, A. D., & Rossi, A. (2011). A master-slave haptic system for neurosurgery. *Applied Bionics and Biomechanics*, *8*, 209–220. https://doi.org/10.3233/ABB-2011-0026

Zhang, T., Du, L., & Dai, X. (2014). Test of robot distance error and compensation of kinematic full parameters. *Advances in Mechanical Engineering*, *6*, 810684.

Zhao, S. (2016). Time derivative of rotation matrices: A tutorial. *arXiv preprint arXiv:1609.06088*.

Zheng, H., Rosal, M. C., Li, W., Borg, A., Yang, W., Ayers, D. C., Franklin, P. D., et al. (2018). A web-based treatment decision support tool for patients with advanced knee arthritis: Evaluation of user interface and content design. *JMIR human factors*, *5*(2), e8568.

Zhi, D. (2015). Towards estimating fiducial localization error of point-based registration in image-guided neurosurgery. *Bio-Medical Materials and Engineering*, *26*(s1), S943–S949.

Zinreich, S. J., Tebo, S., Long, D., Brem, H., Mattox, D., Loury, M., Vander Kolk, C., Koch, W., Kennedy, D., & Bryan, R. (1993). Frameless stereotaxic integration of ct imaging data: Accuracy and initial applications. *Radiology*, *188*(3), 735–742.

Zitová, B., & Flusser, J. (2003). Image registration methods: A survey. *Image and Vision Computing*, *21*, 977–1000. https://doi.org/10.1016/S0262-8856(03)00137-9

A. Literature search method

This appendix demonstrates how an extensive literature search concerning registration methods can be conducted.

A literature search consists of several steps:

1. Databases should be chosen while considering the interpretation differences.
2. A clear scope should be defined, followed by writing a search query.
3. The search query is executed.

A.1. Databases

This literature search recommends three databases: PubMed, Scopus and Web of Science (WOS). The reason for selecting PubMed is its focus on medicine, bio-medicine, life science and clinical trials (Falagas et al., 2008). WOS is added due to its emphasis on science and technology without the medical aspect (2008). Furthermore, WOS incorporates patents. Patents can be attractive for this thesis as companies are more likely to describe new techniques in patents rather than literature. Utilising PubMed and WOS provides broad coverage of medical and non-medical-related papers and patents. Nevertheless, Scopus is added by virtue of its extensive coverage of journals, 12,850, and its origin outside the U.S (2008). The three databases thoroughly cover different journals, papers, sources and topics. Scopus focuses on physical sciences, and health sciences (2008). Google Scholar is merely used to retrieve information since it provides inconsistent accuracy (2008).

A.1.1. PubMed interpretation

PubMed is distinguishable from other databases considering it is human-curated and does not probe full-text articles. Consequently, it will automatically map terms to their same vocabulary. Hence the search ”high blood pressure” will also retrieve articles related to hypertension (Clancy, 2014). In PubMed, including double quotation marks is vital as it interprets separate words as independent search terms. Additionally, an asterisk is exploited to retrieve alternative endings: citizen jur* includes jury and juries.

PubMed will occasionally understand the meaning of some acronyms and abbreviations, though not all. Pubmed inquiries thus require mindful attention.

The Boolean operators in Pubmed are:

1. AND retrieves results that possess all search terms
2. OR retrieves outcomes that include at least one of the terms
3. NOT excludes the retrieval of terms from the intended search

A.1.2. Scopus interpretation

Like PubMed, Scopus will automatically add an AND operator between two separate words, requiring double quotation marks. Scopus likewise excludes full-text articles and utilises the Title, Abstract and Keywords as its main sources. Regarding keywords, both author keywords and subject-specific databases are included (Beatty, 2022). Furthermore, punctuation is ignored, and certain general phrases like ”it” and ”of” are excluded. In addition to the operators AND, OR, and NOT, Scopus uses two proximity operators:

1. PRE/n: An unconnected phrase that allows particular words to be separated. For example, one searches for titles including ”zika virus” while additionally including titles such as ”zika and

dengue virus". In which case the search term becomes zika PRE/2 virus (2022).

2. W/n: An allowance for the difference in word order within a certain range. Search phrase becomes zika W/2 virus and can contain titles like "virus-like zika", "virus, zika", etc. (2022).

Scopus uses two types of punctuation while searching for wildcards:

1. ?: Any single character. Example: wom?n can receive both woman and women.
2. *: Any number of characters, including zero. Example: comput* can return computer, computers, computerisation, and computational.

A.1.3. Web of Science (WOS) interpretation

WOS likewise adds the AND operator between two individual words; hence, double quotation marks are required. On the contrary, full-text screening is an element of WOS (Carter, n.d.). WOS-specific operators, in complement to the traditional operators AND, OR and NOT, are:

1. NEAR: Equivalent to PRE/n function in Scopus; finds terms within a certain number of each other ("Advanced Searching in Web of Science - Advanced Searching Techniques - LibGuides at Tufts University", 2022).
2. SAME: Search terms that must appear in the same sentence ("Advanced Searching in Web of Science - Advanced Searching Techniques - LibGuides at Tufts University", 2022).

Regarding wildcards, the ? and * symbols are equal in functionality to Scopus. Additionally, $ is used to substitute any zero or one character. Thus disease$ can include disease, diseased, and diseases ("Advanced Searching in Web of Science - Advanced Searching Techniques - LibGuides at Tufts University", 2022).

A.2. Scope

It is important to have a clear scope to specify what terms are required, what limitations should be in place, and whether the results are sufficient and useable. The PICO chart can be used to scope a project. PICO stands for:

1. P: The problem, process, population or predicament
2. I: The improvement or intervention
3. C: The comparison
4. O: The outcome

In this subject, the PICO can is expressed as:

1. P: Registration methods for robotic surgery
2. I: High accuracy
3. C: Different techniques; what is the distinction in accuracy
4. O: Various error metrics

So, the search statement is stated:
I am searching for high-accuracy registration methods for robotic surgery. What is the accuracy of different approaches displayed in various error metrics?

A.2.1. Exclusion and inclusion criteria

It is necessary to state and comprehend the limits of the search. Some limitations of this search are:

1. Only material in the English language will be regarded
2. A time frame of the last ten years is asserted
3. Case studies are excluded due to their low statistical value
4. Articles, conference papers and technical reports are considered valuable resources (Patents may be added; however, patents likely require a separate search query)
5. Intra-imaging techniques are excluded

6. No cross-referencing will take place
7. No minimum number of citations is required

A.3. Query

Concerning search terms, there is a trade-off between using adequate terms to represent all critical concepts and avoiding numerous terms that merely conduct limited articles (Clancy, 2014). Moreover, specific phrases should be evaded, such as imprecise words. These could include increased, better, greater, less, worse, etc. (2014).
The queries are built from diverse search terms gathered from known literature. These search terms are displayed in Figure A.1 and classified in different columns. Search terms 1 are the terms used to express the transformation within the registration process. Search terms 2 are terms and synonyms for registration. Search term 3 shows the different types of keywords for various registration methods. The latter can be selected based on the scope of the literature search. If one exclusively wants to find non-invasive methods, one has to exclude the terms such as frame-based and invasive. It is assumed that all registration methods are included. The search terms of column 4 indicate registration usage within the surgical world.

Search term 1	Search term 2	Search term 3	Search term 4
Patient to image	registration	Markerless	Surgery
Robot-world	Calibration	Touch-free	Procedure
Robot	Matching	Frameless	Surgical
Patient	Detection	Frame-based	Intraoperative
Model-based	Localization	Non-invasive	Preoperative
Fiducial	initialising	marker	Operative
Image-guided	Search	Geometrical based	Clinical
Target	Transform	Feature based	
Interaction image	Transformation	Invasive	
Robotic	Correlation	Touch-based	
Model-to-feature	Coregistration	Extrinsic	
Spatial	Initialisation	Intrinsic	
Image to robot		Automatic	
Hand eye		Semiautomatic	
Pair point			
Medical image			
image-to-physical			
Coordinate space			

Figure A.1: An outline of the various search terms utilised for the queries

The queries combine the different terms. The resulting queries are summarised in Figures A.2, A.2 and A.3.

```
(
        ("Patient to image"[Title/Abstract]) OR ("Robot to world"[Title/Abstract]) OR ("Robot-world"[Title/Abstract])
        OR ("Patient-image"[Title/Abstract]) OR (Robot[Title/Abstract]) OR (Robotic[Title/Abstract])
        OR (Patient[Title/Abstract]) OR ("Model-based"[Title/Abstract]) OR (Fiducial[Title/Abstract]) OR ("Image-guided"[Title/Abstract])
        OR (Target[Title/Abstract]) OR (Targeting[Title/Abstract]) OR ("Interaction image"[Title/Abstract])
        OR ("Model-to-feature"[Title/Abstract]) OR (Spatial[Title/Abstract]) OR ("Image-to-robot"[Title/Abstract])
        OR ("Hand eye"[Title/Abstract]) OR ("Pair point"[Title/Abstract]) OR ("Medical Image"[Title/Abstract])
        OR ("Image-to-physical"[Title/Abstract]) OR (Physical[Title/Abstract]) OR ("Coordinate space"[Title/Abstract])
)

AND (
        (Registration[Title/Abstract]) OR (Calibration[Title/Abstract]) OR (Matching[Title/Abstract])
        OR (Detection[Title/Abstract]) OR (Localization[Title/Abstract]) OR (Initiali*[Title/Abstract])
        OR (Search[Title/Abstract]) OR (Transform*[Title/Abstract]) OR (Coregistration[Title/Abstract])
)
AND (
        (Markerless[Title/Abstract]) OR (Extrinsic[Title/Abstract]) OR (Intrinsic[Title/Abstract])
        OR ("Touch-free"[Title/Abstract]) OR ("Touch-based"[Title/Abstract]) OR (Frameless[Title/Abstract])
        OR ("Frame-based"[Title/Abstract]) OR ("Non-invasive"[Title/Abstract]) OR (marker*[Title/Abstract]) OR (Geometrical[Title/Abstract])
        OR (Feature[Title/Abstract]) OR (Invasive[Title/Abstract]) OR (automatic[Title/Abstract])
        OR ("Semi-automatic"[Title/Abstract]) ) AND ( (Surger*[Title/Abstract]) OR (Procedure[Title/Abstract])
        OR (Surgical[Title/Abstract]) OR (Intraoperative[Title/Abstract]) OR (Preoperative[Title/Abstract]) OR (Clinical[Title/Abstract])
)

NOT ("intra-operative imaging"[Title/Abstract])
```

Figure A.2: Pubmed query related to registration methods

```
(
TITLE(
        (
                ("Patient to image") OR ("Robot to world") OR ("Robot-world") OR ("Patient-image") OR (Robot*)
                OR (Patient) OR ("Model-based") OR (Fiducial) OR ("Image-guided") OR (Target*)
                OR ("Interaction image") OR ("Model-to-feature") OR (Spatial) OR ("Image-to-robot")
                OR ("Hand eye") OR ("Pair point") OR ("Medical Image") OR ("Image-to-physical") OR (Physical)
                OR ("Coordinate space")
        )
        AND (
                (Registration) OR (Calibration) OR (Matching) OR (Detection) OR (Localization) OR (Initiali*)
                OR (Search) OR (Transform*) OR (Coregistration)
        )
        AND (
                (Markerless) OR (Extrinsic) OR (Intrinsic) OR ("Touch-free") OR ("Touch-based") OR (Frameless)
                OR ("Frame-based") OR ("Non-invasive") OR (marker*) OR (Geometrical) OR (Feature)
                OR (Invasive) OR (*automatic)
        )
        AND (
                (Surger*) OR (Procedure) OR (Surgical) OR (Surgical) OR (Intraoperative) OR (Preoperative)
                OR (Clinical)) AND NOT ("intra-operative imaging")
        )
)

OR (
        ABS (
                (
                        ("Patient to image") OR ("Robot to world") OR ("Robot-world") OR ("Patient-image")
                        OR (Robot*) OR (Patient) OR ("Model-based") OR (Fiducial) OR ("Image-guided")
                        OR (Target*) OR ("Interaction image") OR ("Model-to-feature") OR (Spatial)
                        OR ("Image-to-robot") OR ("Hand eye") OR ("Pair point") OR ("Medical Image")
                        OR ("Image-to-physical") OR (Physical) OR ("Coordinate space")
                )
                AND (
                        (Registration) OR (Calibration) OR (Matching) OR (Detection) OR (Localization)
                        OR (Initiali*) OR (Search) OR (Transform*) OR (Coregistration)
                )
                AND (
                        (Markerless) OR (Extrinsic) OR (Intrinsic) OR ("Touch-free") OR ("Touch-based")
                        OR (Frameless) OR ("Frame-based") OR ("Non-invasive") OR (marker*) OR (Geometrical)
                        OR (Feature) OR (Invasive) OR (*automatic)
                )
                AND (
                        (Surger*) OR (Procedure) OR (Surgical) OR (Surgical) OR (Intraoperative)
                        OR (Preoperative) OR (Clinical)
                )
                AND NOT ("intra-operative imaging")
        )

      AND KEY (
              (
                      ("Patient to image") OR ("Robot to world") OR ("Robot-world") OR ("Patient-image")
                      OR (Robot*) OR (Patient) OR ("Model-based") OR (Fiducial) OR ("Image-guided") OR (Target*)
                      OR ("Interaction image") OR ("Model-to-feature") OR (Spatial) OR ("Image-to-robot")
                      OR ("Hand eye") OR ("Pair point") OR ("Medical Image") OR ("Image-to-physical")
                      OR (Physical) OR ("Coordinate space")
              )
              AND (
                      (Registration) OR (Calibration) OR (Matching) OR (Detection) OR (Localization)
                      OR (Initiali*) OR (Search) OR (Transform*) OR (Coregistration)
              )
              AND (
                      (Markerless) OR (Extrinsic) OR (Intrinsic) OR ("Touch-free") OR ("Touch-based")
                      OR (Frameless) OR ("Frame-based") OR ("Non-invasive") OR (marker*) OR (Geometrical)
                      OR (Feature) OR (Invasive) OR (*automatic)
              )
              AND (
                      (Surger*) OR (Procedure) OR (Surgical) OR (Surgical) OR (Intraoperative)
                      OR (Preoperative) OR (Clinical)
              )
              AND NOT ("intra-operative imaging")
      )
)
```

Figure A.3: SCOPUS query related to registration methods

```
(TI=
        ((
                (
                        ("Patient to image") OR ("Robot to world") OR ("Robot-world") OR ("Patient-image")
                        OR (Robot*) OR (Patient) OR ("Model-based") OR (Fiducial) OR ("Image-guided")
                        OR (Target*) OR ("Interaction image") OR ("Model-to-feature") OR (Spatial)
                        OR ("Image-to-robot") OR ("Hand eye") OR ("Pair point") OR ("Medical Image")
                        OR ("Image-to-physical") OR (Physical) OR ("Coordinate space")
                )
                AND (
                        (Registration) OR (Calibration) OR (Matching) OR (Detection) OR (Localization)
                        OR (Initiali*) OR (Search) OR (Transform*) OR (Coregistration)
                )
                AND (
                        (Markerless) OR (Extrinsic) OR (Intrinsic) OR ("Touch-free") OR ("Touch-based")
                        OR (Frameless) OR ("Frame-based") OR ("Non-invasive") OR (marker*) OR (Geometrical)
                        OR (Feature) OR (Invasive) OR (*automatic)
                )
                AND (
                        (Surger*) OR (Procedure) OR (Surgical) OR (Surgical) OR (Intraoperative)
                        OR (Preoperative) OR (Clinical)
                )
                NOT ("intra-operative imaging")
        ))
)
OR (
        AB=((
                (
                        ("Patient to image") OR ("Robot to world") OR ("Robot-world") OR ("Patient-image")
                        OR (Robot*) OR (Patient) OR ("Model-based") OR (Fiducial) OR ("Image-guided")
                        OR (Target*) OR ("Interaction image") OR ("Model-to-feature") OR (Spatial)
                        OR ("Image-to-robot") OR ("Hand eye") OR ("Pair point") OR ("Medical Image")
                        OR ("Image-to-physical") OR (Physical) OR ("Coordinate space")
                )
                AND (
                        (Registration) OR (Calibration) OR (Matching) OR (Detection) OR (Localization)
                        OR (Initiali*) OR (Search) OR (Transform*) OR (Coregistration)
                )
                AND (
                        (Markerless) OR (Extrinsic) OR (Intrinsic) OR ("Touch-free")
                        OR ("Touch-based") OR (Frameless) OR ("Frame-based") OR ("Non-invasive") OR (marker*)
                        OR (Geometrical) OR (Feature) OR (Invasive) OR (automatic) OR ("Semi-automatic")
                )
                AND (
                        (Surger*) OR (Procedure) OR (Surgical) OR (Surgical) OR (Intraoperative)
                        OR (Preoperative) OR (Clinical)
                )
                NOT ("intra-operative imaging")
        ))
        AND AK=((
                        (
                                ("Patient to image") OR ("Robot to world") OR ("Robot-world")
                                OR ("Patient-image") OR (Robot*) OR (Patient) OR ("Model-based")
                                OR (Fiducial) OR ("Image-guided") OR (Target*)
                                OR ("Interaction image") OR ("Model-to-feature") OR (Spatial)
                                OR ("Image-to-robot") OR ("Hand eye") OR ("Pair point")
                                OR ("Medical Image") OR ("Image-to-physical") OR (Physical)
                                OR ("Coordinate space")
                        )
                        AND (
                                (Registration) OR (Calibration) OR (Matching) OR (Detection)
                                OR (Localization) OR (Initiali*) OR (Search) OR (Transform*) OR (Coregistration)
                        )
                        AND (
                                (Markerless) OR (Extrinsic) OR (Intrinsic) OR ("Touch-free") OR ("Touch-based")
                                OR (Frameless) OR ("Frame-based") OR ("Non-invasive") OR (marker*)
                                OR (Geometrical) OR (Feature) OR (Invasive) OR (automatic)
                                OR ("Semi-automatic")
                        )
                        AND (
                                (Surger*) OR (Procedure) OR (Surgical) OR (Surgical) OR (Intraoperative)
                                OR (Preoperative) OR (Clinical)
                        )
                        NOT ("intra-operative imaging")
        ))
)
```

Figure A.4: WOS query related to registration methods

A.4. Implementation steps

The implementation steps are not executed because this is out of scope for this master thesis. Nevertheless, each step is explained below. The steps are:

1. Perform queries, store results in Mendeley
2. Repeat and refine queries
3. Identifying and removing duplicates
4. Title-based exclusion
5. Abstract-based exclusion
6. Methodology and conclusion: are data and method trustworthy?
7. Read full-text

Excluding papers includes preserving rationales for exclusion in an overview such as Rayyan.

A.4.1. Perform queries, store results in Mendeley

Performing the queries and exclusion criteria provided a total of 1593 papers for Scopus, 3792 papers for PubMed and 133 papers for WOS. They are resulting in a total of 6,831 papers.

A.4.2. Repeat and refine queries

The discovered papers should be reviewed in this step, adopting the search query accordingly. For illustration, it is deduced from the results that multiple papers indicated no association with robot registration. Considerable results were related to biomedical research, such as non-invasive testing in the prenatal phase. Words such as ”target”, ”cell”, or ”spatial” lead to irrelevant results. Excluding these terms will lead to better results. One can also decide not to exclude them based on search terms, reducing the chance of overlooking relevant papers yet increasing workload.

A.4.3. Identifying and removing duplicates

Uploading papers in Rayyan. Followed by detecting and removing duplicates.

A.4.4. Title-based exclusion

Title-based exclusion consists of scanning the titles and excluding papers based on their title. In this step, it is required to regard the stated exclusion criteria.

A.4.5. Abstract-based exclusion

Abstract-based exclusion consists of scanning the abstracts and excluding papers based on their abstracts.

A.4.6. Methodology and conclusion: are data and method trustworthy?

The researcher reads the methodology and conclusion of the remaining papers. Consider if the data and method are trustworthy and sufficient for scientific knowledge related to the subject.

A.4.7. Read full-text

One reads all papers in full-text and determines inclusion or exclusion.

B. Experimental design

This appendix elaborates on the experimental design setup used in the benchmarking tests.

This chapter elaborates on the experimental design by providing the experimental purpose, task, experimental conditions, dimensions, setting up the robot and the metric measurements in detail.

B.1. Experimental research

The experimental research is used to perform human-factor analysis within the field of HRI for the image-to-patient registration real-world localisation process. The two design will be tested against the old system in its system, human-robot, performance.

B.1.1. Purpose

This experimental research tests the development of an improved human-robot (system) registration procedure for a sub-millimetric bone drilling robot in the lateral skull base. Before the robot can autonomously perform bone drilling, registration needs to take place to map features from one coordinate space to another coordinate space by finding a corresponding transformation matrix. Since sub-millimetric accuracy is needed for drilling the lateral skull, the registration process needs to be of sub-millimetric accuracy. According to J. Wang et al. (2020), as stated, the total accuracy should be 0.5mm or less to provide a safe procedure between the facial nerve and the chorda tympani nerve.

Moreover, the procedure should be easy to carry out and fit within the workflow of the surgeon, thus having a low workload. Since the robot will be applied in a medical environment, trust and usability are important factors for its adoption and success. Therefore, these are considered important requirements within the registration procedure.

In this experiment the performance of the system as is, will be compared to a new design in performing touch-based pair-point robot registration. The system performance is measured in terms of human localisation performance, since this directly influences the systems accuracy, the system workload; usability; and trust. The workload is measured both qualitatively and quantitatively. The usability and trust were measured solely quantitatively.

B.1.2. Task

This task differs slightly per design and can roughly be described with the following steps; steps 3 to 10 are repeated eight times by each participant:

1. Robot in starting position [315 -120 50] and heading [0 0.8829 -0.4695]
2. Start procedure registration
3. Admittance control to find position fiducial screw 1
4. Save position fiducial screw 1
5. Admittance control to find position fiducial screw 2
6. Save position fiducial screw 2
7. Admittance control to find position fiducial screw 3
8. Save position fiducial screw 3
9. Admittance control to find position fiducial screw 4
10. Save position fiducial screw 4
11. Check saved positions

B.1.3. Experimental conditions

The experimental conditions related to this project are described below.

Participant consent and draft groups

Since this is a comparison study between a baseline design and a new design there will be two groups of participants. It is best to assign participants to a group at random. However, since the previous experience with robots can have a large influence on the results, the participants will be asked to score their experience and knowledge on robots on a scale from 1-10 and the participants with higher scores (8-10) as well as medium scores (3-7) and low scores (1-3) will be divided as equally as possible between both groups. Within the high, medium and low scores the participants will be drawn at random. This will be done by putting all low scoring participants in a draft and distributing them between both groups. This will be repeated for medium and high scores. This scoring will be done when participants are asked to participate in the study. Moreover, the groups will also be distributed based on gender and age-group to make both groups as equally distributed as possible. The age groups will be distributed from 20-30, 30-40 and 40+.

They will be provided a consent form in which they are informed about the objectives of the study and how their data will be used for the study. After signing the form, the participants were asked to fill out a pre-questionnaire which consisted of the demographics and robot acquaintance questions. Depending on their age-group and rated experience, the two groups are divided as equally as possible.

Design 1: Single-mode, no feedback admittance control with two operators

This baseline design allowed a human operator to guide the robot end-effector in the translational axis and save fiducial localisation positions in cooperation with a second human operator. Operator 1 is positioned at the computer and controls the GUI. Operator 2 is arranged at the robot manipulator and performs the localisation task. They change functions after finishing the task.

Training

The training for this design is given in Appendix C.

Single-mode, no feedback admittance control

Operator 2 will perform the localisation task in this design with single-mode admittance control. Hence, only one mode is provided to the operator to perform localisation with the control loop explained in Chapter 3.6. The robot manipulator was set at a starting and reference position [315 -120 50] and heading [0 0.8829 -0.4695]. The force positional gain was set to 3.0, and the positional and rotational gains were set to [0.05 0.05].

Localisation

The localisation is executed by operator 1 with the control loop as described. Each fiducial is localised eight times. Once operator 1 has located the position of a fiducial, operator 2 should press the according switches in the GUI to save their data, as pictured in Figure B.1. When all positions are gathered, operator 1 needs to press “Check values” to check if all data is saved and export it to a CSV file.

Survey

After the task, both operators were requested to complete a survey, given in Appendix D.

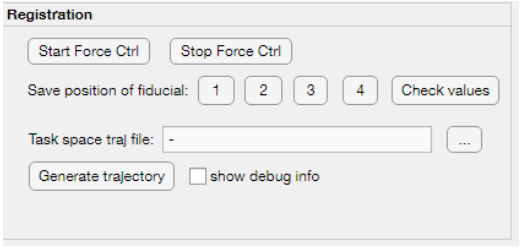


Figure B.1: The GUI presented to the operators acting with the baseline design

Design 2: Multi-modal, multi-feedback admittance control with one operator and supervisor
In this design, one operator and one supervisor execute the task. The operator was placed at the robot manipulator and conducted the admittance control localisation. The supervisor was positioned near the computer and performed checklists with the operator. They switched roles, both performing the localising task.

Training
The training for this design is given in Appendix C

Multi-modal, multi-feedback admittance control
As explained in Figure 5.7, localisation was accomplished by operating the control loop with a foot-operated interface with several buttons. The robot manipulator was set at a starting and reference position [315 -120 50] and heading [0 0.8829 -0.4695]. The gain values are given in Table 5.3

Localisation
The GUI workflow is explained in Appendix M. The localisation was performed by the operator with the control loop, as mentioned. Localising each fiducial eight times.

Survey
After the task, both participants were requested to complete a survey, given in Appendix D.

B.2. Dimensions
The dimensions of the used experimental setup is described in detail below.

B.2.1. Block
The fiducials are attached to an an area of 42.219 x 33.425 x 0.441 mm, as shown in Figure B.2. The block is secured to an adult skull-resembling wooden setup at an angle of attack for the end-effector of 62 degrees and exhibited in Figures B.3, B.4,B.6 and B.5.

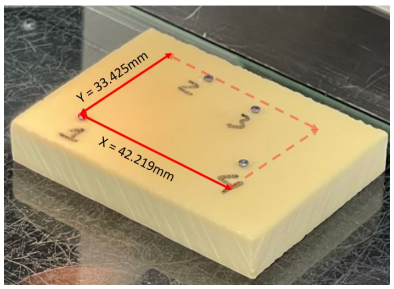


Figure B.2: Arrangement of the four fiducials on a block consisting of material that acts similar to bone during drilling. Locations are measured with a microscope and gave $F1 = [0.0; 0.0; 0.0]mm$, $F2 = [13.016; 33.425; -0.441]mm$, $F3 = [30.393; 30.065; -0.182]mm$ and $F4 = [42.219; 10.185; -0.182]mm$.

B.2.2. Skull-resembling setup
The skull-resembling setup is where the block and fiducials are attached to. It is a wooden block in an area of 21x15x7.5cm as shown in Figures B.3, B.4 and B.5. The angle of attack is resembled in Figure B.5.

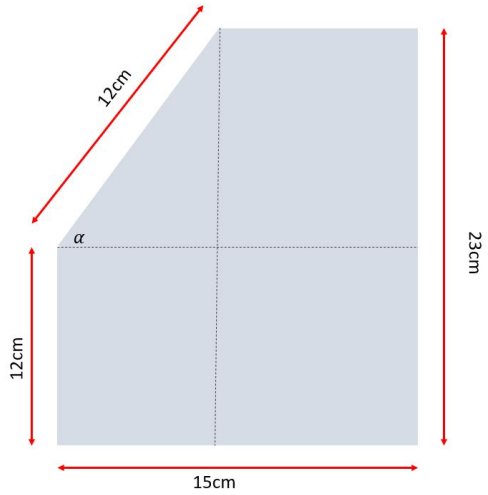


Figure B.3: The wooden skull-resembling block with a dimension of 15x21x7.5cm and a slight incline with an angle of $\alpha = 45$ degrees and is related to Figure B.5.

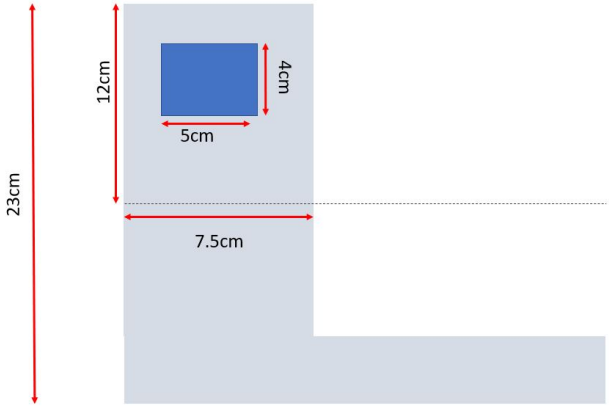


Figure B.4: The wooden skull-resembling block with a dimension of 15x21x7.5cm and the attachment of the block with fiducials on a slight incline with an angle of 45 degrees.

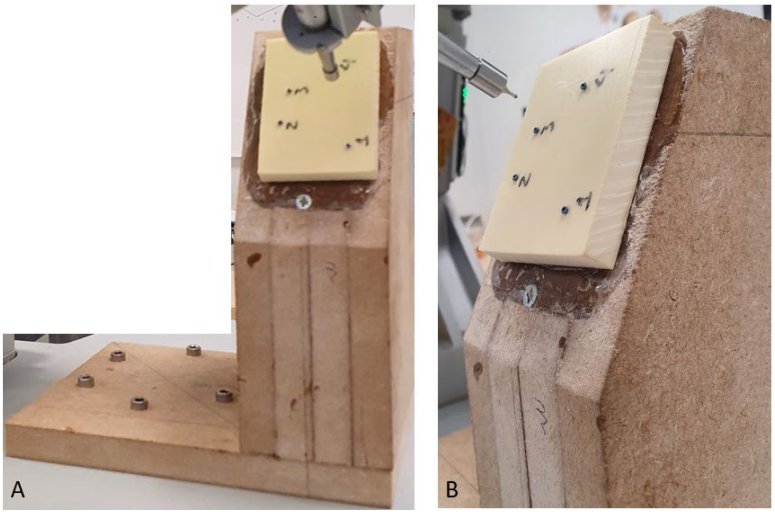


Figure B.5: Setup of fiducials on a wooden block with dimensions similar to the human head.

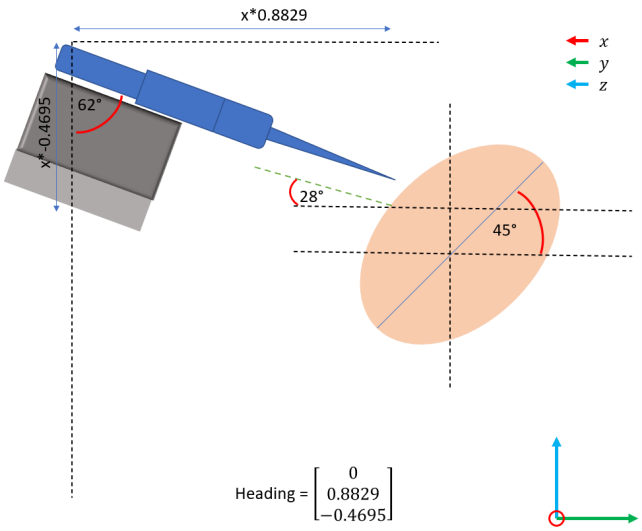


Figure B.6: A schematic overview of the optimal heading angle. Assuming that the patient lies with an angle of 45 degrees to the centre axis of the head, depicted with a dark blue line, it is given that the specific surgical path can be found with an angle of 28 degrees. This requires an angle of 62 degrees for the end-effector to match this angle and perform the best surgery. The coordinate frame is shown with x in red, y in green and z in blue. The registration bit length is given in a y-value of x*0.8829 and a z-value of x*-0.4695, with x being a scaling factor.

B.2.3. Fiducial and tip

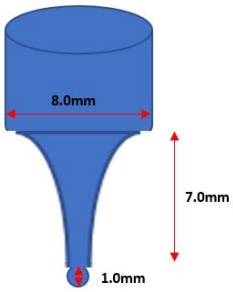


Figure B.7: The tip dimensions of the registration bit used to localise the fiducial screws.

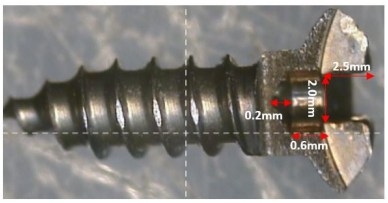


Figure B.8: The most important screw dimensions of the screws used to localise their position.

B.3. Setting up robot

The following steps are required for preparing the robot for testing:

1. Turn on the robot
2. Start Matlab
   Run startup.m in Matlab 2017
3. Load the correct model and open in Simulink
4. Select the target
   Select Target D16 (Previously D13)
5. (Re)connect to target

6. Start (RT) model
7. Setup values
   Set errorbitmasks (MMC), temp errorbitmask (MMC) and reference position (SLRT)
8. Set registration values
   Set the registration gains (SLRT) as described in Chapter 7
9. Set registration bit
   Connect the registration bit to the robot end-effector
10. Set testing block
    Connect the test block to the robot base
11. Start robot
12. Set robot to registration mode

B.4. Metric measurements

The measurements that require more details are displayed. *TRE*
In coordinate frame 1 the fiducial locations are measured with a specialised microscope. Their coordinates are:

$$FC1 = \begin{bmatrix} 0 & 13.016 & 30.393 & 42.219 \\ 0 & 33.425 & 30.065 & 10.185 \\ 0 & -0.441 & -0.182 & -0.182 \end{bmatrix} \tag{B.1}$$

A target position in this frame, a_{target}, is needed to calculate the TRE following

$$\vec{TRE}_{target} = T_P^Q b_{target} - a_{target}. \tag{B.2}$$

The chosen target is the round window, as this is a critical structure in CI surgery. From real CT data, the target position is given with respect to four fiducials placed at the CT coordinates

$$RealCT = \begin{bmatrix} 69.020 & 77.186 & 72.966 & 72.712 \\ 148.848 & 157.838 & 142.160 & 168.832 \\ -1312.142 & -1285.511 & -1291.345 & -1296.026 \end{bmatrix} \tag{B.3}$$

with corresponding target position

$$CTPT = \begin{bmatrix} 43.953 & 167.379 & -1299.922 \end{bmatrix} \tag{B.4}$$

It is concluded that the fiducial placement in the real CT image is equivalent to the fiducial placement in this experiment. Thus the transformation matrix between the real CT and the experimental setup block fiducials is calculated with the DetermineHomogeneousTransformationMatrix script. This transformation matrix is then applied to the real CT target, providing a PT location of

$$PT_{coordinate1} = \begin{bmatrix} 30.476 & 1.743 & -25.952 \end{bmatrix} \tag{B.5}$$

Now the TRE can be calculated by performing registration between frame 1 and frame 2 with the DetermineHomogeneousTransformationMatrix script. This transformation matrix is then applied to the PT of coordinate frame 1 to gather the expected PT in coordinate frame 2 based on the measured locations. To calculate the TRE from each participant the GT fiducial values are use calculate a GT PT position.

A transformation matrix is calculated between the real CT data and the GT fiducial locations for each participant. This matrix is used to calculate the expected GT PT for each participant. Then the TRE is calculated from the vector difference between the expected PT in coordinate frame 2 and the GT PT in coordinate frame 2. This is repeated for each possible combination of fiducial measurement for each

participant. Eight measurements of each fiducial gives 8 to the power of 4 different combinations.

Human workload

The human workload is measured both quantitatively and qualitatively.

Qualitative measure: NASA TLX

The NASA TLX, NASA Task Load Index, created by Hart and Staveland (1988), is a multi-dimensional survey-based measure of workload (Grier, 2015). This survey scores workload based on six sub-scales: mental demand, physical demand, temporal demand, frustration, effort, and performance. Each is scaled from 0 to 100 by marking the desired position. The scale is a 12-cm line divided into 20 equal intervals anchored by bipolar descriptors (High vs low). Thus the 21 tick marks divide the scale from 0 to 100 with an increment of 5. If a subject marks between two ticks, the value is rounded to the right tick. The definitions and rating sheet are exhibited in Figure B.9.

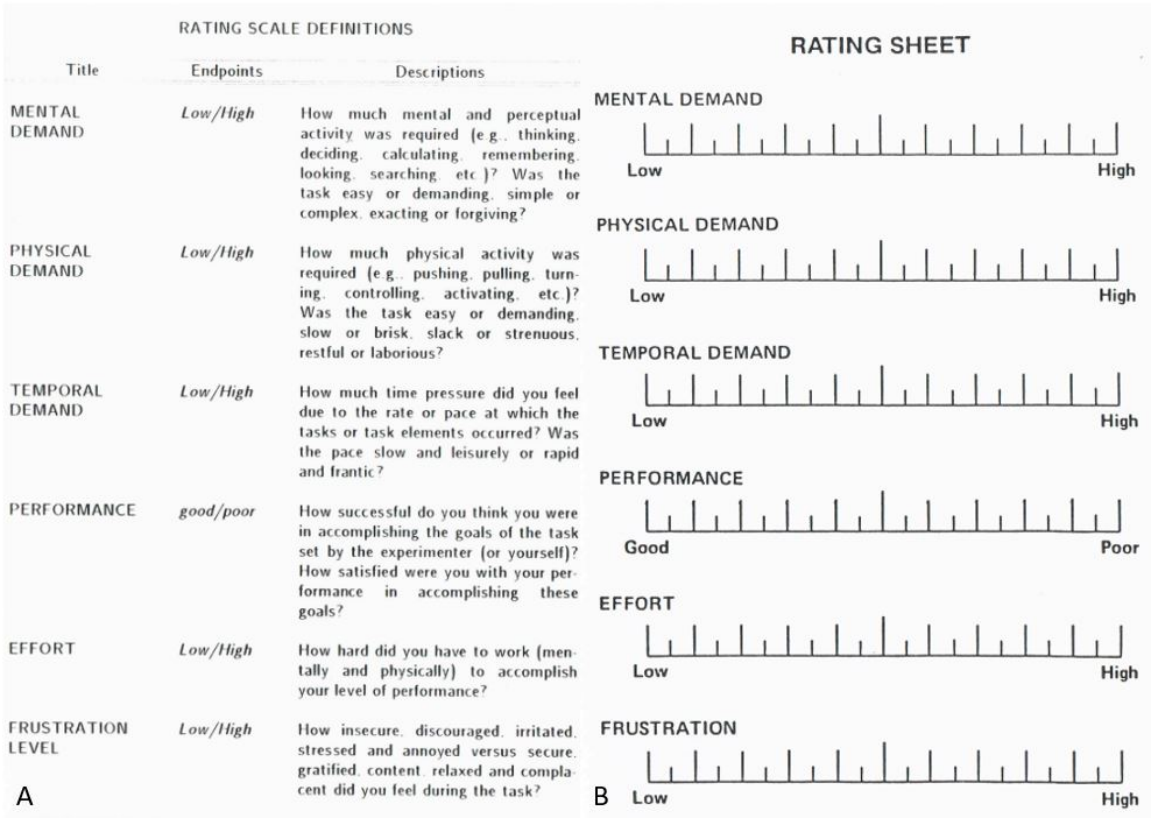


Figure B.9: The definitions (A) and scale (B) of the NASA TLX survey, created by Hart and Staveland (1988). Scaling takes place by marking each scale at the desired position. The scale is a 12-cm line divided into 20 equal intervals anchored by bipolar descriptors (High vs low). Thus the 21 tick marks divide the scale from 0 to 100 with an increment of 5.

Each sub-scale is weighted based on participant questions related to the perceived contribution to the workload. These questions contain a pair-wise comparison among the six sub-scales with scoring cards. The participants had to select the member of each pair that they thought contributed more to the total workload. Then the total number of each factor is tallied, ranging from 0 to 5. This scoring accounts for two potential sources of between-subject variability. Namely, the differences in workload definitions between subjects within a certain task and, secondly, discrepancies in the sources of workload between different tasks. From each sub-scale, the mean was calculated with

$$\text{mean} = \frac{\text{Sum of all question points}}{\text{Number of question points}} \tag{B.6}$$

These mean values were weighted through multiplication with its tally and dividing by 15, the sum of

all weights. The resulting weighted mean is the overall workload, ranging from 0 (low) to 100 (high).

Quantitative measure: Average force
The operator effort is measured quantitatively by the average force. The higher the average force, the higher the workload. The average force was calculated from the total force over the total time:

$$WL = \frac{\sum_{t_{\text{begin}}}^{t_{\text{end}}} F_t}{t_{\text{end}} - t_{\text{end}}} \tag{B.7}$$

with WL the workload for one participant, t_{end} the end-time of the localisation task and t_{begin} the beginning time, and F_t the force magnitude measurement as $|F|$ at some point in time t. Then the WL for each group of participants is analysed and compared to the WL of the other participants.

Usability
Usability analysis can be performed in various ways, such as interviews, behaviour analysis and through questionnaires. For this research, two questionnaires were employed: the Computer System Usability Questionnaire (CSUQ) and the System Usability Scale (SUS).

CSUQ
CSUQ is a 19-item survey that measures usability specifically for human-computer interaction. It is created by J. R. Lewis (1995) and holds a 0.95 overall Cronbach's alpha score. The 19 questions were scored along a bipolar 7-point Likert scale. Afterwards, participants are requested for three negative and three positive aspects of the system. The scores were summed for each participant to acquire their total score. The mean was calculated with Equation 2.3. The complete questionnaire can be found at https://garyperlman.com/quest/quest.cgi and is demonstrated in Figure B.10.

		1	2	3	4	5	6	7		NA
1. Overall, I am satisfied with how easy it is to use this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
2. It was simple to use this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
3. I can effectively complete my work using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
4. I am able to complete my work quickly using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
5. I am able to efficiently complete my work using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
6. I feel comfortable using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
7. It was easy to learn to use this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
8. I believe I became productive quickly using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
9. The system gives error messages that clearly tell me how to fix problems	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
10. Whenever I make a mistake using the system, I recover easily and quickly	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
11. The information (such as online help, on-screen messages, and other documentation) provided with this system is clear	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
12. It is easy to find the information I needed	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
13. The information provided for the system is easy to understand	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
14. The information is effective in helping me complete the tasks and scenarios	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
15. The organization of information on the system screens is clear	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
16. The interface of this system is pleasant	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
17. I like using the interface of this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
18. This system has all the functions and capabilities I expect it to have	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
19. Overall, I am satisfied with this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
		1	2	3	4	5	6	7		NA

Figure B.10: The CSUQ survey, scoring a 19-item survey on a bipolar Likert scale where each statement has to be ranked. Hence, it measures usability on a scale from 1 to 7 (1995).

SUS

The SUS is a 10-item survey designed by Brooke et al. (1996) and shown in Figure B.11. The scale ranges from 1 to 5, and these scores are normalised from 0 (poorest rating) to 4 (best rating). To calculate the participant SUS score, the odd questions are scored positively, whereas the even questions are scored negatively. These are then summed and multiplied by 2.5:

$$\begin{aligned} SUS = 2.5(20 + SUM(SUS01, SUS03, SUS05, SUS07, SUS09) \\ - \text{SUM}(SUS02, SUS04, SUS06, SUS08, SUS10)) \end{aligned} \tag{B.8}$$

The final value is a reflection of the system's usability. Its relevance is based on the distribution of all scores, similar to a grading system. A SUS score of 74 entails that the usability is 70% better than all products tested. The top 10% of the scores are above 80.3. A SUS of 68 is average and reflects a C, whereas a SUS of 51 or lower reflects the bottom 15%, an F.

	The System Usability Scale Standard Version	Strongly Disagree					Strongly Agree
			1	2	3	4	5
1	I think that I would like to use this system frequently.		o	o	o	o	o
2	I found the system unnecessarily complex.		o	o	o	o	o
3	I thought the system was easy to use.		o	o	o	o	o
4	I think that I would need the support of a technical person to be able to use this system.		o	o	o	o	o
5	I found the various functions in this system were well integrated.		o	o	o	o	o
6	I thought there was too much inconsistency in this system.		o	o	o	o	o
7	I would imagine that most people would learn to use this system very quickly.		o	o	o	o	o
8	I found the system very awkward to use.		o	o	o	o	o
9	I felt very confident using the system.		o	o	o	o	o
10	I needed to learn a lot of things before I could get going with this system.		o	o	o	o	o

Figure B.11: The SUS survey, scoring a 10-item survey on a scale where each statement has to be ranked from 1-5 (1996).

Trust

To measure human trust in robots, the Trust Perception Scale-HRI was developed by Schaefer (2016). It is applicable across all robot domains. The Trust Perception Scale-HRI is a 40-item survey with a 14-item sub-scale developed based on six dimensions, as pictured in Figure B.12. The scale was designed as a post-interaction measure to compare changes in trust across multiple conditions. Hence, the survey should be filled in instantly after the interaction. Five items have to be reverse coded, indicated in Figure B.12 with an "a". Then the items are summed and divided by the total number of items, as Equation 2.3., acquiring an overall trust score between 0 and 100%. In this measure of trust, other factors like perceived risk and reliability are included and will not be measured individually.

	0 %	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %	100 %
What % of the time will this robot be . . .											
1. Considered part of the team	O	O	O	O	O	O	O	O	O	O	O
2. Responsible	O	O	O	O	O	O	O	O	O	O	O
3. Supportive	O	O	O	O	O	O	O	O	O	O	O
4. Incompetent[a]	O	O	O	O	O	O	O	O	O	O	O
5. Dependable[b]	O	O	O	O	O	O	O	O	O	O	O
6. Friendly	O	O	O	O	O	O	O	O	O	O	O
7. Reliable[b]	O	O	O	O	O	O	O	O	O	O	O
8. Pleasant	O	O	O	O	O	O	O	O	O	O	O
9. Unresponsive[a,b]	O	O	O	O	O	O	O	O	O	O	O
10. Autonomous	O	O	O	O	O	O	O	O	O	O	O
11. Predictable[b]	O	O	O	O	O	O	O	O	O	O	O
12. Conscious	O	O	O	O	O	O	O	O	O	O	O
13. Lifelike	O	O	O	O	O	O	O	O	O	O	O
14. A good teammate	O	O	O	O	O	O	O	O	O	O	O
15. Led astray by unexpected changes in the environment	O	O	O	O	O	O	O	O	O	O	O
What % of the time will this robot . . .											
16. Act consistently[b]	O	O	O	O	O	O	O	O	O	O	O
17. Protect people	O	O	O	O	O	O	O	O	O	O	O
18. Act as part of the team	O	O	O	O	O	O	O	O	O	O	O
19. Function successfully	O	O	O	O	O	O	O	O	O	O	O
20. Malfunction[a]	O	O	O	O	O	O	O	O	O	O	O
21. Clearly communicate	O	O	O	O	O	O	O	O	O	O	O
22. Require frequent maintenance[a]	O	O	O	O	O	O	O	O	O	O	O
23. Openly communicate	O	O	O	O	O	O	O	O	O	O	O
24. Have errors[a]	O	O	O	O	O	O	O	O	O	O	O
25. Performa a task better than a novice human user	O	O	O	O	O	O	O	O	O	O	O
26. Know the difference between friend and foe	O	O	O	O	O	O	O	O	O	O	O
27. Provide feedback[b]	O	O	O	O	O	O	O	O	O	O	O
28. Possess adequate decision-making capability	O	O	O	O	O	O	O	O	O	O	O
29. Warn people of potential risks in the environment	O	O	O	O	O	O	O	O	O	O	O
30. Meet the needs of the mission/task[b]	O	O	O	O	O	O	O	O	O	O	O

	0 %	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %	100 %
31. Provide appropriate information[b]	O	O	O	O	O	O	O	O	O	O	O
32. Communicate with people[b]	O	O	O	O	O	O	O	O	O	O	O
33. Work best with a team	O	O	O	O	O	O	O	O	O	O	O
34. Keep classified information secure	O	O	O	O	O	O	O	O	O	O	O
35. Perform exactly as instructed[b]	O	O	O	O	O	O	O	O	O	O	O
36. Make sensible decisions	O	O	O	O	O	O	O	O	O	O	O
37. Work in close proximity with people	O	O	O	O	O	O	O	O	O	O	O
38. Tell the truth	O	O	O	O	O	O	O	O	O	O	O
39. Perform many functions at one time	O	O	O	O	O	O	O	O	O	O	O
40. Follow directions[b]	O	O	O	O	O	O	O	O	O	O	O

[a]Represents the reverse coded items for scoring
[b]Represents the 14 item sub-scale items

Figure B.12: The Trust Perception Scale-HRI scores each statement in a percentage from 0 to 100 with increments of 10%. The superscript "a" indicates scores requiring reverse coding, and "b" indicate the 14-item sub-scale (2016).

C. Training, roles and responsibilities

This Appendix describes the training procedure for participants performing with the new or old registration system.

C.1. Training

The training procedure consists of multiple steps for both designs:

1. Procedure and general terminology
2. Roles and responsibilities
3. Walk through workflow
4. Safety buttons
5. Robot errors
6. Case-studies

Each step is explained generally and the differences per design are higlighted.

C.1.1. Procedure and general terminology

For operators and supervisors to work correctly with the robot, it is important for them to understand what the procedure is and what it means for the accuracy of the surgery.

First the purpose of the robot should be explained: *The robot will be used for CI surgery. In this surgery the robot will be used for drilling the skull and finally also for inserting an electrode into the cochlea. This surgery needs sub-millimeter accuracy because there are a facial and taste nerve that are inside the surgical area and can harm the patient when those are damaged.*

Thereafter a broad outline of registration should be explained: *Before the robot can perform a surgery, the head of a patient is scanned in 3D and a surgical plan is made. However, when the patient and robot are inside the operating room, we still need to tell where the patient exactly is. So the robot knows where to start its pre-programmed plan, which should be at exactly the same x,y,z coordinate. If the robot is off by 1cm, then it could be a large problem during the surgery. Therefore telling the robot where the patient is and where to start is very important. This is called registration.*

Thirdly the registration should be explained in more detail: *One way to perform this registration is by finding the same markers in both the real world as well as the CT image, which is called pair-point registration since you match pairs of different points, called markers. The markers are here four what they call fiducial screws inside the patients head near the surgical area.*

Then the task at hand should be described. This differs per design.

Old design

You will control the robot by moving it with your hand, holding the registration bit. The robot will move towards a reference position if no force is applied. Also, when the force is applied, it will slowly move towards a safe position. You must move the registration bit towards the four fiducials and then save its position to the robot through the GUI. Therefore, there will be a second operator at the computer to

perform the saving of the positions. The operator near the robot, needs to hold the robot end-effector in position until the position it is saved. The order does not matter, but you should find each fiducial eight times while moving in between.

New design
To tell the robot the location of these markers in the real world, we use a registration bit with known size and dimensions and the end of the robot arm, called end-effector, to touch the fiducial markers. This will give the robot the coordinates in x, y and z of these fiducials. The movement of the robot arm is done with a so-called admittance control where the force that is applied at the end-effector is translated to a movement of the robot arm. So after moving the arm to the fiducial marker position, the position needs to be saved to the robot. This moving is done by one operator while another person is located near the computer to be in charge of the GUI. You will control the robot by moving it with your hand, holding the registration bit. However, the robot will not move unless you press the foot pedal. This pedal works just like one in a car, where you can increase and decrease the speed by pressing the pedal more or less. When you let go of the pedal, the robot will go into something called fixed mode, here it will keep still for 10 seconds and it can not be moved by exerting a force. After 10 seconds however, it will move slowly back to a reference position for safety, but then still you can not move the robot arm by applying a force to it. So only when you press the pedal the movement of the robots end-effector with admittance control is possible. Then when you arrive at the wanted position, you can let go of the foot pedal to go into fixed mode and then press the left top-button for saving the position to the robot. Moreover, when you think the orientation of the fiducial is off and you want to adjust the angle of the robot end-effector, you can press the right top button to change its orientation. These different ways to control the robot are called modes, we have the fixed, registration, rotation and translation modes. By using the foot pedal and admittance control you have to find the four fiducials. You should find each fiducial eight times and move in between.

Now that the operator task is clear, the checklist and terminology should also be made clear, which also differs per design.

Old design
In summary, you have the robot arm holding a registration bit and four fiducial screws inside the patient's head that require to be located with admittance control. The robot also has a reference position where it will move, so you must hold the registration bit while the other operator saves the data. Another important thing is that the force sensor used for admittance control can have some bias or trend that must be zeroed out when this happens. This is what we call unbiasing.

New design
So in summary, you have a foot pedal which can provide you with different modes of control, the robot arm holding a registration bit and four fiducial screws inside the patients head that need to be located with admittance control. The robot also has a reference position where it will move towards if it is not touched for longer then 10 seconds and you can save the data with the foot pedal. One more thing that must be noted is that the force sensor that is being used for admittance control can have some bias or drift that has to be zeroed out when this happens, this is what we call unbiasing.

C.1.2. Roles and responsibilities

Since the tasks are clear, the roles and responsibilities must be explained to the operators. This differs per design.

Old design
There are two operators involved in the registration process: operator 1 near the robot and operator 2. near the computer

Operator 1

The operator is positioned near the robot manipulator and performs the full registration process. This includes the use of admittance control to position the robot end-effector to find the fiducial locations, saving their positions to the system, repeating it accordingly, checking for safety, etc. The tasks and responsibilities are described below.

Tasks:

- Perform robot localisation with admittance control
- Localise the fiducial screws with high-accuracy
- Hold the end-effector at fiducial position
- Communicate with operator 2 when to save the position
- Pressing the emergency button in case of emergency
- Check patient safety

Tasks:

- Perform robot localisation with admittance control
- Correct and safe use of admittance control
- Prevent robot manipulator damage
- Accurate fiducial localisation
- Correct performing of registration process
- Keep track of safety near robot
- Clear communication with operator 2

Operator 2

Operator 2 is positioned near the computer, and cannot reach the robot manipulator. The main task of this operator is finalising the fiducial localisation by saving their position by pressing a button on the GUI. Its tasks and responsibilities are:

Tasks:

- Save fiducial locations
- Save data at the end
- Stop admittance control when needed
- Stop the robot when needed
- Check patient safety
- Re-starting robot when errors occur

Responsibilities:

- Accurate saving of fiducial positions
- Keep track of system improvements and damage
- Operator safety
- Patient safety
- Clear communication with operator 1

New design

There are two humans involved in the registration process: the operator and the supervisor.

Operator

The operator is positioned near the robot manipulator and performs the full registration process. This includes the use of admittance control to position the robot end-effector to find the fiducial locations, saving their positions to the system, repeating it accordingly, checking for safety, etc. The tasks and responsibilities are described below.

Tasks:

- Perform robot localisation with admittance control
- Localise the fiducial screws with high-accuracy
- Save the fiducial positions

- Evaluate results with supervisor and iterate process when needed
- Pressing the emergency button in case of emergency
- Check patient safety
- Perform checklists with supervisor

Tasks:

- Perform robot localisation with admittance control
- Correct and safe use of admittance control
- Prevent robot manipulator damage
- Accurate fiducial localisation
- Correct performing of registration process
- Keep track of safety near robot
- Clear communication with supervisor

Supervisor

The supervisor is located near the computer and will supervise the operator and be responsible for filling in the checklists and evaluating the results to make sure the registration is performed safely and with high accuracy. The tasks and responsibilities are described below. *Tasks:*

- Perform checklists together with operator
- Review results together with operator
- Save data at the end
- Stop admittance control when needed
- Stop the robot when needed
- Check patient safety
- Re-starting robot when errors occur

Responsibilities:

- Correct filling of checklists
- Evaluation of registration performance
- Keep track of system improvements and damage
- Operator safety
- Patient safety
- Clear communication with operator

C.1.3. Walk through workflow

As mentioned, it is assumed that the setting up of the robot such as setting the starting variables is already done. However, these steps can be found in Section C.2 if needed. The first step is the trainer explaining the procedure step-by-step. This means walking through the GUI, with the workflow as explained before, and using the admittance control and foot-pedal to control the robot end-effector in case of new design and using admittance control in case of old design. Then the operator and the supervisor perform the walk through themselves while the trainer observes. This is repeated until the trainer feels confident in their abilities.

C.1.4. Safety buttons

It is very important that during the training the safety buttons are clearly presented and tested by both the supervisor as well as the operator. Also the *Stop robot* button in the GUI for the new design. This should be tested at multiple stages in the procedure.

C.1.5. Robot errors

Since it is assumed that the operator and supervisor are no technical personnel involved in the robot mechanics, only errors with respect to the robot state are provided here. If errors are outside these

solutions, it should be handed to a technician. The correct robot state for registration is robot state 6. If any other robot state occurs, the possible solutions in order are shown in Table C.1.

Table C.1: Robot state and possible solutions to get the robot state to 6 for the registration process

Robot state	Solution
-1	1. Re-start the model 2. Re-set bitmasks
0	1. Reset EM
1	1. Re-start Robot 2. Reset EM
4	1. Start Robot 2. Re-start Robot 3. Reset EM
5	1. Set to registration mode 2. Re-start robot

C.1.6. Case-studies

For an elaborate training different case-studies can be practiced by the operator and supervisor. Possible case-study scenarios that can be practiced are:

- Forgetting to unbias
- Robot is moved outside workspace
- Fiducial is placed at a different angle
- Only three fiducials are placed
- Registration bit is not mounted
- Pedal is missing
- Registration mode is not working

C.2. Setting up the robot

Before registration can start, the robot should be prepared. The workflow is described as below.

1. Turn on the robot
2. Start Matlab
   Run startup.m in Matlab 2017
3. Load the correct model and open in Simulink
4. Select the target
   Select Target D16 (Previously D13)
5. (Re)connect to target
6. Start (RT) model
7. Set-up values
   Set errorbitmasks, temp errorbitmask and reference position
8. Set registration values
   Set the registration gains as described in Table 2.1
9. Set registration bit
   Set the registration gains as described in Table 2.1
10. Start robot
11. Set robot to registration

D. Survey

This Appendix describes the survey scoring the workload, usability, and trust from the participants.

The survey consists of six sections: description, trust, CSUQ, SUS, NASA TLX and end.

D.1. Description

This part welcomes the participant and asks for the participant and group number.

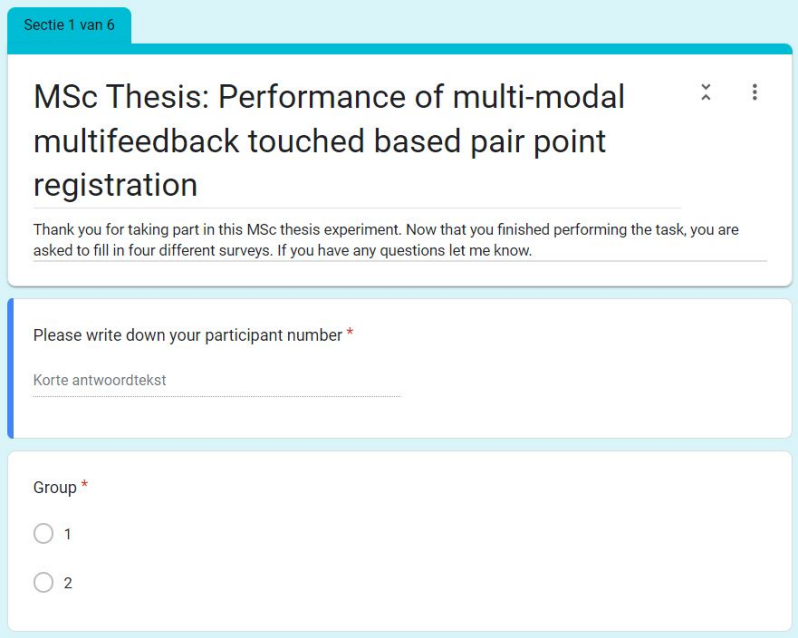


Figure D.1: The first section of the survey conducted of the participant and group number.

D.2. Trust

This part measures the Trust perception scale-HRI through rating multiple statements from 0 to 100%.

Figure D.2: Section two of the survey consisting of part one of the trust perception scale-HRI survey. The scale goes from 0 to 100% and scrolling is needed which is clearly explained to the participants.

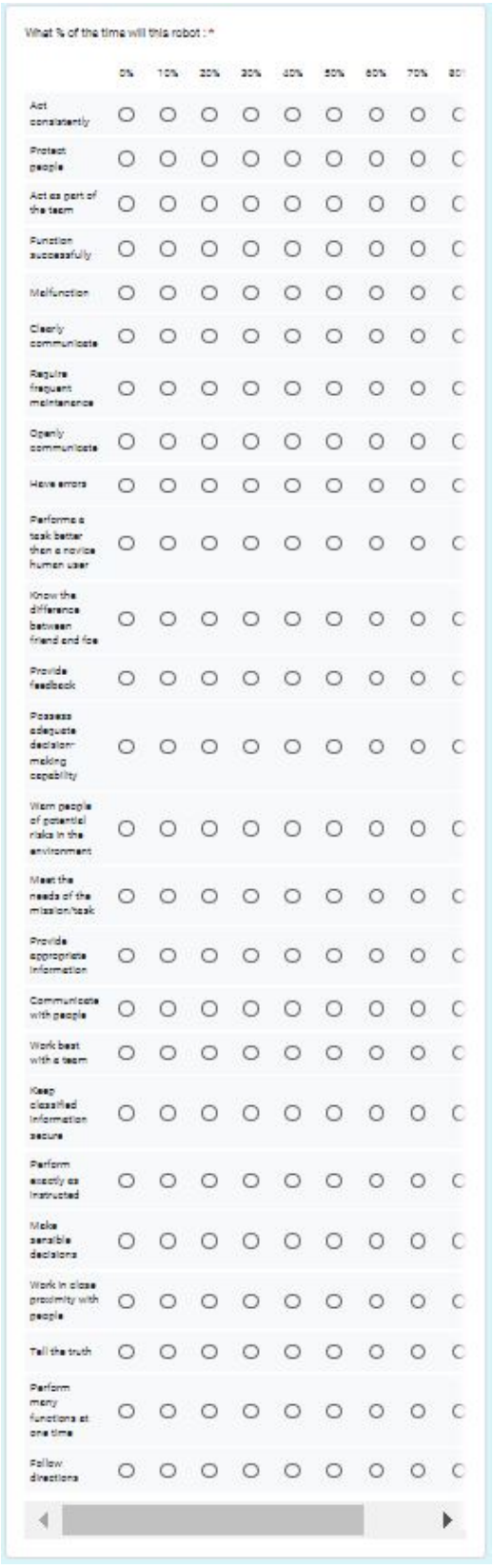


Figure D.3: Section two of the survey consisting of part two of the trust perception scale-HRI survey. The scale goes from 0 to 100% and scrolling is needed which is clearly explained to the participants.

D.3. CSUQ

This part shows the 19-question CSUQ to the participant.

Usefulness, Satisfaction, and Ease of Use Questionnaire

In this survey, please fill in how well you agree or disagree with the statement on a scale from 1 till 7. If the question is not applicable, please fill NA. Then for each question, there is an option to fill in additional comments. Finally, you will be asked to score the system with three negative and three positive aspects.

*

	Strongly disagree 1	2	3	4	5	6	Strongly agree 7	NA
1. Overall, I am satisfied with how easy it is to use this system	○	○	○	○	○	○	○	○
2. It was simple to use this system	○	○	○	○	○	○	○	○
3. I can effectively complete my work using this system	○	○	○	○	○	○	○	○
4. I am able to complete my work quickly using this system	○	○	○	○	○	○	○	○
5. I am able to efficiently complete my work using this system	○	○	○	○	○	○	○	○
6. I feel comfortable using this system	○	○	○	○	○	○	○	○
7. It was easy to learn to use this system	○	○	○	○	○	○	○	○
8. I believe I became productive quickly using this system	○	○	○	○	○	○	○	○
9. The system gives error messages that clearly tell me how to fix problems	○	○	○	○	○	○	○	○
10. Whenever I make a mistake using the system, I recover easily and quickly	○	○	○	○	○	○	○	○

Figure D.4: Section three of the survey consisting of the 19-question CSUQ on a likert scale from 0 to 7.

Figure D.5: Section three of the survey consisting of the 19-question CSUQ on a likert scale from 0 to 7.

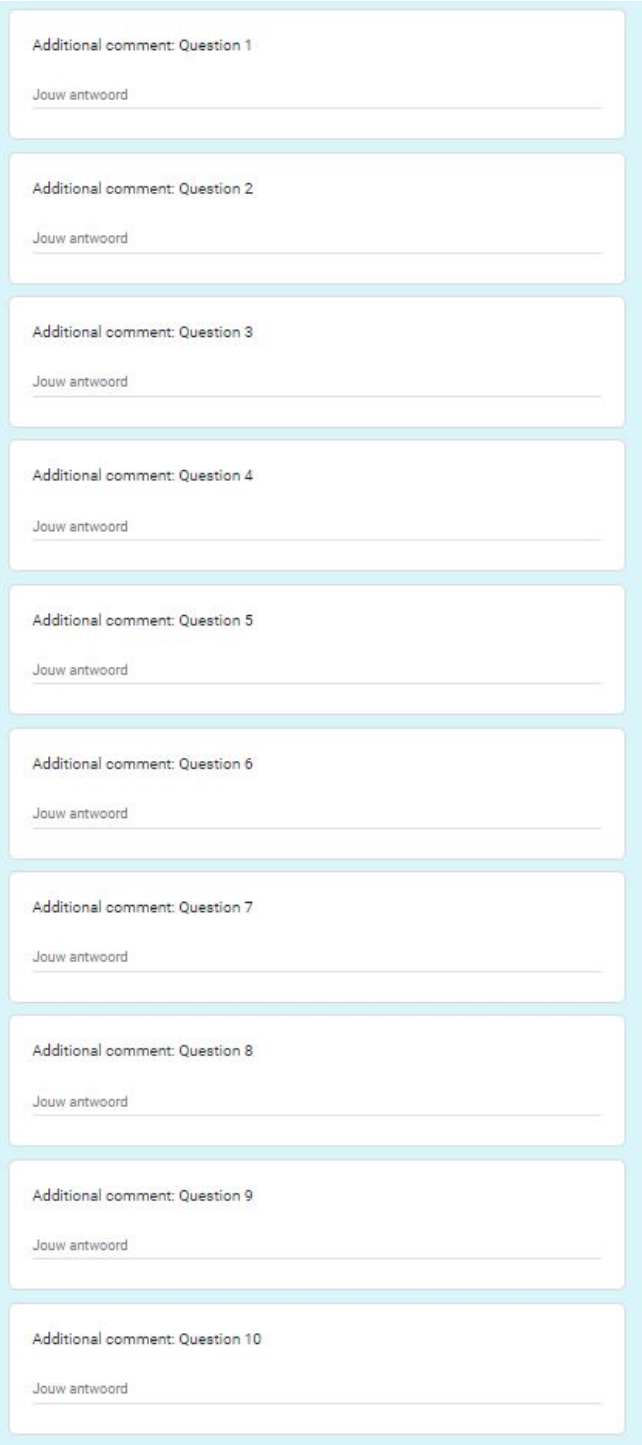


Figure D.6: Section three of the survey consisting of the 19-question CSUQ, their additional comments and negative and positive aspects

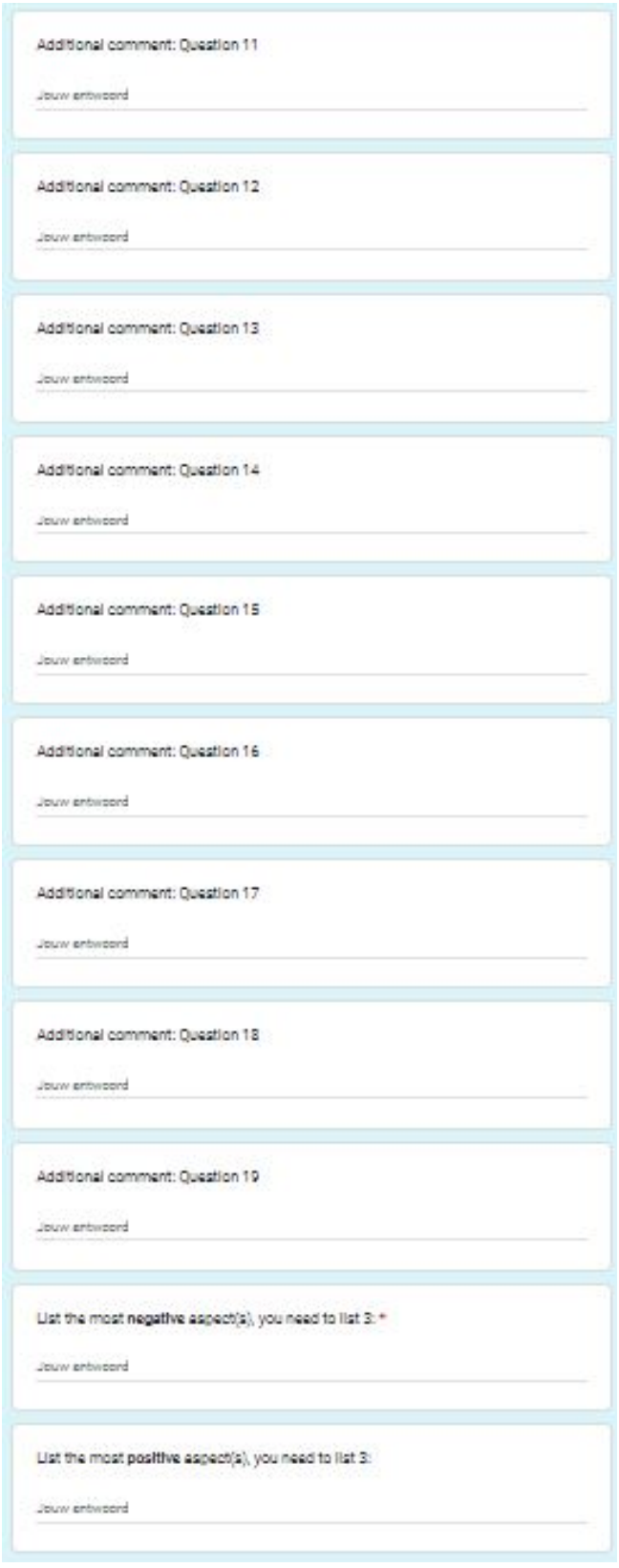


Figure D.7: Section three of the survey consisting of the 19-question CSUQ, their additional comments and negative and positive aspects

D.4. SUS

This part shows the 10-question SUS to the participant.

System Usability Scale (SUS)

In this survey please fill in how well you agree or disagree with the statement on a scale from 1 till 5.

*

	Strongly disagree 1	2	3	4	Strongly agree 5
I think that I would like to use this system frequently	○	○	○	○	○
I found the system unnecessarily complex	○	○	○	○	○
I thought the system was easy to use	○	○	○	○	○
I think that I would need the support of a technical person to be able to use this system	○	○	○	○	○
I found the various functions in this system were well integrated	○	○	○	○	○
I thought there was too much inconsistency in this system	○	○	○	○	○
I would imagine that most people would learn to use this system very quickly	○	○	○	○	○
I found the system very cumbersome to use	○	○	○	○	○
I felt very confident using the system	○	○	○	○	○
I needed to learn a lot of things before I could get going with this system	○	○	○	○	○

Figure D.8: Section four of the survey consisting of the 10-question SUS on a likert scale from 0 to 5.

D.5. NASA TLX

This part refers to the NASA TLX online survery and requests participant to summarise their scores.

Figure D.9: Section five of the survey consisting of the NASA TLX survey. The survey links through the NASA TLX survey. The scores have to be filled in afterwards.

D.6. End

This part thanks the participant and saves the data.

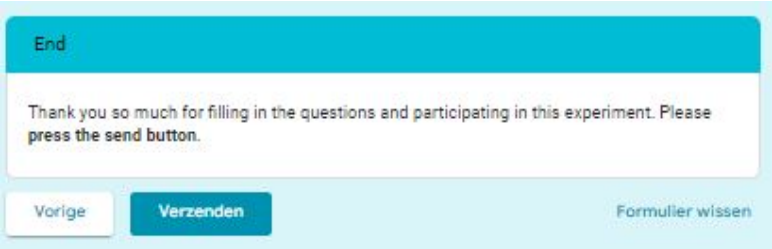


Figure D.10: Section six of the survey, finishing the survey.

E. Ethics

This Appendix describes the ethical documents assuring ethical work.

This experiment involves humans, thus requesting ethical considerations. Consequently, a consent form, a data management plan and a checklist for human research were used and approved by the TU Delft ethics committee.

E.1. Consent form

The consent form request an active consent of the participant.

Consent form

You are being invited to participate in a research study titled "Performance of multi-modal multifeedback touched based pair point registration". This study is being done by Susanna Halman in supervision of Luka Peternel from the TU Delft and in collaboration with Eindhoven Medical Robotics with supervision of Arash Arjmandi.

Purpose

The purpose of this research study is comparing accuracy, workload, usability and trust of participants with different systems in completing the task of touched based pair point localization with robot force control.

Procedure and tasks

During this experiment, you will perform fiducial localization in the robot space by positioning the robot end-effector with force control. Thereafter, you will be asked to fill in three different surveys consisting of different types of questions on a scale. This will take you approximately 60-90 minutes to complete. Before this, you will be given some training with the robot, task and the safety procedures so that you understand your role, task and know how to respond if something happens during this experiment.

Data management and privacy

The direct personal information [name] will be linked to a participant number inside an encrypted file and locked away safely.

Your indirect personal data: age-group, gender and previous knowledge on robotics systems, will be used for dividing the participants over different groups and getting the average demographic information for each group. Moreover, it can be used to find relations between test results and age-group or previous knowledge on robotic systems.

This direct and indirect personal data is also called the personally identifiable information (PII). The rest of the test data, the personally identifiable research data (PIRD), will be anonymized and will be used for scoring the different designs based on different metrics and this will be included in the Thesis report. Your results or input quotes can be quoted anonymously for thesis reporting. The anonymized data will be shared with Eindhoven Medical Robotics and TU Delft for more research if found necessary after the project is finished. The data will not be published publicly.

Potential risks and mitigations

During this experiment you will be working with a robot manipulator that can cause hazardous situations. The design has many safety measures such as automatic brakes, emergency stop buttons and limits to robot speed and the amount of force it can exert which limits the risks to a minimum. These emergency stop buttons and other mitigations will be discussed during the training procedure. You are also free to stop or pause the experiment at any time when you feel uncomfortable.

This experiment is a collaboration between the TU Delft and Eindhoven Medical robotics and in no way can either be held responsible or liable for any injuries that might occur.

Due to covid-19 the experiment the researcher is obligated to report it to the participant when feeling sick and then rescheduling has to take place.

Participants rights

Your participation in this study is entirely voluntary and you can withdraw at any time. You are free to omit any questions. You are free to request a removal or rectification of your personal data before the experiments start, whereafter the data will be removed automatically.

Participants obligations

In participating for this experiment, you have the obligations to follow the training procedure. Moreover, no pictures are allowed to be taken and sensitive information should not be shared. Due to covid-19 it is also obligatory for the participant that if the participant is feeling unwell, they should report this and rescheduling of the experiment should take place.

Contact details

For further questions or remarks you can contact Susanna Halman at s.t.halman@student.tudelft.nl or +316-83213582.

Figure E.1: Part one of the consent form with the written information.

PLEASE TICK THE APPROPRIATE BOXES	Yes	No
A: GENERAL AGREEMENT – RESEARCH GOALS, PARTICPANT TASKS AND VOLUNTARY PARTICIPATION		
1. I have read and understood the study information dated [/ /], or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	☐	☐
2. I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	☐	☐
3. I understand that taking part in the study involves fiducial localization in the robot space by positioning the robot end-effector with force control and filling in related surveys. I also understand that this means my anonymized data will be used in the reporting of the thesis findings and shared with the aforementioned parties (See Data management and privacy).	☐	☐
4. I understand that the study will end [30/09 /22] and that my participation will take place and end at [/ /].	☐	☐
B: POTENTIAL RISKS OF PARTICIPATING (INCLUDING DATA PROTECTION)		
5. I understand that taking part in the study involves the risk of physical discomfort or injury. I understand that these risks will be mitigated by multiple safety measures such as safety brakes, emergency stop buttons, and design limits. I also understand I am allowed to pause or stop the experiment as I feel the need or feel uncomfortable without having to provide a reason.	☐	☐
6. I understand that taking part in the study also involves collecting specific personally identifiable information (PII): name, age-group, gender and associated personally identifiable research data (PIRD) previous knowledge on robots, average localization error, end-effector position over time, produced force over time, SUS results, USE results, NASA TLX results and Trust Perception Scale-HRI results with the potential risk of my identity being revealed.	☐	☐
7. I understand that the following steps will be taken to minimize the threat of a data breach, and protect my identity in the event of such a breach: anonymization of the data, encrypting personal data and limited access to data.	☐	☐
8. I understand that personal information collected about me that can identify me: name, age-group and gender, will not be shared beyond the involved team consisting of Arash Arjmandi, Luka Peternel and Susanna Halman.	☐	☐
10. I understand that I am not allowed to take pictures or share information with people outside of the involved team as aforementioned.	☐	☐
11. I understand that I am obligated to follow a training before taking part in the experiment.	☐	☐
12. I understand that rescheduling due to covid-19 can take place in this experiment and when I am feeling sick, I am obligated to report this to the experimental researcher.	☐	☐
13. I understand that Eindhoven Medical Robotics or TU Delft cannot be held liable in any way if an injury occurs.	☐	☐

Figure E.2: Part two of the consent form where participants have to provide active consent

PLEASE TICK THE APPROPRIATE BOXES	Yes	No
C: RESEARCH PUBLICATION, DISSEMINATION AND APPLICATION		
14. I understand that after the research study the de-identified information I provide will be used for thesis reporting.	☐	☐
15. I agree that my responses, views or other input can be quoted anonymously in research outputs	☐	☐
D: (LONGTERM) DATA STORAGE, ACCESS AND REUSE		
16. I give permission for the de-identified data previous knowledge on robotic systems, age-group, gender average localization error, end-effector position over time, produced force over time, SUS results, USE results, NASA TLX results and Trust Perception Scale-HRI results that I provide to be archived and shared with Eindhoven Medical Robotics so it can be used for future research and learning.	☐	☐

Signatures

_____ _____ _____

Name of participant Signature Date

I, as researcher, have accurately shown the information sheet to the potential participant and gave them time to read through it thoroughly and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

_____ _____ _____

Name of Researcher Signature Date

Study contact details for further information: [*Susanna Halman, +31683213582, s.t.halman@student.tudelft.nl*]

Figure E.3: Part two of the consent form where participants have to provide active consent

E.2. HREC Checklist

This thesis requires a Human Research Ethics Committee (HREC) approval. This checklist includes research information and an extensive risk assessment and mitigation plan.

Delft University of Technology
HUMAN RESEARCH ETHICS
CHECKLIST FOR HUMAN RESEARCH
(Version January 2022)

IMPORTANT NOTES ON PREPARING THIS CHECKLIST

1. An HREC application should be submitted for every research study that involves human participants (as Research Subjects) carried out by TU Delft researchers
2. Your HREC application should be submitted and approved **before** potential participants are approached to take part in your study
3. All submissions from Master's Students for their research thesis need approval from the relevant Responsible Researcher
4. The Responsible Researcher must indicate their approval of the completeness and quality of the submission by signing and dating this form OR by providing approval to the corresponding researcher via email (included as a PDF with the full HREC submission)
5. There are various aspects of human research compliance which fall outside of the remit of the HREC, but which must be in place to obtain HREC approval. These often require input from internal or external experts such as Faculty Data Stewards, Faculty HSE advisors, the TU Delft Privacy Team or external Medical research partners.
6. You can find detailed guidance on completing your HREC application here
7. Please note that incomplete submissions (whether in terms of documentation or the information provided therein) will be returned for completion **prior to any assessment**
8. If you have any feedback on any aspect of the HREC approval tools and/or process you can leave your comments here

I. Applicant Information

PROJECT TITLE:	**MSc thesis: Performance of multi-modal multifeedback touched based pair point registration**
Research period: *Over what period of time will this specific part of the research take place*	**Total:** **01-03-2022 – 30-09-2022** **Specific:** **08-08-2022 – 30-09-2022**
Faculty:	**Mechanical, Maritime and Materials Engineering**
Department:	**Cognitive Robotics**
Type of the research project: *(Bachelor's, Master's, DreamTeam, PhD, PostDoc, Senior Researcher, Organisational etc.)*	**Master Thesis**
Funder of research: *(EU, NWO, TUD, other – in which case please elaborate)*	**X**
Name of Corresponding Researcher: *(If different from the Responsible Researcher)*	**Susanna Halman**
E-mail Corresponding Researcher: *(If different from the Responsible Researcher)*	**s.t.halman@student.tudelft.nl**
Position of Corresponding Researcher: *(Masters, DreamTeam, PhD, PostDoc, Assistant/ Associate/ Full Professor)*	**Master student**
Name of Responsible Researcher: ***Note:*** *all student work must have a named Responsible Researcher to approve, sign and submit this application*	**Luka Peternel**
E-mail of Responsible Researcher: *Please ensure that an institutional email address (**no Gmail, Yahoo, etc.**) is used for all project documentation/ communications including Informed Consent materials*	**l.peternel@tudelft.nl**
Position of Responsible Researcher : *(PhD, PostDoc, Associate/ Assistant/ Full Professor)*	**Assistant Professor**

II. Research Overview

NOTE: *You can find more guidance on completing this checklist here*

a) Please summarise your research very briefly (100-200 words)

What are you looking into, who is involved, how many participants there will be, how they will be recruited and what are they expected to do?

Add your text here – (please avoid jargon and abbrevations)
During this experiment, two groups of each 16 humans will perform position localization of four different positions of a robot end-effector with force control in groups of two. This means applying a force whereafter the robot will move to the position. Each group will use a different design and those designs will be compared to each other in terms of accuracy, workload, usability and trust. These results will be gathered through surveys and data from the robot. Each position will have to be located 20 times per participant. As mentioned, the task will be performed in groups of two where their roles (either supervisory or executing) will switch during the experiment. Before the procedure, the participants will be trained in using the system. The groups will be divided based on their age, gender and prior knowledge on robots and both groups will be made as similar as possible in demographic metrics.

b) **If your application is an additional project** related to an existing approved HREC submission, please provide a brief explanation including the existing relevant HREC submission number/s.

Add your text here – (please avoid jargon and abbrevations)

c) **If your application is a simple extension of, or amendment to,** an existing approved HREC submission, you can simply submit an HREC Amendment Form as a submission through LabServant.

III. Risk Assessment and Mitigation Plan

NOTE: *You can find more guidance on completing this checklist here*

Please complete the following table in full for all points to which your answer is "yes". Bear in mind that the vast majority of projects involving human participants as Research Subjects also involve the collection of **Personally Identifiable Information (PII)** and/or **Personally Identifiable Research Data (PIRD)** which may pose potential risks to participants as detailed in Section G: Data Processing and Privacy below.

To ensure alignment between your risk assessment, data management and what you agree with your Research Subjects you can use the last two columns in the table below to refer to specific points in your Data Management Plan (DMP) and Informed Consent Form (ICF) – **but this is not compulsory**.

It's worth noting that **you're much more likely to need to resubmit your application if you neglect to identify potential risks**, than if you identify a potential risk and demonstrate how you will mitigate it. If necessary, the HREC will always work with you and colleagues in the Privacy Team and Data Management Services to see how, if at all possible, your research can be conducted.

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT – what risks could arise?** *Please ensure that you list ALL of the actual risks that could potentially arise – do not simply state whether you consider any such risks are important!*	**MITIGATION PLAN – what mitigating steps will you take?** *Please ensure that you summarise what actual mitigation measures you will take for each potential risk identified – do not simply state that you will e.g. comply with regulations.*	**DMP**	**ICF**
A: Partners and collaboration						
1. Will the research be carried out in collaboration with additional organisational partners such as: • One or more collaborating research and/or commercial organisations • Either a research, or a work experience internship provider[1] *[1] If yes, please include the graduation agreement in this application*	**X**		**In this project, a collaboration is established with Eindhoven Medical Robotics. This company provides the technology with which implementations are made and testing takes place. Most testing will take place among company employees. The results are related to the robot performance.**	First of all, the data results will be anonymized and only anonymized information will be shared with the company. Therefore, the company can not filter individual employee or participant performance in any way from the data. In addition the results are not individually important for the experiment conclusion or a reflection of their employee performance in any way. A graduation agreement was signed by three involved parties at the beginning of the master thesis. With respect to data storage the data will be stored by the student during the project and shared with the company after the project with a CC0 license. The company will store the data on an encrypted file only accessible to employees.		
2. Is this research dependent on a Data Transfer or Processing Agreement with a collaborating partner or third party supplier? *If yes please provide a copy of the signed DTA/DPA*		**X**				
3. Has this research been approved by another (external) research ethics committee (e.g.: HREC and/or MREC/METC)? *If yes, please provide a copy of the approval (if possible) and summarise any key points in your Risk Management section below*		**X**				
B: Location						
4. Will the research take place in a country or countries, other than the Netherlands, within the EU?		**X**				
5. Will the research take place in a country or countries outside the EU?		**X**				
6. Will the research take place in a place/region or of higher risk – including known dangerous locations (in any country) or locations with non-democratic regimes?		**X**				

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT – what risks could arise?** *Please ensure that you list ALL of the actual risks that could potentially arise – do not simply state whether you consider any such risks are important!*	**MITIGATION PLAN – what mitigating steps will you take?** *Please ensure that you summarise what actual mitigation measures you will take for each potential risk identified – do not simply state that you will e.g. comply with regulations.*	**DMP**	**ICF**
C: Participants						
7. Will the study involve participants who **may** be vulnerable and possibly (legally) unable to give informed consent? (e.g., children below the legal age for giving consent, people with learning difficulties, people living in care or nursing homes,).		**X**				
8. Will the study involve participants who **may** be vulnerable under specific circumstances and in specific contexts, such as victims and witnesses of violence, including domestic violence; sex workers; members of minority groups, refugees, irregular migrants or dissidents?		**X**				
9. Are the participants, outside the context of the research, in a dependent or subordinate position to the investigator (such as own children, own students or employees of either TU Delft and/or a collaborating partner organisation)? *It is essential that you safeguard against possible adverse consequences of this situation (such as allowing a student's failure to participate to your satisfaction to affect your evaluation of their coursework).*	**X**		**Reputational loss of employees or fellow students can take place if they do not take part.**	The data will be anonymized to prevent that the data can be linked to individual people. Additionally, in the explicit consent it is ensured that the participants are provided with sufficient information to make an informed decision on whether or not to participate. Moreover, they are provided with the ability to stop participation from the project at any time without needed explanation.		
10. Is there a high possibility of re-identification for your participants? (e.g., do they have a very specialist job of which there are only a small number in a given country, are they members of a small community, or employees from a partner company collaborating in the research? Or are they one of only a handful of (expert) participants in the study?		**X**				
D: Recruiting Participants						
11. Will your participants be recruited through your own, professional, channels such as conference attendance lists, or through specific network/s such as self-help groups		**X**				
12. Will the participants be recruited or accessed in the longer term by a (legal or customary) gatekeeper? (e.g., an adult professional working with children; a community leader or family member who has this customary role – within or outside the EU; the data producer of a long-term cohort study)		**X**				
13. Will you be recruiting your participants through a crowd-sourcing service and/or involve a third party data-gathering service, such as a survey platform?		**X**				
14. Will you be offering any financial, or other, remuneration to participants, and might this induce or bias participation?		**X**				
E: Subject Matter *Research related to medical questions/health may require special attention. See also the website of the CCMO before contacting the HREC.*						
15. Will your research involve any of the following: • Medical research and/or clinical trials • Invasive sampling and/or medical imaging • Medical and *In Vitro Diagnostic Medical* Devices Research		**X**				
16. Will drugs, placebos, or other substances (e.g., drinks, foods, food or drink constituents, dietary supplements) be administered to the study participants? *If yes see here to determine whether medical ethical approval is required*		**X**				
17. Will blood or tissue samples be obtained from participants?		**X**				

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT – what risks could arise?** *Please ensure that you list ALL of the actual risks that could potentially arise – do not simply state whether you consider any such risks are important!*	**MITIGATION PLAN – what mitigating steps will you take?** *Please ensure that you summarise what actual mitigation measures you will take for each potential risk identified – do not simply state that you will e.g. comply with regulations.*	**DMP**	**ICF**
If yes see here to determine whether medical ethical approval is required						
18. Does the study risk causing psychological stress or anxiety beyond that normally encountered by the participants in their life outside research?		**X**				
19. Will the study involve discussion of personal sensitive data which could put participants at increased legal, financial, reputational, security or other risk? (e.g., financial data, location data, data relating to children or other vulnerable groups) *Definitions of sensitive personal data, and special cases are provided on the TUD Privacy Team website.*		**X**				
20. Will the study involve disclosing commercially or professionally sensitive, or confidential information? (e.g., relating to decision-making processes or business strategies which might, for example, be of interest to competitors)		**X**				
21. Has your study been identified by the TU Delft Privacy Team as requiring a Data Processing Impact Assessment (DPIA)? *If yes please attach the advice/ approval from the Privacy Team to this application*		**X**				
22. Does your research investigate causes or areas of conflict? *If yes please confirm that your fieldwork has been discussed with the appropriate safety/security advisors and approved by your Department/Faculty.*		**X**				
23. Does your research involve observing illegal activities or data processed or provided by authorities responsible for preventing, investigating, detecting or prosecuting criminal offences *If so please confirm that your work has been discussed with the appropriate legal advisors and approved by your Department/Faculty.*		**X**				
F: Research Methods						
24. Will it be necessary for participants to take part in the study without their knowledge and consent at the time? (e.g., covert observation of people in non-public places).		**X**				
25. Will the study involve actively deceiving the participants? (For example, will participants be deliberately falsely informed, will information be withheld from them or will they be misled in such a way that they are likely to object or show unease when debriefed about the study).		**X**				
26. Is pain or more than mild discomfort likely to result from the study? And/or could your research activity cause an accident involving (non-) participants?		**X**				
27. Will the experiment involve the use of devices that are not 'CE' certified? *Only, if 'yes': continue with the following questions:*	**X**		**The used robot manipulator is not CE certified and built by the company itself.** **It could result in risks of unexpected physical discomfort or injury to the participants.**	The used device for this experiment is a highly regulated robot by the company that has gone through extensive quality control checks and reviewing pipelines. Multiple safety measures such as automatic brakes and emergency stop buttons and extensive protection and safety measures in the robot control software are provided inside the robot as well as outside the robot. The robot is using electrical wiring for medical equipment, which has standard safety build-ins. Note: the robot will not be used in any medical or clinical setting during this experiment.		

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT – what risks could arise?** *Please ensure that you list ALL of the actual risks that could potentially arise – do not simply state whether you consider any such risks are important!*	**MITIGATION PLAN – what mitigating steps will you take?** *Please ensure that you summarise what actual mitigation measures you will take for each potential risk identified – do not simply state that you will e.g. comply with regulations.*	**DMP**	**ICF**
				In addition, extensive training is part of the experiment to prevent unexpected behavior to occur. At all times a project involved person will be attending the experiment. Added footpedal is a standard hardware equipment. Additionally, in the explicit consent it is ensured that the participants are provided with sufficient information to make an informed decision on whether or not to participate.		
• Was the device built in-house?	**X**					
• Was it inspected by a safety expert at TU Delft? *If yes, please provide a signed device report*		**X**				
• If it was not built in-house and not CE-certified, was it inspected by some other, qualified authority in safety and approved? *If yes, please provide records of the inspection*		**X**				
28. Will your research involve face-to-face encounters with your participants and if so how will you assess and address Covid considerations?	**X**		**Covid-19 spreads are a risk to the patients performing experiments.**	In the explicit consent it is ensured that the participants are provided with the obligation to report being sick and reschedule the experiment. Moreover, the researcher is also obligated to report and reschedule when feeling sick. During the experiment, the doors will be kept open and a ventilator will be used to ensure good ventilation of the room. Moreover, the distances between the two participants will be kept at 1,5 m unless they work together in close proximity on a daily basis. Additionally, in the explicit consent it is ensured that the participants are provided with sufficient information to make an informed decision on whether or not to participate.		
29. Will your research involve **either:** a) "big data", combined datasets, new data-gathering or new data-merging techniques which might lead to re-identification of your participants **and/or** b) artificial intelligence or algorithm training where, for example biased datasets could lead to biased outcomes?		**X**				
G: Data Processing and Privacy						
30. Will the research involve collecting, processing and/or storing any directly identifiable PII (Personally Identifiable Information) including name or email address that will be used for administrative purposes only? (eg: obtaining Informed Consent or disbursing remuneration)	**X**		**The data could be traced back to the individual person and their performance could be evaluated by other companies or individuals.**	Anonymisation and pseudonymisation of data takes place. The key where translation between anonymized data and personalized data can take place, will be safed in an encrypted file. Additionally, the personalized data is limited to name only. Moreover, the data will not be shared publicly after the project and therefore this minimizes the risk of data leaks or data being available for other companies or indivuals.		

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT – what risks could arise?** *Please ensure that you list ALL of the actual risks that could potentially arise – do not simply state whether you consider any such risks are important!*	**MITIGATION PLAN – what mitigating steps will you take?** *Please ensure that you summarise what actual mitigation measures you will take for each potential risk identified – do not simply state that you will e.g. comply with regulations.*	**DMP**	**ICF**
				Lastly, when a breach is noticed, the participants will be notified as soon as possible. Additionally, in the explicit consent it is ensured that the participants are provided with sufficient information to make an informed decision on whether or not to participate.		
31. Will the research involve collecting, processing and/or storing any directly or indirectly identifiable PIRD (Personally Identifiable Research Data) including videos, pictures, IP address, gender, age etc and **what other Personal Research Data** (including personal or professional views) will you be collecting?	**X**		**The indirect personal data that will be gathered is age group, previous experience with robots and gender.** **The data could indirectly be traced back to the individual person and their performance could be evaluated by other companies or individuals.**	Anonymisation and pseudonymisation of data takes place. The key where translation between anonymized data and personalized data can take place, will be safed in an encrypted file. Additionally, the personalized data is limited to gender, age group and experience. Age group makes it more difficult to trace back the person compared to age. Moreover, the data will not be shared publicly after the project and therefore this minimizes the risk of data leaks or data being available for other companies or indivuals. Lastly, when a breach is noticed, the participants will be notified as soon as possible. Additionally, in the explicit consent it is ensured that the participants are provided with sufficient information to make an informed decision on whether or not to participate.		
32. Will this research involve collecting data from the internet, social media and/or publicly available datasets which have been originally contributed by human participants		**X**				
33. Will your research findings be published in one or more forms in the public domain, as e.g., Masters thesis, journal publication, conference presentation or wider public dissemination?	**X**		After the thesis is finalized, the data and thesis will not be published publicly. Only after the embargo is exceeded the data will become available publicly. Publicly available data can lead to higher risks in evaluation of a person.	Anonymisation and pseudonymisation of data takes place. The key where translation between anonymized data and personalized data can take place, will be safed in an encrypted file. Moreover, all the personal data is limited to a minimum. Additionally, in the explicit consent it is ensured that the participants are provided with sufficient information to make an informed decision on whether or not to participate.		
34. Will your research data be archived for re-use and/or teaching in an open, private or semi-open archive?	**X**		The data will be shared with Eindhoven Medical Robotics where it will be used for re-use, research and development.	The data will be saved on an encrypted disk only accessible to Eindhoven Medical Robotics employees. Only participants involved in the project through TU Delft or Eindhoven Medical Robotics are allowed to participate in the experiment. Additionally, in the explicit consent it is ensured that the participants are provided with sufficient information to make an informed decision on whether or not to participate.		

H: More on Informed Consent and Data Management

NOTE: You can find guidance and templates for preparing your Informed Consent materials) here

Your research involves human participants as Research Subjects if you are recruiting them or actively involving or influencing, manipulating or directing them in any way in your research activities. This means you must seek informed consent and agree/ implement appropriate safeguards regardless of whether you are collecting any PIRD.

Where you are also collecting PIRD, and using Informed Consent as the legal basis for your research, you need to also make sure that your IC materials are clear on any related risks and the mitigating measures you will take – including through responsible data management.

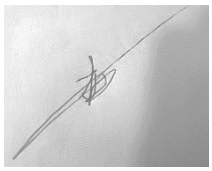
Got a comment on this checklist or the HREC process? You can leave your comments here

IV. Signature/s

Please note that by signing this checklist list as the sole, or Responsible, researcher you are providing approval of the completeness and quality of the submission, as well as confirming alignment between GDPR, Data Management and Informed Consent requirements.

Name of Corresponding Researcher (if different from the Responsible Researcher) (print)
Susanna Halman

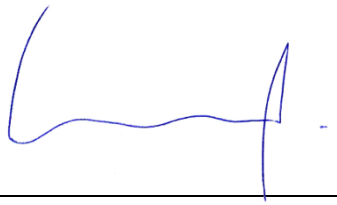
Signature of Corresponding Researcher:



Date:
05-08-2022

Name of Responsible Researcher (print)
Luka Peternel

Signature (or upload consent by mail) Responsible Researcher:



Date: 12-08-2022

V. Completing your HREC application

Please use the following list to check that you have provided all relevant documentation

Required:

- **Always:** This completed HREC checklist
- **Always:** A data management plan (reviewed, where necessary, by a data-steward)
- **Usually:** A complete Informed Consent form (including Participant Information) and/or Opening Statement (for online consent)

Please also attach any of the following, if relevant to your research:

Document or approval	Contact/s
Full Research Ethics Application	After the assessment of your initial application **HREC will let you know if and when you need to submit additional information**
Signed, valid Device Report	Your Faculty HSE advisor
Ethics approval from an external Medical Committee	TU Delft Policy Advisor, Medical (Devices) Research
Ethics approval from an external Research Ethics Committee	Please append, if possible, with your submission
Approved Data Transfer or Data Processing Agreement	Your Faculty Data Steward and/or TU Delft Privacy Team
Approved Graduation Agreement	Your Master's thesis supervisor
Data Processing Impact Assessment (DPIA)	TU Delft Privacy Team
Other specific requirement	Please reference/explain in your checklist and append with your submission

E.3. Data management plan

The data management plan allows correct saving and handling of personal data.

Plan Overview

A Data Management Plan created using DMPonline

Title: MSc Thesis: Performance of multi-modal multifeedback touched based pair point registration with force control

Creator: Susanna Halman

Affiliation: Delft University of Technology

Template: TU Delft Data Management Plan template (2021)

Project abstract:

The purpose of this research study is comparing accuracy, workload, usability and trust of participants with different systems in completing the task of touched based pair point localization with robot force control. Participants will be asked to perform fiducial localization in the robot space by positioning the robot end-effector with force control. Thereafter, they will be asked to fill in three different surveys consisting of different types of questions on a scale. Before this, they will be given some training with the robot, task and the safety procedures so that they understand their role, task and know how to respond if something happens during this experiment.

ID: 104819

Start date: 01-03-2022

End date: 07-10-2022

Last modified: 18-08-2022

MSc Thesis: Performance of multi-modal multifeedback touched based pair point registration with force control

0. Administrative questions

1. Name of data management support staff consulted during the preparation of this plan.

My faculty data steward, [Bjorn, Surname of the data steward], has reviewed this DMP on [date].

2. Date of consultation with support staff.

2022-08-05

I. Data description and collection or re-use of existing data

3. Provide a general description of the type of data you will be working with, including any re-used data:

Type of data	File format(s)	How will data be collected (for re-used data: source and terms of use)?	Purpose of processing	Storage location	Who will have access to the data
Participant Demographics: Age-group, Gender and experience with robots	.csv	Short survey	To report the average age-group and gender of the involved participants and divide the groups as equally as possible with respect to gender, age-group and experience. The experience will furthermore be converted to a group to later find possible relations between results and experience. Age-groups will also be used to relate results to age-group.	Personal laptop and external hard drive after transforming experience to experience groups.	Susanna Halman and Eindhoven Medical Robotics
Name	Written consent form	Consent form	To get participant consent	Scanned to personal laptop with fingerprint-only access and external hard drive that will be locked away. Will be saved in an encrypted file. Hard copies will be removed. The consent forms will also be saved in an encrypted file only accesible to Eindhoven Medical Robotics.	Susanna Halman and Eindhoven Medical Robotics
Qualitative metrics for usability	.csv	Through the USE Questionnaire and the system Usability Scale (SUS). Data is anonymous, only linked to participant number which is not connected to name or ID of particpant.	To score the different designs with the average usability score of all participants	Personal laptop with fingerprint-only access and external hard drive that will be locked away. Will also be saved in an encrypted file only accesible to Eindhoven Medical Robotics.	Susanna Halman and Eindhoven Medical Robotics
Qualitative metrics for workload	.csv	Through the NASA TLX survey. Data is anonymous, only linked to participant number which is not connected to name or ID of particpant.	To score the different designs with the average workload score of all participants.	Personal laptop with fingerprint-only access and external hard drive that will be locked away. Will also be saved in an encrypted file only accesible to Eindhoven Medical Robotics.	Susanna Halman and Eindhoven Medical Robotics

Qualitative metrics for trust	.csv	Through the Trust Perception Scale-HRI. Data is anonymous, only linked to participant number which is not connected to name or ID of particpant.	To score the different designs with the average trust score of all participants	Personal laptop with fingerprint-only access and external hard drive that will be locked away. Will also be saved in an encrypted file only accesible to Eindhoven Medical Robotics.	Susanna Halman and Eindhoven Medical Robotics
Quantitative metrics of workload	.csv	Through the measurement of force over time. Data is anonymous.	To score the different designs with the average workload score of all participants	Personal laptop with fingerprint-only access and external hard drive that will be locked away. Will also be saved in an encrypted file only accesible to Eindhoven Medical Robotics.	Susanna Halman and Eindhoven Medical Robotics
Quantitative metrics of localization accuracy	.csv	The fiducial localization error by using the localization position. Data is anonymous.	To score the different designs with the root mean square fiducial localization error.	Personal laptop with fingerprint-only access and external hard drive that will be locked away. Will also be saved in an encrypted file only accesible to Eindhoven Medical Robotics.	Susanna Halman and Eindhoven Medical Robotics
Quantitative metrics of robot end-effector position, mode, time, force and torque	.csv	Robot data from the SLRT. Data is anonymous.	To gain insight in possible discussion points.	Personal laptop with fingerprint-only access and external hard drive that will be locked away. Will also be saved in an encrypted file only accesible to Eindhoven Medical Robotics.	Susanna Halman and Eindhoven Medical Robotics

4. How much data storage will you require during the project lifetime?

- < 250 GB

II. Documentation and data quality

5. What documentation will accompany data?

- Data will be deposited in a data repository at the end of the project (see section V) and data discoverability and re-usability will be ensured by adhering to the repository's metadata standards
- README file or other documentation explaining how data is organised
- Methodology of data collection

The data will be accompanied with a methodology of data collection, a README file on how the data is organised in files and how scripts are used.
Discovery metadata and DataCite Metadata Schema will be used to ensure discoverability and adhere to repository standards.
This documentation will also be placed under an embargo in accordance to graduation agreement appendix 2, article 4.

III. Storage and backup during research process

6. Where will the data (and code, if applicable) be stored and backed-up during the project lifetime?

- Another storage system - please explain below, including provided security measures

Personal laptop and external hard drive. Personal laptop is only accessible with fingerprint scanner. External hard-drive is provided as back-up and will be locked away physically and will only be accesible to the researcher. The anonymized data will also be saved in an encrypted file only accesible to Eindhoven Medical Robotics.
The key that allows linking the anonymised data to the personalized data will be kept in an encrypted file together with the consent forms at the fingerprint-only accesible laptop. These will also be saved in an encrypted file only accesible to Eindhoven Medical Robotics.

IV. Legal and ethical requirements, codes of conduct

7. Does your research involve human subjects or 3rd party datasets collected from human participants?

- Yes

8A. Will you work with personal data? (information about an identified or identifiable natural person)

If you are not sure which option to select, ask your Faculty Data Steward for advice. You can also check with the privacy website or contact the privacy team: privacy-tud@tudelft.nl

- Yes

Direct personal data: Name for consent form
Indirect personal data: Age group, gender, previous experience with robotic systems.

8B. Will you work with any other types of confidential or classified data or code as listed below? (tick all that apply)

If you are not sure which option to select, ask your Faculty Data Steward for advice.

- Yes, I work with other types of confidential or classified data (or code) - please explain below

The code will be confidential or partly confidential due to condifdential company information according to the graduation agreement appendix 2, articles 1,2 and 3.

9. How will ownership of the data and intellectual property rights to the data be managed?

For projects involving commercially-sensitive research or research involving third parties, seek advice of your Faculty Contract Manager when answering this question. If this is not the case, you can use the example below.

The datasets underlying the published master thesis will be uploaded to the TU Delft repository following the TU Delft Research Data Framework Policy. During the active phase of research, the project leader from TU Delft will oversee the access rights to data (and other outputs), as well as any requests for access from external parties.
Before publication, an embargo will be placed upon the data according to graduation agreement appendix 2, article 4.
IP rights will be managed according to the graduation agreement appendix 2, article 5.

10. Which personal data will you process? Tick all that apply

- Other types of personal data - please explain below
- Signed consent forms
- Gender, date of birth and/or age

Previous experience with robotic systems

11. Please list the categories of data subjects

Study participants will be including working force participants.
Due to condifdential company information, only Company employees or TU Delft employees are allowed to participate in the experiment.

12. Will you be sharing personal data with individuals/organisations outside of the EEA (European Economic Area)?

- No

15. What is the legal ground for personal data processing?

- Informed consent

16. Please describe the informed consent procedure you will follow:

All study participants will be asked for their written consent for taking part in the study and for data processing before the start of the research. The participants will need to sign an explicit consent.

17. Where will you store the signed consent forms?

- Same storage solutions as explained in question 6

The consent forms will be located in an encrypted file.

18. Does the processing of the personal data result in a high risk to the data subjects?

If the processing of the personal data results in a high risk to the data subjects, it is required to perform a Data Protection Impact Assessment (DPIA). In order to determine if there is a high risk for the data subjects, please check if any of the options below that are applicable to the processing of the personal data during your research (check all that apply).
If two or more of the options listed below apply, you will have to complete the DPIA. Please get in touch with the privacy team: privacy-tud@tudelft.nl to receive support with DPIA.
If only one of the options listed below applies, your project might need a DPIA. Please get in touch with the privacy team: privacy-tud@tudelft.nl to get advice as to whether DPIA is necessary.
If you have any additional comments, please add them in the box below.

- None of the above applies

22. What will happen with personal research data after the end of the research project?

- Anonymised or aggregated data will be shared with others

The anonymised data will be shared with Eindhoven Medical Robotics after the experiment and there the data will be saved on an encrypted drive only accessible to employees.
The data will not be shared anywhere else and will be placed under an embargo according to graduation agreement appendix 2, article 4. No data will be available publicly.

25. Will your study participants be asked for their consent for data sharing?

- Yes, in consent form - please explain below what you will do with data from participants who did not consent to data sharing

Participants that do not conform data sharing, will be excluded from the experiment.

V. Data sharing and long-term preservation

27. Apart from personal data mentioned in question 22, will any other data be publicly shared?

- No other data can be publicly shared - please explain below why data cannot be publicly shared
- All other non-personal data (and code) underlying published articles / reports / theses
- All other non-personal data (and code) produced in the project

The data and code will not be publicly shared. They shall be shared with an embargo according to the graduation agreement appendix 2, article 4 where part of the code will completly be left out completely due to condifdential company information according to the graduation agreement appendix 2, articles 1,2 and 3.

29. How will you share research data (and code), including the one mentioned in question 22?

- No data can be publicly shared - please explain below

The data and code will not be publicly shared. They shall be shared with an embargo according to the graduation agreement appendix 2, article 4 where part of the code will be left out completely due to condifdential company information according to the graduation agreement appendix 2, articles 1,2 and 3.

VI. Data management responsibilities and resources

33. Is TU Delft the lead institution for this project?

- Yes, leading the collaboration - please provide details of the type of collaboration and the involved parties below

Collaboration with Eindhoven Medical Robotics according to the graduation agreement.

34. If you leave TU Delft (or are unavailable), who is going to be responsible for the data resulting from this project?

Luka Peternel (L.Peternel@tudelft.nl) will be responsible for the data resulting from this project after leaving the TU Delft.
The data shared with Eindhoven Medical Robotics will be falling responsible to Arash Arjmandi. The anonymised data will be shared with Eindhoven Medical Robotics after the experiment and there the data will be saved on an encrypted drive only accessible to employees.

35. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

The data will not be shared publicly in accordance to the graduation agreement appendix 2, article 4.

.

F. STOTA registration methods

This Appendix provides detailed knowledge of the found STOTA registration methods.

Figure 3.21 provides a summary of the found STOTA registration methods. The STOTA literature is analysed per registration method. This thesis focuses solely on the real-world coordinate frame, disregarding the image-based localisation in the STOTA papers. No methodological literature search was performed. Thus this information might lack some STOTA papers. Appendix A describes the steps for an elaborated literature search.

F.1. Invasive extrinsic registration methods

The paper from Gerber et al. (2013) achieves the lowest TRE of 0.101 ± 0.040mm, utilising a semi-automatic fiducial detection technique for pair-point-based bone-anchored fiducial registration. The paper employs a ball-in-cone localisation method requiring the human operator to localise each fiducial screw coarsely, illustrated in Figure F.1. Then automatically positioning the end-effector into the fiducial using a force sensor measuring the contact force along the x-, y- and z-axis. It minimises the contact forces in the x- and y-axis while achieving a specific constant contact force in the z-axis with an additional bias force of 1N to prevent the end-effector from settling for local minima (2013). An optical tracking camera and the robot position encoders track the robot end-effector to increase the positioning accuracy.

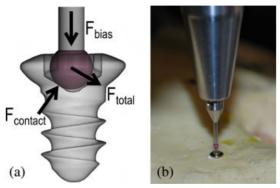


Figure F.1: The semiautomatic BFR detection technique described by Gerber et al. (2013). The paper employs a ball-in-cone localisation method requiring the human operator to localise each fiducial screw coarsely, seen in (b). Then automatically positioning the end-effector into the fiducial using a force sensor measuring the contact force along the x-, y- and z-axis, seen in (a). It minimises the contact forces in the x- and y-axis while achieving a specific constant contact force in the z-axis with an additional bias force of 1N to prevent the end-effector from settling for local minima

Gerber et al. (2013) tested the technique with ten fiducials on a phantom, resulting in an FLE of 0.005 ± 0.004mm exclusive tracking error (N = 45) and 0.046 ± 0.029mm inclusive tracking error (N = 42), as depicted in Figure F.2. Thirty different initial positions and orientations did not result in a significant difference in TRE, yet influenced the positioning time.

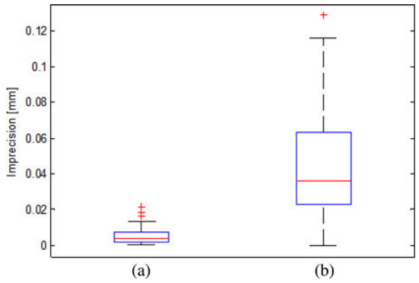


Figure F.2: Physical fiducial localisation error for BFR for ten fiducial screws gave a result of (a) 0.005 ± 0.004mm exclusive tracking error (N = 45) and (b) 0.046 ± 0.029mm inclusive tracking error (N = 42) (2013).

Machetanz, Grimm, Wang, et al. (2021) also uses BFR. This paper performs either BFR or LSR registration for 57 robot-assisted surgery patients. For BFR, A surgeon positioned a pointer attached to the robot end-effector to localise five fiducial locations. The fiducial and pointer both contain an internal diameter of 2.0mm, resulting in an average TRE of 0.7 $\pm$ 0.5mm for an entry point PT (2021). This result depends on the PT position, where more frontal positioning achieves higher accuracy than temporal or parietal positions and PT distance to fiducial centroids summarised in Figure F.3. Machetanz, Grimm, Wang, et al. (2021) concludes these BFR results are significantly lower than LSR on the same patients, thus increasing the registration accuracy. Machetanz, Grimm, Wang, et al. (2021) states that BFR complicates the surgical workflow.

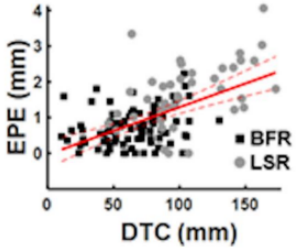


Figure F.3: The relation between the distance-to-centroid (DTC) on the x-axis with the Entry Point Error (EPE) on the y-axis for both BFR and LSR (2021). For both methods, the EPE increases as the DTC increases. Hence, the further away the surface points or bone fiducials centroids are from their PT, the larger the expected TRE.

Soteriou et al. (2016) compares BFR to automatic registration, LSR and anatomic landmark registration. The automatic registration employs the electromagnetic navigation system InstaTrak 3500 Plus (2016). Four skull models containing 26 screws are used for registration and TRE measurements. BFR uses four of those 26 screws, yielding a TRE of 0.94 $\pm$ 0.06mm. Automatic registration provided a TRE of 1.41 $\pm$ 0.04mm, LSR a TRE of 1.59 $\pm$ 0.14mm and 5.15 $\pm$ 0.66mm and 4.37 $\pm$ 0.73mm for the anatomic landmark registration with four or five landmarks respectively. Soteriou et al. (2016) also tested an oral split and concluded it inapplicable.

Luebbers et al. (2008) also compared multiple methods, among which the BFR. The other methods used for comparison are a dental-mounted splint, an LSR method called Z-touch, and a combination of the dental split with two fiducial screws. Z-touch is a class-I device which is moved free by the surgeon. It shows a visible red dot on the patient's skin and uses an infrared camera to provide live tracking of objects (Eggers et al., 2006). One hundred seventy landmarks on a skull were used to calculate the TRE values. These TRE values were separated based on the head area delivering the results in Figure F.4. The performances differ per area, but generally, the split performs the worst, and the screws and combination of a dental split with two fiducials score best. Luebbers et al. (2008) concludes that each setting requires different methods.

Most extrinsic registration methods rely on BFR, as it is the most accurate one (Su et al., 2022). However, J. Wang et al. (2020) mentions that BFR increases the patient's surgical trauma and the likelihood of infection, which is not beneficial to postoperative rehabilitation.

	Splint	Z-touch®	Screws	Splint & lat. orb.
Periorbital				
Min	0.4	0.2	0.1	0.2
Max	1.6	1.6	1.3	0.9
Avg	**1.1**	**1.0**	**0.7**	**0.6**
Stdev	0.3	0.4	0.3	0.2
n	46	46	46	46
Viscerocranium				
Min	0.5	0.3	0.1	0.2
Max	2.3	2.1	1.6	1.8
Avg	**1.3**	**1.2**	**0.8**	**0.8**
Stdev	0.4	0.4	0.3	0.3
n	40	40	40	40
Neurocranium				
Min	1.1	0.1	0.3	0.4
Max	3.4	2.0	2.1	2.5
Avg	**2.3**	**1.1**	**1.1**	**1.2**
Stdev	0.5	0.4	0.4	0.5
n	87	87	87	87

Figure F.4: The TRE in mm for different registration techniques: Split being a dental splint, Z-touch an LSR technique, BFR consisting of five fiducial screws, and Splint & lat.orb being the combination of two BFR with the dental split (2008). The TRE values are calculated based on 170 landmarks divided over a skull and segmented based on their areas on the head. The best performing method based on average TRE for the periorbital area is the Splint & lat. orb. The best-performing methods based on average TRE for the viscerocranium area are the Splint & lat. orb and the BFR For the neurocranium area, the best performing methods based on average TRE are the BFR, and the Z-touch method (2008). However, looking at the maximum TRE value, the best performing method is splint & lat. orb for the periorbital area, BFR for the viscerocranium, and Z-touch for the neurocranium area (2008)

The other invasive extrinsic registration method is frame-based, as shown in Figure 3.2. The paper of Machetanz, Grimm, Wuttke, et al. (2021) compares a frame-based surgery with a BFR surgery. It states a significant difference between the target entry point error of the BFR and frame-based surgery, 0.7 ± 0.5mm and 1.5 ± 0.6mm, respectively. However, no significant difference was found for another target point inside the skull, with TRE being 1.5 ± 0.8mm and 1.6 ± 0.8mm, respectively.

F.2. Non-invasive extrinsic registration methods

J. Wang et al. (2020) summarises non-inasive methods and their accuracies in Figure F.5.

Registration method	Invasive or not	Target registration error/target point error	Application
Bone-bed and the short process of malleus[23]	Minimal invasive	0.30–1.29 mm	Cochlear Implantation
Laser Surface Scanning[32]	Non-invasive	1.25 ± 0.64 mm	Neurosurgery
Headset[33]	Non-invasive	1.44 ± 0.24 mm	Lateral skull base surgery
Headband[34]	Non-invasive	1.46 ± 0.15 mm	Lateral skull base surgery
Granular jamming cap[34]	Non-invasive	0.56–1.40 mm	Skull base surgery
LED mask[35]	Non-invasive	0.92 ± 0.13 mm	Sinus and skull base surgery
Skin adhesive markers[36]	Non-invasive	2.49 ± 1.07 mm	Neurosurgery
Anatomical landmarks[37]	Non-invasive	0.93 ± 0.31 mm	Orthognathic surgery
Dental splint[38]	Non-invasive	0.55 ± 0.28 mm	Frontolateral skull base surgery
Template-assisted marker positioning[39]	Non-invasive	1.2 ± 0.12 mm	Temporal bone surgery
Our study	Non-invasive	0.671 ± 0.268 mm	Cochlear Implantation

Figure F.5: The TRE of different non-invasive registration methods and their applications (J. Wang et al., 2020). With **23** (Ke et al., 2016), **32** (Ballesteros-Zebadúa et al., 2016), **33** (Marmulla et al., 2005), **34** (Wellborn et al., 2017), **35** (Grauvogel et al., 2017), **36** (Woerdeman et al., 2007), **37** (Sun et al., 2013), **38** (Ledderose et al., 2012), **39** (Oka et al., 2014).

J. Wang et al. (2020) considers these accuracies insufficient for CI surgery and proposes another method with a visual-positioning frame connected to a mouthguard as shown in Figure F.6. The four marks attached to this mouthguard are detected with a Polaris stereo camera. This method acquires an EPE of 0.471 ± 0.276mm and an average cochlear PT of 0.671 ± 0.268mm with a maximum of

1.177mm and a minimum of 0.318mm. Executing eight surgical trajectories based on the registration damaged the chorda tympani in one case, whereas it had to be sacrificed in three cases. In two cases, the external auditory canal was damaged. In all cases, the facial nerves were undamaged (2020). The markers deviate with an average displacement of 0.03mm and 0.07 degrees.

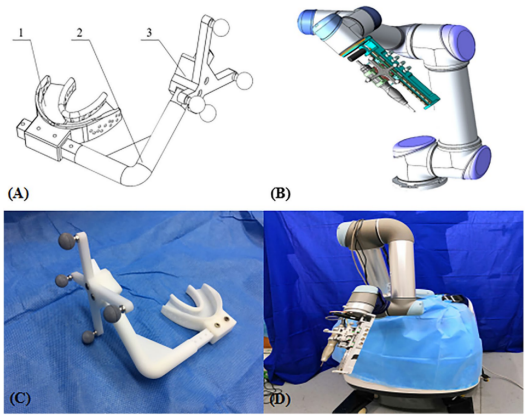


Figure F.6: A non-invasive registration method that uses a dental split attached to a marker frame. (A) Shows the device that consists of a dental support holder (1), a connecting tube (2), and a marker navigation frame (3) with four marker balls. These four marks are detected with a Polaris Stereo camera and calculate the transformation matrix to determine the position of the drill with respect to the patient. (B) Shows the drill attached to the end-effector of a UR5, 6DOF, robotic arm. (C) Shows the dental split device in the OR. (D) Shows the robotic system that is installed on an operating platform and that can be adjusted in its direction (2020).

One option similar to J. Wang et al. (2020) would be to optically track the patient with a visual frame such as stated or by electromagnetic tracking of multiple positions that can be used as a coordinate frame (Teatini et al., 2018). Nevertheless, it was concluded that electromagnetic tracking with electromagnetic sensors is less accurate than optical tracking by the Polaris camera, as mentioned by J. Wang et al. (2020).

This research found other non-invasive methods that provide similar results and methods as stated in Figure F.6, such as the paper by Labadie et al. (2004) that used a dental split and achieved a mean TRE of 0.73 $\pm$ 0.25mm in the application of CI surgery. Alternatively, the paper by Varma and Eldridge (2006) reports an ultrasound registration system with a TRE of 0.3 to 5.1mm. The paper by Opdenakker et al. (2017) uses a unique dental split and headband for paediatric application but lacks accurate results.

One unique non-invasive registration method that differs from the already found results is the Neuroclate device described by Cardinale et al. (2017) and shown in Figure F.7. The Neuroclate consists of five ruby balls (red arrows), the Neuroclate FMs, attached to the FM frame (black arrow) with carbon-fibre rods (green arrow). It attaches a frame to a laser tool holder (white arrow) mounted to the arm of a 5DOF passive robot. The Neuroclate moves close to the patient, and intraoperative imaging provides the 3D data of the balls. Cardinale et al. (2017) tested the Neuroclate in a clinical study with eight patients, calculating the TRE based on 20 surgical trajectories resulting in a mean TRE of 0.67 $\pm$ 0.29mm for Neurolocate-based trajectories compared to 0.76 $\pm$ 0.34mm for frame-based registration.

Adhesive marker registration still needed to be considered. Mongen and Willems (2019) describes their accuracy with an average TRE of 2.49 $\pm$ 0.86mm for seven adhesive markers, measured from eleven patients. The inaccuracy comes from the marker's size and skin mobility (2019). Other papers mention kin markers falling off or slipping over time (Güler et al., 2013; Omara et al., 2014).

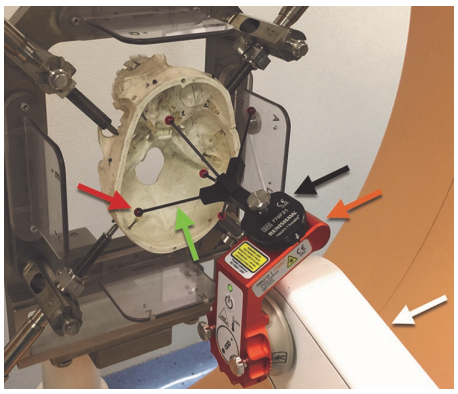


Figure F.7: An overview of the Neurolocate registration device as described by Cardinale et al. (2017). This device has five ruby balls (red arrows) which are referred to as the Neurolocate FMs, and are attached to an FM frame (black arrow) with carbon-fibre rods (green arrow). This frame is attached to a laser tool holder (white arrow), which is mounted to the arm of a 5DOF passive robot. Then the Neurolocate is moved very close to the patient's head; an intraoperative CT scan provides 3D data, and registration is performed with a specific software module that selects the centre of the FMs. The five positions are sent to the robot, and the transformation matrix is calculated (2017).

F.3. Intrinsic Point-based registration methods

Shamir et al. (2009) uses nine anatomical landmarks in addition to three skin-applied fiducials for registration that require locating by a neurosurgeon with a tracked pointer. This results in an average FLE of 1.6 - 3.0mm based on 45 repeated measurements on ten patient heads (2009). Shamir et al. (2009) compared the results of an expert neurosurgeon with a novice neurosurgeon, providing an average FLE of 1.6 $\pm$ 1.5mm and FLE of 2.2 $\pm$ 1.3mm, respectively. Based on TRE estimation calculations of Fitzpatrick and West (2001), the TRE was estimated at 4.1 $\pm$ 1.6mm.

Schneider et al. (2018) yields higher accuracies for anatomical landmark registration by selecting different combinations of landmarks, placing four on the mastoid cortex, eight on the external auditory canal and middle ear. Schneider et al. (2018) uses two human cadaverous temporal bones implanted with four bone-anchored fiducial screws, utilising the BFR method from Gerber et al. (2013), as the ground truth values to compare 14 different strategies, seven through anatomical landmarks and seven through LSR. Figure F.8 shows the seven anatomical landmark strategies.

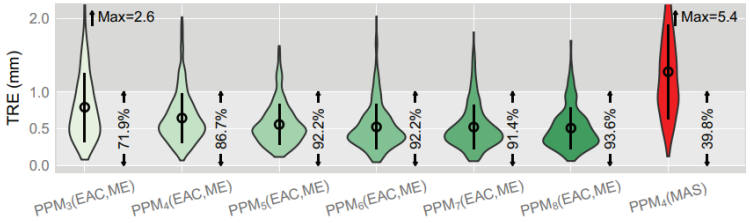


Figure F.8: TREs of a different combination of landmark areas and several landmarks as described by Schneider et al. (2018). The different possible areas are the mastoid cortex (MAS), the external auditory canal (EAC), and the Middle Ear (ME). The number of landmarks chosen ranges from three to eight. PPM is used for pair-point matching, thus referring to pair-point-based registration. For each combination, the TRE is shown in a violin plot, where the point range indicates the mean and the standard deviation. The numbers on the right of each violin plot indicate the percentage of TRE that were below 1mm. For the two highest TREs, the maximum TRE is also shown next to the violin plot. The name subscript for each strategy indicates the number of landmarks used. Each violin plot consists of 360 data points (2018).

(2018) acquires the best-performing landmark-based registration, with a mean of TRE of 0.51 $\pm$ 0.28mm, through eight landmarks located in the external auditory canal and middle ear (2018). The worst method, with a mean TRE of 5.4mm, used four fiducials in the mastoid bone. Further, Schneider et al. (2018) concludes that TRE decreased with an increase in the number of landmarks and increased

significantly with an increasing distance-to-centroid (DTC), as shown in Figure F.9.

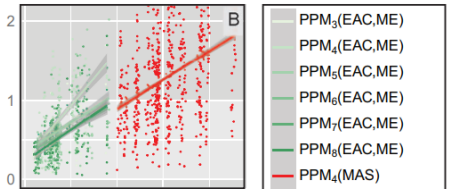


Figure F.9: Effect of the distance between the landmark and PT centre on the TRE. The TREs for seven different landmark combination methods on the y-axis is plotted against the distance between their centroid and the target position on the x-axis. Each dot corresponds to a TRE for a specific target and a combination. The line trends show an increase in TRE with an increase in distance for all registration strategies (2018).

Omara et al. (2014) follows a similar approach Schneider et al. (2018), defining three different configurations out of nine landmarks from different locations on the human head. The first configuration, the supine, uses eight landmarks from the front and side of the head. The second configuration, the lateral, uses six landmarks from one side, the front and back of the head. The last configuration, the prone, uses five landmarks from the back and sides of the head. Omara et al. (2014) tests these configurations in a lab and clinical environment. The lab environment tests use 20 case studies from two different hospitals, ten cases each, to simulate the configurations and calculate the TRE from 100 different PTs, as shown in Figure F.10. The clinical environment locates the landmarks three to four times to provide an FLE value ranging from 1.6 to 5.5mm and a TRE of 3.5 ± 0.17mm.

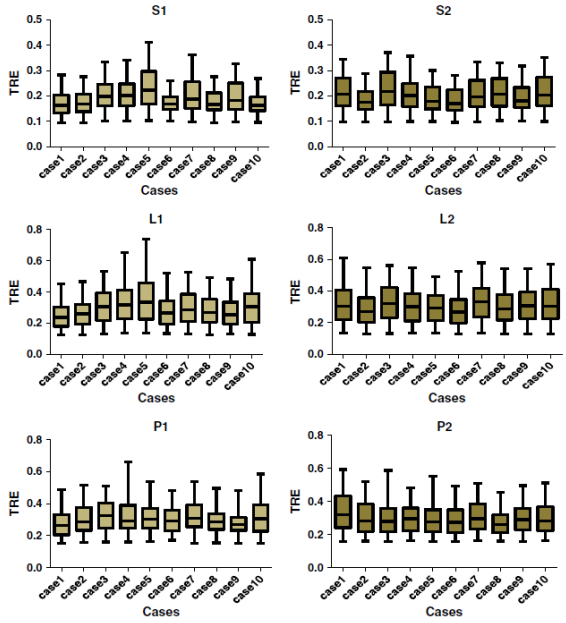


Figure F.10: The TRE values for landmark-based registration for two different hospitals, indicated by subscript 1 or 2, and three different configurations S, L and P, each presented with ten cases. The central mark indicates the median TRE for each box, and the bottom and top edges indicate the 25th and 75th percentiles of TRE, respectively. The whiskers extend to the most extreme data points not considered outliers (2014).

Another paper that uses anatomical landmarks and obtains higher accuracies than the previous is written by Güler et al. (2013), even though concluding insufficient for high-precision surgery. Güler et al. (2013) chooses different numbers of anatomical landmarks on a plastic skull and an anatomical specimen. User placement of a tracked probe provides landmark positions, yielding an FLE of 0.61mm for the plastic skull and 0.9mm for the anatomic specimen. This paper is the sole paper mentioning the human influence on the FLE. They state a higher human localisation error in anatomical markers

compared to fiducial screws (2013).

The paper of Sun et al. (2013) looks at the difference in TRE between two humans localising six anatomical landmarks on the front of a skull (2013). After training each human three times, calculating the TRE from 85 locations resulted in a mean TRE of 0.93 ± 0.31mm with an operator difference of 0.15mm.

This thesis did not find any papers on using geometrical landmarks for image-to-patient registration.

F.4. Intrinsic Surface-based registration methods

Four previously mentioned papers compared LSR with other registration methods. Machetanz, Grimm, Wang, et al. (2021) compares BFR to LSR for 57 robot-assisted surgery patients. For LSR resulting in an average TRE of 1.7 ± 0.9mm., depending on the DTC as shown in Figure F.3 and available imaging data.

Schneider et al. (2018) compared 14 different strategies, seven through anatomical landmarks and seven through LSR. Four fiducials provide coarse registration, and 50 surface points provide fine registration selecting them from different areas and providing seven distinct methods. In addition, different numbers of surface points from multiple areas were also compared. Figure F.11 demonstrates the results.

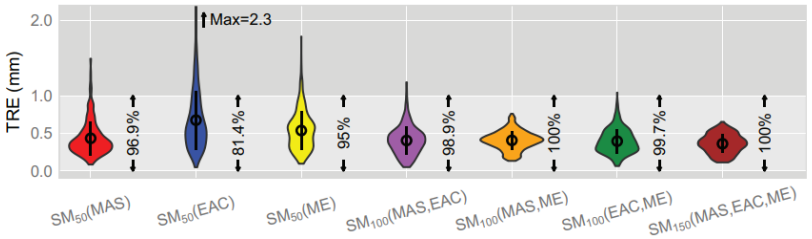


Figure F.11: TREs of a different combination of surface points from different areas and different numbers of surface points as described by Schneider et al. (2018). The different possible areas are the mastoid cortex (MAS), the external auditory canal (EAC), and the Middle Ear (ME). The number of surface points ranges from 50 to 150. SM is a term used for surface matching, thus referring to LSR. For each combination, the TRE is shown in a violin plot, where the point range indicates the mean and the standard deviation. The numbers on the right of each violin plot indicate the percentage of TRE that were below 1mm. The maximum TRE is also shown next to the violin plot for the highest TRE. The name subscript for each strategy indicates the number of surface points used. Each violin plot consists of 360 data points (2018).

(2018) investigates the effect of the number of scanned surface points on the registration accuracy by down-sampling the scanned surfaces to 5, 12, 22, 32, 42, and 50 surface points. Figure F.13 summarises these results (2018). Using all surfaces on all three parts provides the lowest TRE of 0.36 ± 0.13mm. The maximum TRE was stated as 2.3mm. Increasing the number of surface points decreases the TRE. More than 32 points do not significantly improve the TRE further (2018). Schneider et al. (2018) concludes a significant increase in TRE with increasing DTC, as shown in Figure F.12 (2018). Soteriou et al. (2016) and Luebbers et al. (2008) find mean TRE values for LSR of 1.59 ± 0.14mm and 1.1 ± 0.4mm, respectively.

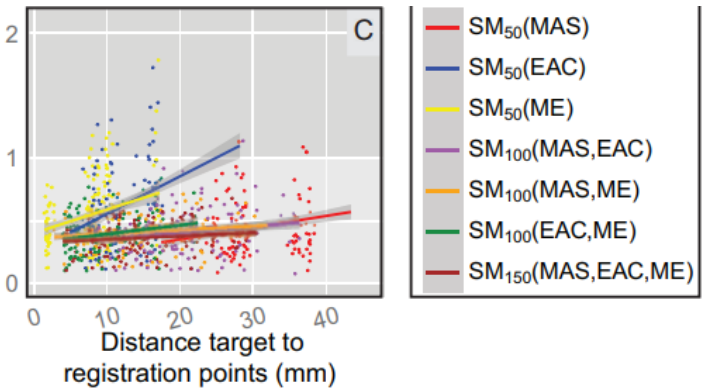


Figure F.12: Effect of the DTC for PCs and the PT on the TRE. The TREs for seven different PC combination methods on the y-axis is plotted against the distance between their centroid and the target position on the x-axis. Each dot corresponds to a TRE for a specific target and a combination. The line trends are depicted for each method (2018).

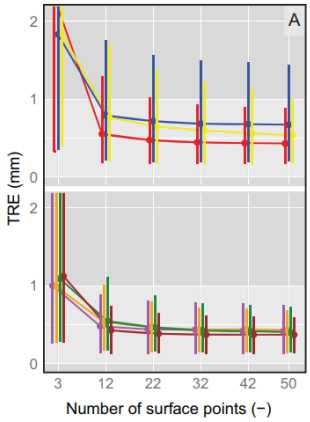


Figure F.13: TRE on the y-axis with respect to the number of surface points on the x-axis for seven different LSR methods. The colour scheme can be found in F.12. The upper Figure shows the results of LSR applied to only one anatomical region. The lower Figure shows the results of LSR methods that use multiple anatomical areas to make it more visible. The trendline indicates the mean TRE, whereas the range indicates the 5th and 95th percentile of TRE (2018).

Fan et al. (2020) also uses surface-based registration. This paper overcomes the problem or ICP results depending on the initial pose by matching two PCs with a robust, less accurate method before applying ICP. A common way is by additionally selecting three or four anatomical landmarks manually, which helps overcome the problem but also increases the registration time and limits possible automation. Therefore, Fan et al. (2020) proposes a method for automatic registration without the need for landmarks, based on coarse registration with a method called 4PCS, followed by fine registration with ICP. The 4PCS algorithm matches two PCs based on key markers obtained with a 3D Harris corner detector, obtaining an initial pose that can be used in fine ICP registration.

The paper compares their method with two other methods: 1) method only based on 4PCS and ICP algorithm, thus without the use of key points and 2) method based on manual coarse registration and automatic fine registration (2020). The proposed method does not outperform the other methods significantly with respect to the accuracy, as shown in Figure F.14. Fan et al. (2020) concludes that the proposed method does, however, have a significant speed advantage (2020).

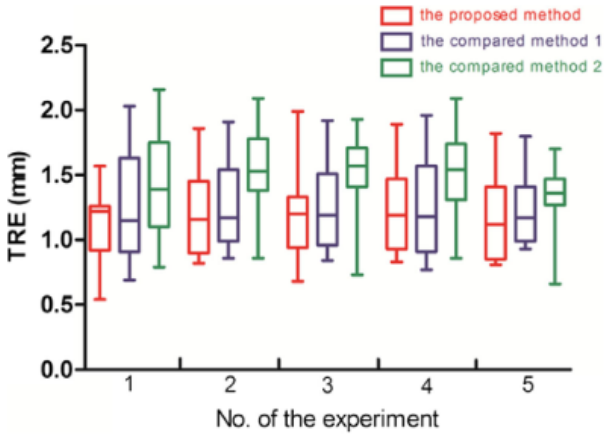


Figure F.14: TRE box plots providing the results of Fan et al. (2020), comparing their method with two other methods: 1) method only based on 4PCS and ICP algorithm and 2) method based on manual coarse registration and automatic fine registration (2020). The central mark indicates the median TRE for each box, and the bottom and top edges indicate the 25th and 75th percentiles of TRE, respectively. The whiskers extend to the most extreme data points not considered outliers.

Su et al. (2022) describes a similar method using an RGB-D camera to localise five facial landmarks with a deep learning tool called Adaboost (2022). These landmarks align the surfaces coarsely, whereafter in-depth matching takes place with ICP, obtaining an average TRE of 3.0mm for a camera to a phantom distance smaller than 1000mm.

The paper of Cao et al. (2016) uses a surface-based registration method without an imaging device to retrieve PCs. The surgeon moves a needle attached to the robot end-effector along a 3D reconstructed kneecap surface to record multiple surface points randomly, matching them with ICP. Cao et al. (2016) added five validation points fixed in location to measure the TRE values, providing a minimum TRE of 2.3mm for a PC of 40 points in the real world and a PC of 60 points in the 3D model. The chosen method induces incorrect positions due to gravity and installation errors (2016).

Lastly, Mongen and Willems (2019) performed surface matching based on the requirements of the navigation system at hand. Comparing three navigation systems resulted in an average TRE of 5.35 $\pm$ 1.64mm.

The LSR method accuracy highly depends on the available imaging data and can lead to an increase in intraoperative preparation time (Machetanz, Grimm, Wang, et al., 2021). Besides, clinical studies with LSR show additional problems related to the non-rigidity of surfaces, such as facial expressions (Omara et al., 2014). Lastly, the addition of LSR in addition to BFR does not improve TRE significantly (Perwög et al., 2018).

F.5. Other methods

Point-to-point ICP considers two points on different PCs and minimises their distance. However, point-to-plane ICP can also be used for registration, which includes information about the normal vector. The surface normals are computed from the target PC, and the error vector of two points close together onto the surface, as shown schematically in Figure F.15. Point-to-plane matching is more accurate than point-to-point matching and can improve accuracy with 0.05 to 0.2mm; no STOTA registration uses this method (Park and Subbarao, 2003).

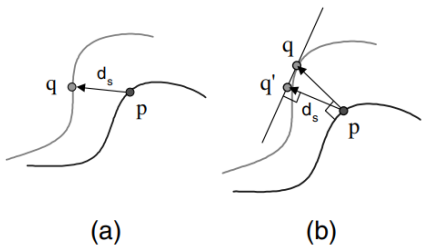


Figure F.15: The difference between (a) point-to-point registration and (b) point-to-plane registration. On one surface, the closest point q is searched for a control point p on another. This requires a minimal distance d_s between p and q for (a). For (b) the algorithm searches for a intersection $q^{'}$ from a normal vector of a control point p on another surface (2003).

Liao et al. (2013) describes other ICP methods, such as Projection-to-Volume Registration, Slice-to-Volume Registration, Video-to-Volume Registration and Volume-to-Volume Registration. The first three consider a 2D to 3D image registration. Due to deformities, these methods have high-accurate issues, and no one method seems superior compared to the others (2013).

One method unique from all others found is called a physics-based shape matching (PBSM) and is schematically shown in Figure F.16. PBSM focuses on mitigating problems related to soft-tissue registration using intraoperative imaging and preoperative images to match surface shapes. Suwelack et al. (2014) proposes to tackle this problem with an electrostatic–elastic model. The solution to this electrostatic–elastic problem is given with a finite element model to model the possibilities with the lowest energy cost. This registration method was tested on three different livers and yielded FLE values that represent the euclidean distance of the computed deformation at each point to the ground truth mesh (2014). These FLE values were 0.43, 0.32, and 0.46mm. It is concluded that this method outperforms other proposed surface registration methods for non-rigid surfaces.

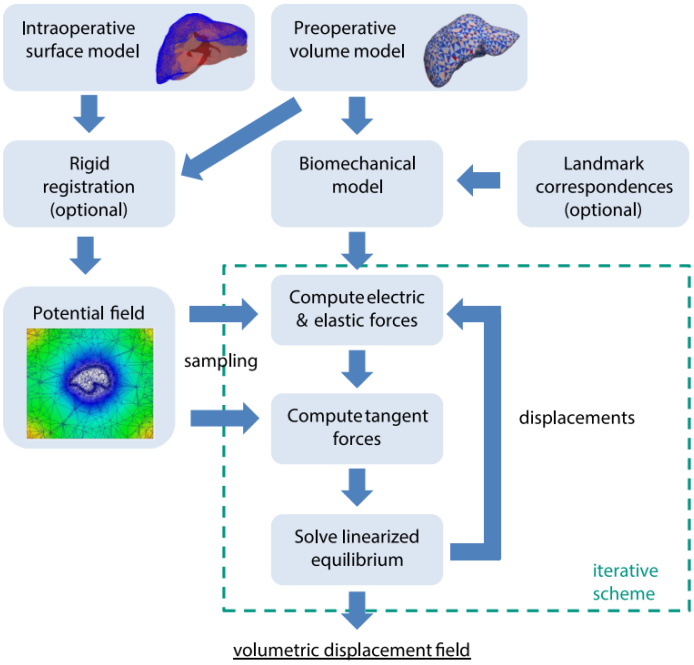


Figure F.16: The PBSM registration method for non-rigid surfaces as proposed by Suwelack et al. (2014). An intraoperative surface model is matched to a preoperative surface model. Rigid registration occurs between the preoperative and intraoperative surfaces, which is translated into a potential field. A biomechanical model is made from the preoperative model, using landmarks as reference points. This biomechanical model is then used to calculate elastic forces and tangent forces. This is matched with the potential field with the help of a finite element model to find an energy equilibrium. This last part is iterated as needed and results in a displacement field of the non-rigid surface (2014).

G. Additional knowledge

This Appendix provides additional background knowledge yielded during this project and used to create the scoring and exclusion criteria in Chapter 4.

This chapter starts with additional background knowledge related to HRI, followed by robot control related knowledge.

G.1. HRI related knowledge

HRI is already described in Section 3.4 but more knowledge was found that supported the scoring and exclusion criteria that is non-critical knowledge.

G.1.1. Levels of robot autonomy

Robots can perform tasks with diverse levels of robot autonomy (LORA). The level impacts how humans and robots interact (Beer et al., 2014). How HRI and LORA are related is widely discussed. HRI describes different types of human roles. These can be classified as human-in-the-loop (HITL) or human-out-the-loop (HOTL) (Nahavandi, 2017).

HITL denotes the robot needs the human to perform a certain task (2017). Perhaps the robot must stop and wait for new input, or the human acts as a supervisor. The latter is referred to as Human-on-the-loop (2017). One example of HITL is a sawing assignment where the robot supplies the force while the human is responsible for positioning the saw and initiating the movement. HOTL means the robot can act independently, without supervision, in performing its task. A robot vacuum is one example. The human role impacts both the LORA and HRI.

Concerning the LORA, a framework developed by Beer et al. (2014) describes a 10-point taxonomy of automation. Robot autonomy is defined in various ways by literature. This thesis employs the definition of Beer et al. (2014): *"The extent to which a robot can **sense** the environment, **plan** based on that environment, and **act** upon that environment, with the intent of reaching some goal (either given to or created by the robot) without external control."*.

Autonomy and the LORA are highly dependent on the task to be executed by the robot. A robot being autonomous in performing one specific task does not indicate it can perform other tasks autonomously (2014). For example, a vacuum cleaner robot might be fully autonomous in vacuum cleaning. Extending the task to the movement of misplaced objects, the robot is no longer considered autonomous.

The 10-point taxonomy of LORA divides tasks into three elements: sense, plan and act (2014; R. R. Murphy, 2019). Sense denotes that the robot is aware of its environment. Plan indicates the robot's behaviour adjustability based on its senses. Act represents performing a task based on the plan (Beer et al., 2014). In bone drilling, the robot perceives the bones needed to drill and senses obstacles that require avoidance (Sense); Next, the robot provides a path for drilling (Plan); Lastly, the drilling task is conducted (Act). These three components can be allocated to the human or the robot in different degrees. Differences in allocation provide distinct LORA, as shown in Figure G.1.

Autonomy is a continuum; a robot could be defined between two LORA. Besides, the robot's LORA can alter throughout specific interactions and is not considered static.

LORA	Sense	Plan	Act	Description
Manual	H	H	H	The human performs all aspects of the task including sensing the environment, generating plans/options/goals, and implementing processes.
Tele-operation	H/R	H	H/R	The robot assists the human with action implementation. However, sensing and planning is allocated to the human. For example, a human may teleoperate a robot, but the human may choose to prompt the robot to assist with some aspects of a task (e.g., gripping objects).
Assisted Tele-operation	H/R	H	H/R	The human assists with all aspects of the task. However, the robot senses the environment and chooses to intervene with task. For example, if the user navigates the robot too close to an obstacle, the robot will automatically steer to avoid collision.
Batch Processing	H/R	H	R	Both the human and robot monitor and sense the environment. The human, however, determines the goals and plans of the task. The robot then implements the task.
Decision Support	H/R	H/R	R	Both the human and robot sense the environment and generate a task plan. However, the human chooses the task plan and commands the robot to implement actions.
Shared Control With Human Initiative	H/R	H/R	R	The robot autonomously senses the environment, develops plans and goals, and implements actions. However, the human monitors the robot's progress and may intervene and influence the robot with new goals and plans if the robot is having difficulty.
Shared Control With Robot Initiative	H/R	H/R	R	The robot performs all aspects of the task (sense, plan, act). If the robot encounters difficulty, it can prompt the human for assistance in setting new goals and plans.
Executive Control	R	H/R	R	The human may give an abstract high-level goal (e.g., navigate in environment to a specified location). The robot autonomously senses environment, sets the plan, and implements action.
Supervisory Control	H/R	R	R	The robot performs all aspects of task, but the human continuously monitors the robot, environment, and task. The human has override capability and may set a new goal and plan. In this case, the autonomy would shift to executive control, shared control, or decision support.
Full Autonomy	R	R	R	The robot performs all aspects of a task autonomously without human intervention with sensing, planning, or implementing action.

Figure G.1: Proposed 10-point taxonomy of Levels of Robot Autonomy for HRI, dividing the task into three elements: Sense, Plan and Act. Sense denotes that the robot is aware of its environment. Plan indicates the robot's behaviour adjustability based on its senses. Act represents performing a task based on the plan (2014). These can be allocated to either the human (H) or the Robot (R) or shared between both (H/R). The LORA depends on this allocation and ranges from manual to full autonomy with a total of 10 levels (2014). Each level contains a description. Autonomy is a continuum, a robot could be defined between two LORA, and its LORA can change over time.

G.1.2. Human-robot Collaboration

Human-robot collaboration (HRC) focuses on finding the proper task division between humans and robots founded on their complementary strengths and limitations as described by Taylor (2006) and portrayed in Figure G.2. Utilising these skills to their best advantage can drive success and task performance in HRI (2006). HRC should be considered at all levels of the surgical team since it is shown that robot-assisted surgeries affect communication within a surgical team and requires distinct and unique training procedures (Abdelaal et al., 2019).

G.1.3. HRI for the medical field

Robots are utilised at an increasing rate in the medical field, where the interactions are exceptionally direct. The leading application is surgery (Okamura et al., 2010). The use of robots can drive the trend of minimally invasive surgery (MIS) (J. Liu et al., 2019). MIS is surgery performed with thin instruments and small incisions to create less trauma, reduce hospital stay and recovery time for the patient, and simplify clinical workflows (Machetanz, Grimm, Wang, et al., 2021; Westebring–van der Putten et al., 2008). This type of surgery comes with supplementary challenges and complications, which can be mitigated with the acquisition of computer-integrated surgery (CIS) or image-guided surgery (IGS) (Taylor, 2006).

	Strengths	Limitations
Humans	Excellent judgment Excellent hand-eye coordination Excellent dexterity (at natural "human" scale) Able to integrate and act on multiple information sources Easily trained Versatile and able to improvise	Prone to fatigue and inattention Tremor limits fine motion Limited manipulation ability and dexterity outside natural scale Cannot see through tissue Bulky end-effectors (hands) Limited geometric accuracy Hard to keep sterile Affected by radiation, infection
Robots	Excellent geometric accuracy Untiring and stable Immune to ionizing radiation Can be designed to operate at many different scales of motion and payload Able to integrate multiple sources of numerical & sensor data	Poor judgment Hard to adapt to new situations Limited dexterity Limited hand-eye coordination Limited haptic sensing (today) Limited ability to integrate and interpret complex information

Figure G.2: Complementary Strengths and Limitations of Robots and Humans (2006). The first row relates to human strengths and limitations in the medical field. The last row connects to robot strengths and limitations in the medical field. Human and robot strengths and limitations are largely complementary. For instance, a robot is tireless when a human is prone to fatigue and inattention. Integrating these strengths and limitations to their best capacities can lead to successful HRI tasks (2006).

CIS is the concept where computer technologies are employed for surgery in either planning, guiding or performing surgeries. Computers can enhance the surgeon's technical capabilities by making the procedure more accurate, safe or less invasive (2006). IGS is part of CIS and represents pre-operative or intraoperative imaging to oblige the surgeon in executing its task. Different imaging techniques such as computed tomography (CT) or magnetic resonance imaging (MRI) acquire these images (Azagury et al., 2015).

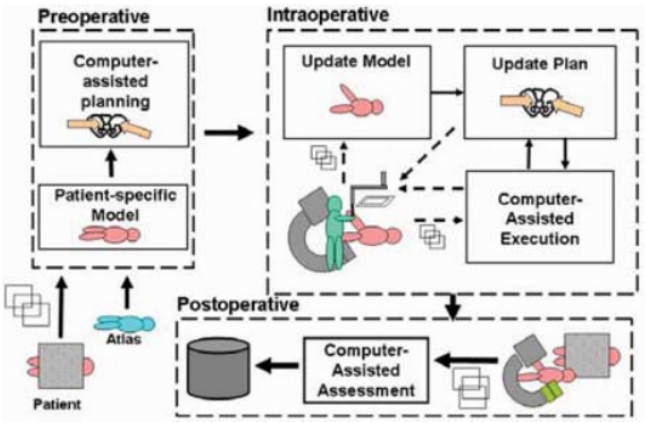


Figure G.3: Work and information flow of CIS. There are three distinct phases: pre-operative, intraoperative and post-operative. In the pre-operative phase, a patient-specific model is developed through imaging techniques such as CT or MRI. From this model, a surgical plan is conducted. During the intra-operative phase, the patient model is loaded and updated as required. Then the surgery can be executed with the aid of the surgical plan. In the post-operative phase, the surgery results can be compared to the surgical plan to define its success (2006).

Figure G.3 displays the work and information flow in CIS. The surgical workflow consists of a pre-operative, intraoperative and post-operative phase (2006). The pre-operative phase collects information about the patient, constructs a diagnosis and schedules surgery. In CIS, this encloses computer-assisted planning of the surgery (2006). This information is used in the operating room (OR) to guide the surgeon, primarily with 3D localisation of the surgery tools and specific landmarks. The plan can be revised if required where-after the surgery takes place. The definitive results can be compared to the initial plan to specify the success of the surgery.

LORA in the medical field

In the medical field, six LORA are distinguished, from 0 to 5, as portrayed in Figure G.4 (Yang et al., 2017). At the higher LORA, the robot is a medical device practising medicine.

Figure G.4: Various LORA used in the medical field. The LORA range from 0: No autonomy to 5: Full automation, presenting six levels. Level 0, the robot is perceived as a medical tool, whereas level 5 denotes a robot that can conduct surgery without HITL (2017).

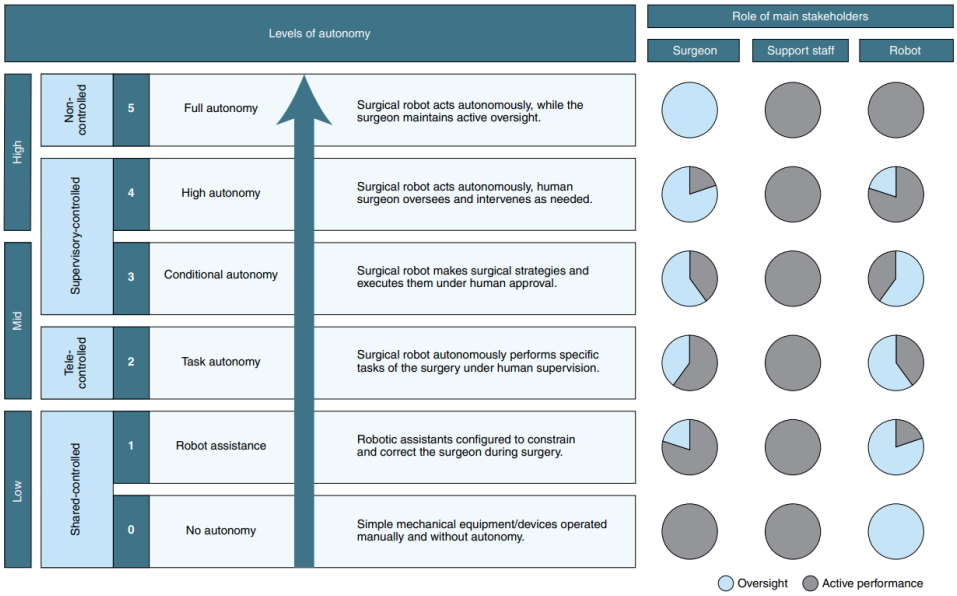


Figure G.5: Distinct LORA used in the medical field. The LORA range from no autonomy to full automation with six levels. Level 0, the robot is perceived as a medical tool, whereas level 5 denotes a robot that can conduct surgery without HITL (Yang et al., 2017). Each level maps the robot, surgeon and support staff roles. Each role is thought either an active participant, shown in grey, or an oversight participant, shown in blue. The participation of the surgeon shifts from active participation towards oversight with higher LORA. For the robot, the opposite is valid. The support staff has an active role at all levels (Fosch-Villaronga et al., 2021).

These LORA are furthermore expressed by Fosch-Villaronga et al. (2021), with the expansion of medical support staff. The support staff carries out tasks like patient positioning, anaesthetising, and performing control checks. The robot, surgeon and support staff roles are mapped over the distinct LORA as demonstrated in Figure G.5. Fosch-Villaronga et al. (2021) states that humans will not be eliminated with the increase of automation but will partake in an active role. Accordingly, HRI will permanently play an influential part and should be considered carefully.

Challenges in HRI for the medical field

HRI in the medical field offers additional challenges. The primary challenge is safety. To tackle these challenges, Wolf and Shoham (2009) notes design requirements for medical robots, illustrated in Figure G.6. These are extended with general human-robot design requirements, producing the list:

- Effective control
- Limited workspace
- Immunity against magnetic interference
- Fail-safe mode
- Safe behaviour near singularities
- Limit force and force feedback
- Full control option
- Footprint
- Human-robot Collaboration

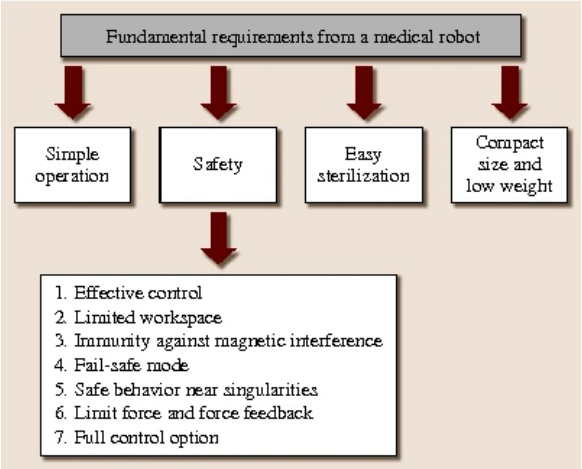


Figure G.6: Requirements for medical robots (2009). The most significant is safety, which directs to seven requirements: effective control, limited workspace, immunity against magnetic interference, fail-safe mode, safe behaviour near singularities, limited force and force feedback, and full control option. Besides, the robot should be designed as uncomplicated as possible and with easy sterilisation. Sterilisation indicates that the robot should be easily sterilised or covered with a drape. Last, the size and weight should be minimised (2009).

Effective control

There are two types of surgery robots with different HRI (Taylor, 2006). The first category is where the robot functions as an extension of the surgeon, the most common form being tele-operation (Abdelaal et al., 2019). These robots expand human capabilities, such as reducing tremors and improving accuracy by scaling down human movements (Taylor, 2006). The human and robot share their control, providing input into the system, called shared control.

In the second category, the robot and surgeons work together with the surgeon as a team. The robot performs individual tasks such as holding or injecting and retracting tools. These robots are called auxiliary surgical supports (2006). These robots predominantly have a control interface, though tasks are executed separately. This is called traded control, but if it takes place at a high level, it can be expressed as supervisory control (Abdelaal et al., 2019).

For these two types, force and speed should be restricted for the robot within safety thresholds to avoid collisions (Wolf and Shoham, 2009). A third category is an autonomous robot that performs assignments without human control or supervision. The proper control method improves performance, reliability and safety (De Santis et al., 2008).

Limited workspace
Limiting the workspace minimises collisions with humans and between robot links (Wolf and Shoham, 2009). Additionally, it limits singularities. Singularities are robot arm positions that different configurations can convey. Singularities can result in hazardous situations and should be evaded at all times.

Immunity against magnetic interference
In the OR, many tools are present and can have magnetic interference with the robotic system (2009). Consequently, the robot should have interference immunity.

Fail-safe mode
A fail-safe mode averts additional damage from occurring. For illustration, during a power failure, the robot should move to another position and remain in its position until the power is revived (2009).

Safe behaviour near singularities
Singularities should be avoided since they can yield harmful behaviour like quick oscillation. Therefore the path planning and workspace should be limited to non-singularity points at all times (2009).

Limit force and force feedback
The force the robot can apply during a task should be restricted (2009). In some procedures, like bone cutting, restricting can be complex due to demanding high force to execute the task.

Full control option
A full control option denotes that the surgeon should be authorised to take over full control at any stage within a detailed time limit (2009).

Footprint
The OR is full of equipment and tools and contains restricted space for new equipment. Hence a robot should have the least possible footprint and weight so it can be re-positioned easily (2009).

G.1.4. Human factors

The prominent role of humans in HRI demands analysis of introduced errors with the Human Factor Analysis and Classification System (HFACS) (Shappell and Wiegmann, 2000). This framework represents human errors on four levels of failure, which united lead to substantial errors. The levels are (1) Organisational influences, (2) Unsafe supervision, (3) Precondition for unsafe acts, and (4) Unsafe acts of operators. Since each level impacts the subsequent, this is called the Swiss cheese model as portrayed in Figure 3.24 (2000). This framework regards both active and latent failures. The latter can go overlooked for a prolonged time and were disregarded before the Swiss cheese model despite their substantial role in accidents. The level of unsafe acts of operators is already explained in Section 3.4.1, the others are explained here.

Organizational influences
At the highest level, items such as financial cuts for training or firm emphasis on efficiency rather than accuracy significantly influence the formation of accidents in the descending levels. There are three classes of organisational influences:

- Resource management
  This holds corporate-level decision-making concerning the allocation and maintenance of assets such as human, financial and equipment resources. This could lead to cuts in training and poorly maintained equipment (2000).
- Organisation climate
  Items like organisational structure, responsibility, communication channels, organisational

policies and culture, and accountability are indications of the organisational climate. A climate with limited communication channels, lack of responsibility divisions, or lack of focus on safety can lead to an absence of reporting and potential dangers (2000).

- Organisational processes
  These are organisational decisions and regulations that guide everyday activities, such as using proper methods for quality checks between workforce and management (2000).

Unsafe supervision

In organisations, supervisors are used to supervising, organising, guiding and planning operators and their organisational tasks. Unsafe supervision failures diverge into four categories:

- Inadequate supervision
  The role of a supervisor is to feed chances to succeed by providing guidance, training, motivation, leadership and a role model. If not provided, it could influence the occurrence of accidents (2000).
- Planned inappropriate operations
  These errors occur when tasks are scheduled at a rate that puts the operator at risk. For example, coupling two unskilled operators jointly (2000).
- Failed to correct the problem
  Particular deficiencies in training, equipment, or other corresponding safety errors are known to the supervisor, though they remain unsolved. For instance, failing to correct inappropriate behaviour or disregarding software updates (2000).
- Supervisory violations
  This emerges when existing rules and regulations are disregarded by the supervisor (2000).

Preconditions for unsafe acts

This level of failure concerns conditions encompassing the humans that have to carry out a task. These conditions include poor communication or coordination routines and mental fatigue. These preconditions for unsafe acts failures consist of three categories:

- Environmental factors
- Condition of operator
- Personnel factors

Environmental factors

Environments can influence the operator and therefore drive failures. These are divided into two sub-categories:

- Physical environment
  Includes both the operational setting, such as robot positioning and the ambient environment, such as lighting. (2000).
- Technological environment
  Factors that possess a variety of design choices (2000).

Condition of operator

The initial state or condition of the operator can prevent or increase failures. These are divided into three sub-categories:

- Adverse mental state
  Adverse mental states could be fatigue, distraction, task fixation, or other stressors.
  Furthermore, personal traits like overconfidence or lack of motivation can play a part (2000).
- Adverse physiological state
  States such as physical tiredness and medical abnormalities (2000).
- Physical/mental limitations

This involves the operator's individual physical or mental limitations. For instance, in the dark human sight is restricted. Moreover, it could be that an individual human operator is not fitted to execute the task; for the localisation task, someone with extreme tremors would be unsuited (2000).

Personnel factors

Personnel factors are related to operator planning and scheduling. The personnel factors are divided into two sub-categories:

- Crew resource management

  Adequate communication and coordination are essential. Coordination between operators is crucial; if two operators have to work together, their roles and responsibilities must be clear. For instance, briefing an operator before and after a task could limit accidents. (2000).
- Personnel readiness

  These errors occur when human operators fail to prepare physically and mentally for the task —for instance, not adhering to rest requirements, running a marathon or self-medication. These violations are not viewed as unsafe acts of operator violations since they do not occur inside the task environment but rather beforehand. Operators should have a sensible judgement on whether they can perform an assignment (2000).

G.2. Robot control

Control systems are used to control the movements and functions of robots. The control system controls the output with a simple block diagram as shown in Figure G.7. This control system is made up of a controller and a plant. In case of robot control, the controller is the motor that actuates the robot and the plant is the robot itself. This is summarised in Figure G.8. Control systems can be classified as either open or closed-loop systems. In robot control, closed-loop systems are the most used and only these will be discussed further.

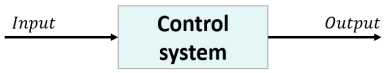


Figure G.7: A simple schematic block diagram of a control system. A control system controls the output based on certain inputs. An example of a control system could be cruise control, where the output, the speed, is controlled by the control system based on different inputs such as wind, changes in road types, changes in road grade.

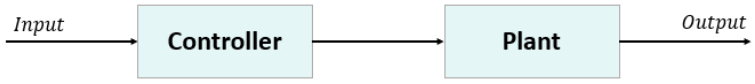


Figure G.8: A simple schematic block diagram of a control system where the control system is made up of a controller, being the motors that actuate the robot, and a plant, being the robot. In the example of cruise control the controller would also be the motor that drives the speed of the car and the plant would be the car itself. The signal going from the controller to the plant is also referred to as the actuating signal.

Closed-loop control systems use feedback from the output into the input to adjust the control system accordingly. This feedback signal is compared to the input signal to provide an error signal that is used to adjust the controller accordingly. This is shown in Figure G.9.

For robot control of robot arms, it is key to understand that a robot arm consists of a robot base, multiple links, multiple joints and an end-effector. This is displayed in Figure G.10. The end-effector has a certain position [x y z] in space and the joints, q, have a certain joint angle with respect to the links L. In controlling a robot arm, both the end-effector position as well as the joint angles are important.

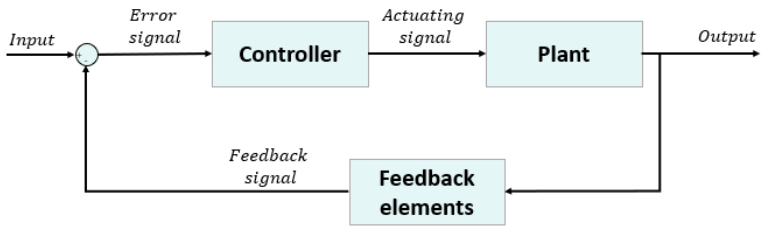


Figure G.9: A simple schematic block diagram of a closed-loop control system where the control system is made up of a controller, being the motors that actuate the robot, and a plant, being the robot. From the plant some feedback is used to get an error signal that is a measure of the difference between the desired and the actual position. This error signal is then used to adjust the controller accordingly. In the example of cruise control, the feedback that is used would be the actual speed of the car. This speed is then compared to the input speed and depending on the difference the controller is adjusted. For instance, if a car is driving down-hill the actual speed will increase with respect to the reference speed. Then the controller will adjust accordingly to decrease the speed of the car so that it matches the reference speed as good as possible.

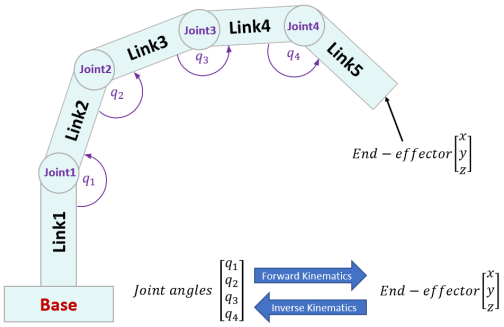


Figure G.10: A schematic robot arm consisting of a robot base, multiple links, multiple joints and an end-effector. The end-effector has a certain position [x y z] in space and the joints, q, have a certain joint angle with respect to the links L.

Robot can be mainly controlled in two different ways: position controlled or force controlled. In this thesis only position controlled robots will be considered. Position controlling the robot means a certain end-effector position is given as a reference, to which the robot should move (Owen-Hill, 2016; Siciliano and Villani, 1999). How this is done will be discussed in more detail.

Two other main definitions in robot control are kinematics and dynamics. Kinematics relates to motions of the robot arm without considering any forces. A kinematic model describes the end-effector position in space with relation to the robot base, with the assumption that no forces are applied. In kinematics two types of mathematical processes are distinguished: inverse kinematics and forward kinematics. Forward kinematics being the calculation of the end-effector position x based on joint angles q of the robot. Whereas inverse kinematics is the calculation of the robot joint angles q based on the end-effector position x.

In contrast, dynamics is the part where the motion and forces that cause this motion are considered. Once the kinematic model can provide us with a certain configuration of the robot at a certain end-effector position, forces should be applied to make the robot move in that position. Dynamics can also be divided into forward dynamics or inverse dynamics. Forward dynamics is the calculation of the joint accelerations or certain motions of the robot based on the external forces, joint states and joint torques. Inverse dynamics on the other hand is the calculation of the forces or torques that are required to produce a certain motion.

G.2.1. Position controlled robots

In position controlled robots, as mentioned, the reference position of the end-effector is controlled by the motors inside the robot arm. These motors send out an electrical current based on a reference position as measured by encoders (Keemink et al., 2018; Owen-Hill, 2016). The control-loop of a position controlled robot is shown in Figure G.11.

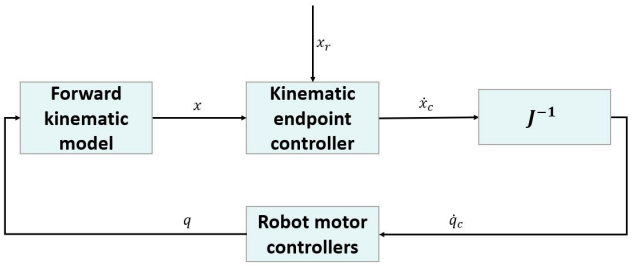


Figure G.11: Position-controlled control loop with x the actual end-effector position, $x_r ef$ the reference end-effector position, $\dot{x}_c$ the reference endpoint velocity, q the joint angles, $\dot{q}_c$ the reference joint velocities, and J the Jacobian matrix.

The input to the endpoint controller are the robot actual x and reference x_r end-effector position. The controller then provides the output of a calculated end-effector velocity to move from the current position to the reference position. This is done according to the formula $x = K_p(x_r - x)$, with K_p the proportional gain in 1/s and x the velocity needed to move from the current to the reference position.

This end-effector velocity is used to calculate the joint velocities using an inverse Jacobian with the formula $q = J^{-}1x$. Thus reflecting inverse kinematics. The Jacobian can be used to change velocities between joint and endpoint or change joint forces to endpoint forces. The joint velocities $\dot{q}_c$ are then send to the robot motor controllers to adjust their output accordingly and get the wanted joint angles. This reflects the inverse dynamics. Dynamic controls either through impedance or admittance and in this project the type of dynamics is impedance control. This means the robot takes a measured position or motion, in this case the joint velocities $\dot{q}_c$ and calculates a force that is used as output to move the robot to the correct position. These joint angles are thus used to move the robot to the joint angles q. From these joint angles, the end-position is calculated x using forward kinematics. This closes the control loop and is then repeated continuously.

Position controller robots have the advantage of being fairly cheap and easy to implement. However, they are difficult to use in a dynamic environment and have interactions with their environment. This can be counteracted by implementing a force sensor and using the force values acted upon by a human operator to calculate the reference position in this loop and create interactive behaviour (Keemink et al., 2018).

G.2.2. DH parameters

For forward kinematics, a kinematic model is needed to accurately described the end-effector position based on the joint angles of the robot. This is mostly done using transformation matrices from the robot base plate frame to the end-effector frame. One way to provide this is with DH parameters, also called Denavit–Hartenberg parameters. This limits the number of parameters the describe the transform between two adjacent coordinate frames to four being θ (deg), d (mm), α (deg), and a (mm), which represents the rotation around the z-axis, translation along the z-axis, rotation around the x-axis of the new frame, and translation along the x-axis of the new frame respectively. More information can be found in Siciliano et al. (2009).

H. EA and HFA

This Appendix provides an extensive error- and human-factor analysis for pair-point-based registration.

After choosing the concept from a broad selection, it should now be made concrete in a design. Since it should consider a high-accuracy design, it is important to know what errors play a role in registration and what factors contribute to those errors. The detailed error- and human-factor analysis will lead to design implementations.

Knowingly, the TRE is the most important measure for accuracy and system output for performing registration. The TRE is directly influences by the FLE, therefore the factors that contribute to the FLE are analysed. There are three main categories that contribute to the FLE, as summarised in Figure 5.4:

- *Registration Algorithm*
  The registration algorithm is the algorithm that calculates the final transformation matrix.
- *Image Localisation*
  Image localisation is the process of finding the position of the fiducials in the image coordinate system.
- *Robot Localisation*
  Robot localisation is the process of finding the position of the fiducials in the robot coordinate system.

H.1. Registration algorithm

There are different algorithms that can calculate the final transformation matrix. What algorithm is chosen, is mostly dependant on the different types of FLE distributions (**shamir2009localisation**). Different papers thus describe different algorithms as shown in Figure H.1. In this pair-point registration method, the leas squares method is chosen by an algorithm called the *DetermineHomogeneous-TransformationMatrix*, made by Rolf Gaasbeek. This algorithm finds the homogeneous transformation between two coordinate systems. A homogeneous transformation matrix means that the matrix combines both the rotation and translation vector into one single matrix as also mentioned in Equation 3.6 (Automaticaddison, 2020). At the same time, the algorithm is limited to only rotational and transnational movements.
The input for the algorithm are two matrices called coordinates1 and coordinates2. These matrices consist of points in a 3D coordinate system. Lets assume system A is the coordinate system of the robot world and coordinate system B is the coordinate system of the CT image. This can then be correlated to coordinates1 for the robot world coordinate system and coordinates2 for the CT image coordinate system. Each matrix consists of the coordinates for four points in 3D space, representing the four fiducials $A = \{a_1, a_2, a_3, a_4\}$ and $B = \{b_1, b_2, b_3, b_4\}$. This leads to the matrices

$$coordinates1 = \begin{bmatrix} a_{1x} & a_{2x} & a_{3x} & a_{4x} \\ a_{1y} & a_{2y} & a_{3y} & a_{4y} \\ a_{1z} & a_{2z} & a_{3z} & a_{4z} \end{bmatrix} \tag{H.1}$$

and

$$coordinates2 = \begin{bmatrix} b_{1x} & b_{2x} & b_{3x} & b_{4x} \\ b_{1y} & b_{2y} & b_{3y} & b_{4y} \\ b_{1z} & b_{2z} & b_{3z} & b_{4z} \end{bmatrix} \tag{H.2}$$

TRE estimation method	Isotropic/ anisotropic FLE distribution	Homogeneous/ heterogeneous FLE distribution	Biased/ unbiased FLE distribution	Registration method
1. Fitzpatrick et al. [6]	*isotropic*	*homogeneous*	*unbiased*	*least squares*
2. Wiles et al. [18]	*anisotropic*	*homogeneous*	*unbiased*	*least squares*
3. Danilchenko et al. [12]	*anisotropic*	*heterogeneous*	*unbiased*	*weighted least squares*
4. Moghari et al. [17], Ma et al.[16]	*anisotropic*	*heterogeneous*	*unbiased*	*maximum likelihood*
5. Moghari et al. [15]	*anisotropic*	*heterogeneous*	*biased*	*maximum likelihood*

Figure H.1: Comparison of five different methods to estimate the TRE based on their assumptions of the FLE distribution and the chosen registration method (2011). With the sources being: [6] J. M. Fitzpatrick, J. B. West, and C. R. Maurer, Jr., "Predicting error in rigid-body point-based registration," IEEE Trans. Med. Imag., vol. 17, no. 5, pp. 694–702, Oct. 1998. [12] A. Danilchenko and J. M. Fitzpatrick, "General approach to first-order error prediction in rigid point registration," IEEE Trans. Med. Imag., vol. 30, no. 3, pp. 679–693, Mar. 2011. [15] M. H. Moghari and P. Abolmaesumi, "Understanding the effect of bias in fiducial localisation error on point-based rigid-body registration," IEEE Trans. Med. Imag., vol. 29, no. 10, pp. 1730–1738, Oct. 2010. [16] B. Ma, M. H. Moghari, R. E. Ellis, and P. Abolmaesumi, "Estimation of optimal fiducial target registration error in the presence of heteroscedastic noise," IEEE Trans. Med. Imag., vol. 29, no. 3, pp. 708–723, Mar. 2010. [18] A. D. Wiles, A. Likholyot, D. D. Frantz, and T. M. Peters, "A statistical model for point-based target registration error with anisotropic fiducial localiser error," IEEE Trans. Med. Imag., vol. 27, no. 3, pp. 378–390, Mar. 2008.

The algorithm first calls the Absor function, which stands for absolute orientation function. This is also called the Horn's method and is written by Matt (2022). This Absor function finds the transformation matrix based on a least squares method for at least 3 pairs of matching points in two different coordinate systems. Horn et al. (1988) validated that his solution holds good for two of the most common representations of rotation: unit quaternions and orthonormal matrices. The objective of the Horn's method is to find the rotation, translation and scales that transform a point a_i in coordinate system A, to its equivalent point b_i in coordinate system B (1988). This is based on the equation

$$b_i = s * R * a_i + t \tag{H.3}$$

With s the scale, R the rotation and t the translation. In matching the two points a_i and b_i there will be a residual error e_i, which is expressed by

$$e_i = b_i - s * R * a_i - t \tag{H.4}$$

The method described by Horn et al. (1988) then minimises the sum of squares of these residual errors

$$(t, s, R) = \arg\min \sum_{i=1}^{n} ||e_i||^2 \tag{H.5}$$

with n the number of pair points, here the number of fiducials, which is four. This minimising problem is then solved in the following steps.

Step 1: Compute the centroids of A and B:

$$\bar{A} = \frac{1}{n}\sum_{i=1}^{n} a_i \tag{H.6}$$

$$\bar{B} = \frac{1}{n}\sum_{i=1}^{n} b_i \tag{H.7}$$

with $\bar{A}$ and $\bar{B}$ being the centroids. These are shown in Figure H.2.

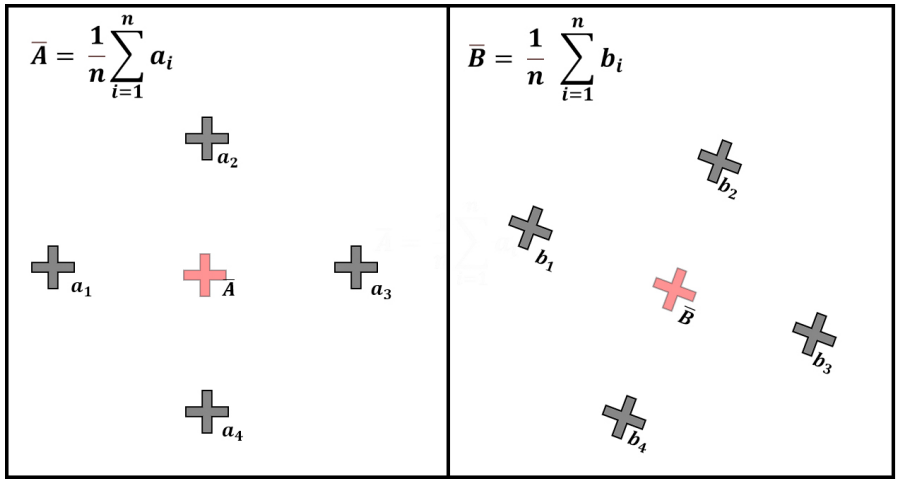


Figure H.2: The centroid positions $\bar{A}$ and $\bar{B}$, shown in light red, based on the fiducial positions $A = \{a_1, a_2, a_3, a_4\}$ and $B = \{b_1, b_2, b_3, b_4\}$ as shown in grey.

Step 2: Define points with respect to centroid:

$$a_i^{'} = a_i - \bar{A} \tag{H.8}$$

$$b_i^{'} = b_i - \bar{B} \tag{H.9}$$

which is shown in Figure H.3.

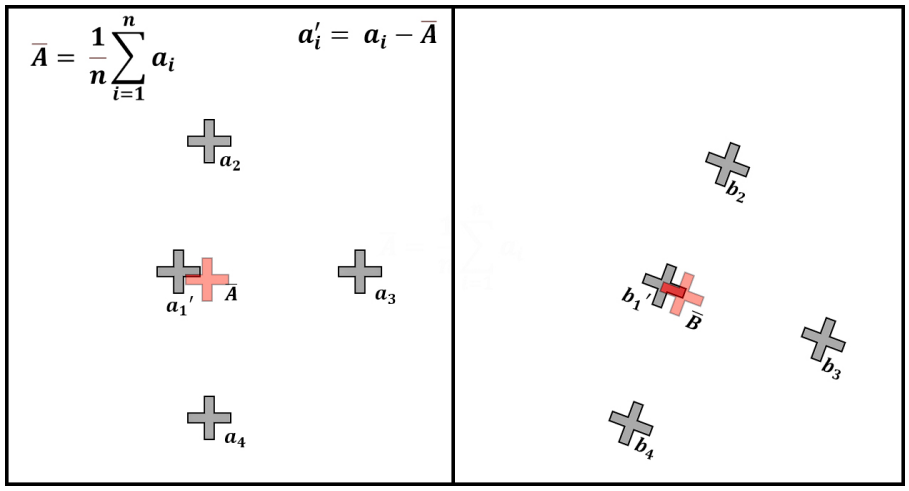


Figure H.3: Defining the points with respect to the centroid. An example of defining $a_1^{'}$ and $b_1^{'}$ with respect to their centroids $\bar{A}$ and $\bar{B}$, as shown in light red, with the formulas H.8 and H.9.

Step 3: Calculate residual error with respect to centroid:

$$e_i = b_i^{'} - s * R * a_i^{'} - t^{'} \tag{H.10}$$

with

$$t^{'} = t - \bar{B} + s * R * \bar{A} \tag{H.11}$$

Step 4: To minimise, fulfill the statement:

$$t^{'} = 0 \tag{H.12}$$

$$0 = t - \bar{B} + s * R * \bar{A} \tag{H.13}$$

$$t = \bar{B} - s * R * \bar{A} \tag{H.14}$$

and

$$\begin{aligned} e_i &= b_i^{'} - s * R * a_i^{'} - 0 \\ &= b_i^{'} - s * R * a_i^{'} \end{aligned} \tag{H.15}$$

Step 5: Solve for s by reducing Equation H.3 to:

$$\begin{aligned} (s, R) &= \arg\min \sum_{i=1}^{n} ||e_i||^2 \\ &= \arg\min_{s,R} \sum_{i=1}^{n} ||b_i^{'} - s * R * a_i^{'}||^2 \end{aligned} \tag{H.16}$$

with

$$e_i = \frac{1}{\sqrt{s}} b_i^{'} - \sqrt{s} * R * a_i^{'} \tag{H.17}$$

this becomes

$$(s, R) = \arg\min_{s,R} \frac{1}{s} \sum_{i=1}^{n} ||b_i^{'}||^2 + s \sum_{i=1}^{n} ||a_i^{'}||^2 - 2 \sum_{i=1}^{n} b_i^{'} * (R * a_i^{'}). \tag{H.18}$$

This can be written in the form of

$$\frac{1}{s} S_b = s^2 S_a + 2 * D \tag{H.19}$$

$$\begin{aligned} S_b &= s^2 S_a + 2 * s * D \\ &= 2 * s * D + s^2 S_a \end{aligned} \tag{H.20}$$

with S_b and S_a the sums of squares of the measurement vectors relative to the centroids and D the sum of the dot products of the corresponding coordinates in coordinate system B with the rotated coordinates in coordinate system A. This can be further written out as

$$\left(s\sqrt{S_a} - \frac{D}{\sqrt{S_a}}\right)^2 + \frac{(S_b S_a - D^2)}{S_a}. \tag{H.21}$$

Minimising with respect to s gives

$$\left(s\sqrt{S_a} - \frac{D}{\sqrt{S_a}}\right)^2 = 0 \quad \text{v} \quad s = \frac{D}{S_a} \tag{H.22}$$

Thus

$$s = \sqrt{\frac{\sum_{i=1}^{n} ||b_i'||^2}{\sum_{i=1}^{n} ||a_i'||^2}} \tag{H.23}$$

Step 6: Solve for R by substituting Equation H.22 into Equation H.18

$$(s, R) = \arg\min_{R,2} \left(\sqrt{\sum_{i=1}^{n} ||b_i'||^2 \sum_{i=1}^{n} ||a_i'||^2} - \sum_{i=1}^{n} b_i' * (R * a_i') \right) \tag{H.24}$$

Step 7: Calculate t according to H.14.

Step 8: Calculate homogeneous transformation matrix as

$$[s * R, t'[00...1]] \tag{H.25}$$

The determineHomoenousTransformationMatrix function also includes the option to work with fiducials that do not have a specific order. So if the order in the CT is different from the order of fiducials in the robot, then this algorithm will solve that. It will try all options of marker orders and return the one with the lowest error.

H.2. Image localisation

The image localisation side of the error-analysis includes localising the four fiducials in the CT image. This is done using 3D slicer software and a markup. The human performing this localisation is also called the operator, but is not necessarily the same operator that performs the robot localisation. The CT localisation operator has to mark all four fiducials with the markup in order. The fiducials are reflected clearly in the CT image as shown in Figure H.4.

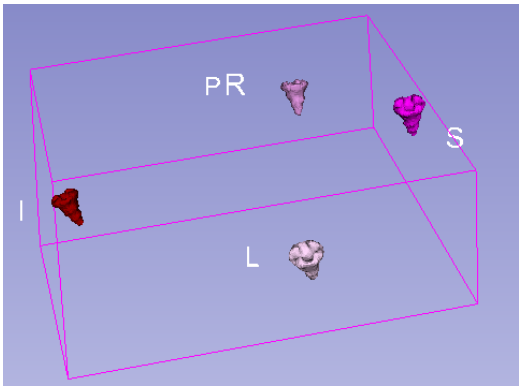


Figure H.4: An example of 3D Slicer UI with four fiducials in 3D space that have to be located during imaging localisation. Due to confidentiality reasons, no patient anatomy was loaded in this example.

Errors that can contribute to the CT localisation can add to the FLE of the CT coordinate system, referred to as coordinate system B. Since CT localisation is not the focus of this project, no detailed

analysis will be performed here, but a quick overview of possible error contributors are stated. These are classified in three categories:

- **CT State**
  This is the state of the CT image itself and errors introduced due to the CT resolution and quality.
- **Fiducial Marking**
  This are the errors introduced by the fit of the fiducial and markup with respect to each other and the amount of fiducials.
- **Operator**
  This is the state of the operator that performs the localisation on the CT side, this operator marks the different fiducials in a 3D software.

CT State
The CT scan consists of 2D images stacked upon each other, called slices. This scan has to show the fiducials and based on this scan the operator should be able to select them. The errors introduced by the CT scan state can be classified as resolution, artifacts and image noise errors. CT images have a higher resolution and geometric accuracy, and allow more exact registration than MRI data. MRI is also more sensitive to large motion artifacts and need longer scanning time and generally have lower resolution (Widmann et al., 2009). In this project it is assumed that CT is always the used method, but it should be noted that MRI most likely has the same type of errors that will be occurring. With respect to the CT state, two main categories add to the CT localisation error:

- *Resolution / slice thickness*
  The resolution of the CT scan is shown to influence the CT FLE (Regodic et al., 2020). The best achievable accuracy is directly correlated to the image resolution. The better the CT resolution in voxel size, the smaller the contribution to the FLE (W. Liu et al., 2009; Regodic et al., 2020). This is shown in Figure H.5. This is also emphasized by M. J. Murphy (1999), who states that a reduced slice thickness from 3mm to 1mm led to an increase in localisation accuracy with a factor 2.
- *Artifacts*
  CT artifacts are common in clinical CT and can affect the CT image (Boas, Fleischmann, et al., 2012). There are different types of artifacts such as image noise, scatter, motion, beam hardening, cone-beam, helical, ring and metal artifacts (2012). Some artifacts are shown in Figures H.6 and H.7. If artifacts occur at the position of a fiducial, this could influence the detection of the center position of the fiducial.

Fiducial and tip design
The fiducial marking is the process of placing a markup inside each of the fiducials in the 3D Slicer software. For this to be highly accurate, the design of the fiducial, tip and the software place a large role. The factors that add to the errors categorised as fiducial and tip design errors are:

- *Software interface*
  The software interface can play a role in the FLE since the way of scrolling through the CT slices and gaining control of the markup placement can greatly help or hammer in the localisation of the fiducials.
- *Fiducial placement*
  With respect to the fiducial placement, the same goes as for the robot localisation, which will be explained in more detail later. How the fiducials are placed with respect to the

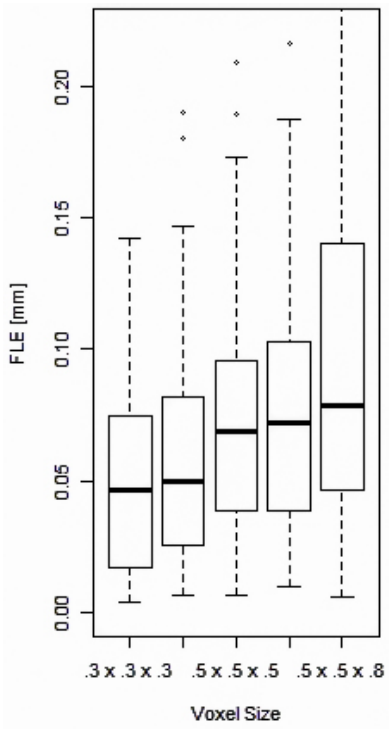


Figure H.5: Boxplots of CT FLE in mm of a fiducial screw depending on the voxel size of the CT image ranging from 0.3 x 0.3 x 0.3mm^2, 0.3 x 0.3 x 0.6mm^2, 0.5 x 0.5 x 0.5mm^2, 0.5 x 0.5 x 0.6mm^2 and 0.5 x 0.5 x 0.8mm^2 (2020).

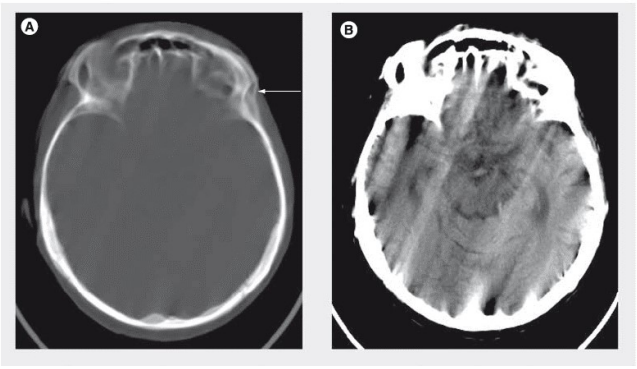


Figure H.6: The effect of motion on the CT image. Motion can cause blurring or double images, as shown on the left side (A), or long streaks as shown on the right side (B) (2012).

target point will influence the TRE. It could also be that if the fiducial is positioned in a more difficult angle for localisation that the accuracy around that fiducial is less compared to other fiducials. The fact that the accuracy can differ per fiducial is shown by Regodic et al. (2020) in Figure H.8.

- *Fiducial amounts*

  It was shown that the number of fiducials is correlated to the accuracy and TRE by multiple papers (Van Wyk and Marvel, 2017; Zhi, 2015). Its influence on the TRE will also be elaborated in more detail in the robot localisation errors.

- *Fiducial dimensions*

  This error also plays a role in the robot localisation and will be explained in more detail there. The smaller the fiducial, the smaller the max FLE. The fiducial dimensions should also be symmetrical to avoid a bias in a certain direction for CT localisation. Also it seems that a spherical fiducial type with a larger size compared to a screw fiducial type of a smaller size results in a smaller FLE as shown in Figure H.9. Thus the shape and size

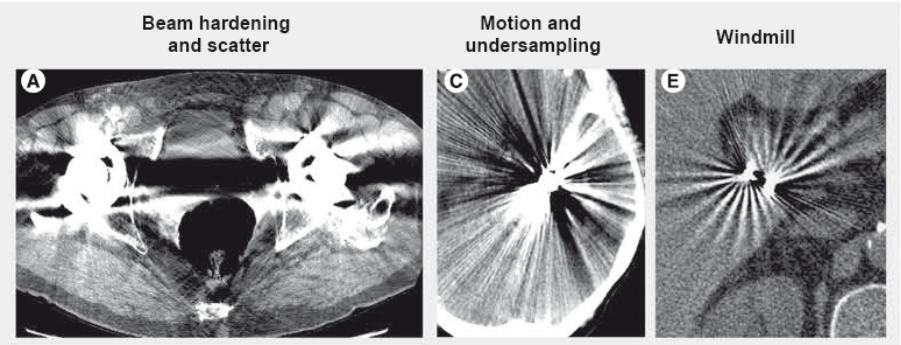


Figure H.7: An example of different types of CT artifacts. A) An example of beam hardening and scatter. B) An example of motion and under sampling CT artifacts. C) An example of a windmill artifact in CT imaging (2012).

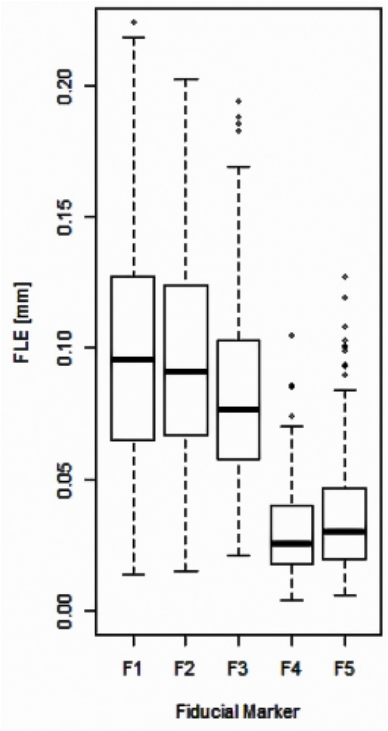


Figure H.8: Boxplots of CT FLE in mm of different fiducials screws F1, F2, F3, F4 and F5 (2020).

matters for the FLE. Therefore it should be considered that if the fiducial dimensions are changed this will influence both the FLE of CT localisation as well as robot localisation and thus largely influence the TRE.

- *Markup dimensions*

  The fit of the markup within the fiducial marker is an important configuration to minimise the FLE. Since boundaries of placement can not be felt but only seen, it is difficult to fit a markup with a perfect fit. Markups can be placed through the wall of a fiducial as shown in Figure H.10. Here it seems that the markup is placed in the center of the fiducial, but when the side view is used it is clearly visible that the fiducial wall is penetrated and thus the markup is placed more to that size. However, if the markup would be significantly smaller as compared to the fiducial walls, it would be difficult to distinct the exact same distance between the fiducial wall and the markup in all directions.

Operator State

In the found literature there is a lack of research on the role of the human operator in localising the fiducials in both the CT as well as the operator. Therefore this project will include the human factor

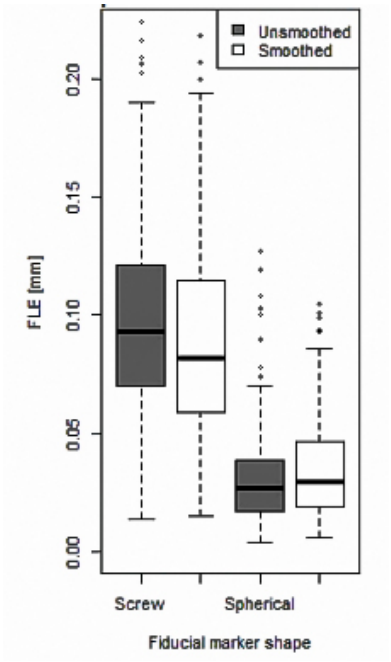


Figure H.9: Boxplots of CT FLE in mm of different types of fiducials screws. It compares either a screw shaped or a shperical shaped fiducial that is either unsmoothed or smoothed (2020).

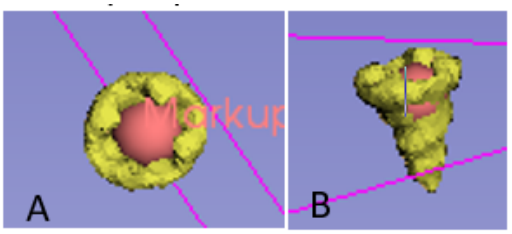


Figure H.10: An example of placing a markup inside the fiducial screw in the 3D Slicer interface. It can be seen that from the above view it seems the markup is placed perfectly, however from the side it is clearly visible the markup is going through the wall of the fiducial screw. Due to confidentiality reasons, no patient anatomy was loaded in this example.

of the operator performing the CT localisation. For CT localisation, machine learning algorithms are proven to be useful and might be able to replace human operators. However, this is not the case in the technology at hand. In order to notice what errors contribute to the FLE and TRE the human factor analysis as mentioned in Chapter 3 should be considered. Here only unsafe acts of the operator performing CT localisation which includes both errors and violations will be taken into account, thus errors like CT maintenance or inspection will not be included since CT localisation is not the main topic. The errors and violations that can add to the FLE and TRE are:

- *Omitted step in procedure*

  The operator could willingly or unwillingly forget a step in the procedure or mark the wrong fiducial with the wrong markup by switching the order.

- *Poor localisation technique*

  When the operator does not perform well in localising the fiducials with the markups due to a bad localisation technique.

- *Failed to prioritise attention*

  When the operator is distracted, rushed or tired this could increase the FLE.

- *Perceptual errors*

Since there is not haptic feedback provided to the operator, clear visibility is important. If the visibility is impaired, this could lead to large FLE for CT localisation.

- *Incorrect use of software*

  When the operator uses the software wrongly by selecting the wrong model or interface.

- *Failed to correct for administrative errors*

  When the wrong patient information is loaded, this should be noticed by the operator and corrected. Other administrative errors can also occur like saving the data wrongly.

- *Failed to correct for CT errors*

  Mitigating for CT errors can be done in part by the operator by enhancing the image quality and contrast, if the operator fails to do so this could lead to large FLE.

H.3. Robot localisation

The robot localisation that is taken place is done by using the robot end-effector in a admittance control loop in this case. This is showed in a simplified image in Figure H.11

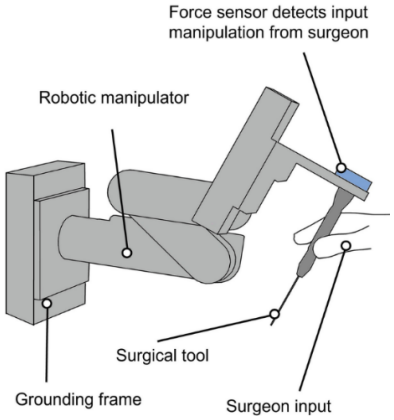


Figure H.11: An schematic example of robot localisation with admittance control. The surgeon holds the robot and a force sensor translates the input to a motion of the surgical tool. The surgeon needs to position the surgical tool inside the fiducial screws as accurately as possible (2021).

The human performing this admittance control task is called the operator and this operator positions the robot to the four fiducials that are attached to the patient on the operating table. Errors that can contribute to the robot localisation can add to the FLE of the robot coordinate system depends on a lot of things like the accuracy of the tracking system, the design, number and arrangement of the fiducial markers, and the image data (Eggers et al., 2006). These errors are classified in four categories:

- **Robot state**

  This is the state of the robot itself and errors introduced due to the robot state.

- **Patient state**

  This is the state of the patient lying in the OK and errors related to this state.

- **Operator state**

  This is the state of the operator that performs the localisation on the robot side, this operator guides the robot to the fiducials.

- **Fiducial and tip design**

  This are the errors introduced by the fit of the fiducial and registration tip with respect to each other and their individual dimensions and amount.

Robot state

From the robot localisation, the robot itself can introduce positioning errors that can lead to an addition in FLE for registration. These errors are classified in either kinematic or non-kinematic errors.

Kinematic errors:

Kinamtic errors are errors that are introduced in the system due to imperfect geometry, alignment and dimensions calculations of robot components. This is a measure for the discrepancies between the physical robot end-effector and its kinematics model, which means the difference between the calculated and actual location of the end-effector (Woodside et al., 2020). These kinematic errors cause incorrect values of the Denavit-Hartenberg (DH) parameters in the kinematic model of the manipulator (Franaszek and Cheok, 2020). Kinematic errors, directly add to FLE values when touching the fiducials in robot space since there will be a difference between the ground truth position of the robots end-effector and the measured position of the end-effector. This leads to a certain systematic error which we will call systematic error 1. If no other errors occur, this will lead to

$$FLE_1 = FLE_2 = FLE_3 = FLE_4 \tag{H.26}$$

and this is shown in Figure H.12. This systematic error will influence the TRE largely and its influence will be researched more with a simulation.

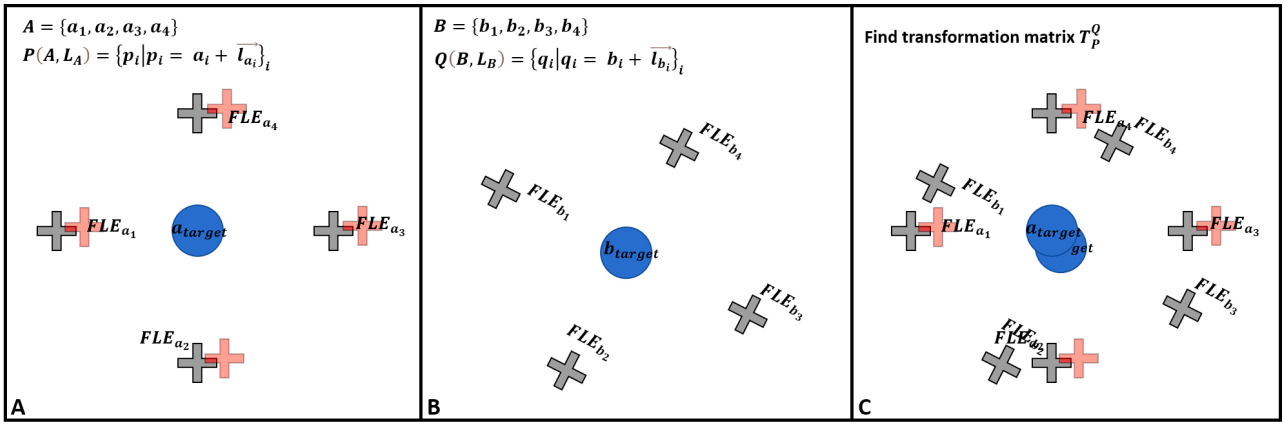


Figure H.12: An example of the influence of the systematic FLE error in the robot localisation. This will lead to the same FLE for all fiducials in the real-world coordinate system and this will be placed on the fiducials of the imaging coordinate system. Then the target position will likely also be biased in the same direction. More on systematic errors will be explained later.

Kinematic errors can be validated and measured with the help of a laser tracker system as described by Zhang et al. (2014). Kinematic errors are introduced by:

- *localisation and calibration*

  Robot localisation is the process of determining where a robot is located in its environment, here especially the localisation of the end effector is important. This localisation depends on the kinematics model and calculations of the robot. If these calculations contain errors then the end-effector position will have errors as well. In this case this is done using forward kinematics and finding DH parameters with calibration. Unrealistic DH parameters can lead to a localisation error of the robot end-effector.

- *Drill or bit change*

  If a new drill or bit is used on the robot end-effector who has a different size or dimension as compared to the previous drill, this could lead to both calibration as well as localisation errors since this new drill might need to adhere to another model as well as re-calibration.

- *Assembling or setup*

  During assembly, when the robot joints, links and end-effector are not correctly assembled it will differ from the robots internal model and thus the end-effector calculated position will differ from its actual position. In setting up the robot, the sterilization process can also add errors if it is not done carefully and damages the registration bit.

- *Wear*

  Due to wear of the robot manipulator the robots model could also differ from the actual geometrical characteristics of the robot.

Non-Kinematic errors:
These are errors attributed to the mechanical characteristics of the robots component or the effect of forces acting on the robot. These errors are mostly not systematic since they can differ per fiducial due to a certain force being more or less in time, but can also be systematic if a certain bending in the arm link takes place during the entire registration process.

A lot of non-kinematic errors can be analysed using a finite element analysis (FEA), which is a method to predict the reaction of an engineering design to external forces (Shanmugasundar et al., 2019). The FEA solves partial differential equations in space by dividing a larger system into smaller parts called finite elements. It needs a detailed model of the robot manipulator but when this is done correctly, it can show the deformation and stress on the manipulator in response to certain external forces. And it should be said that due to it being a model, the FEA will never be 100% accurate, but it is shown to perform well in many applications. An example of an FEA is shown in Figure H.13.

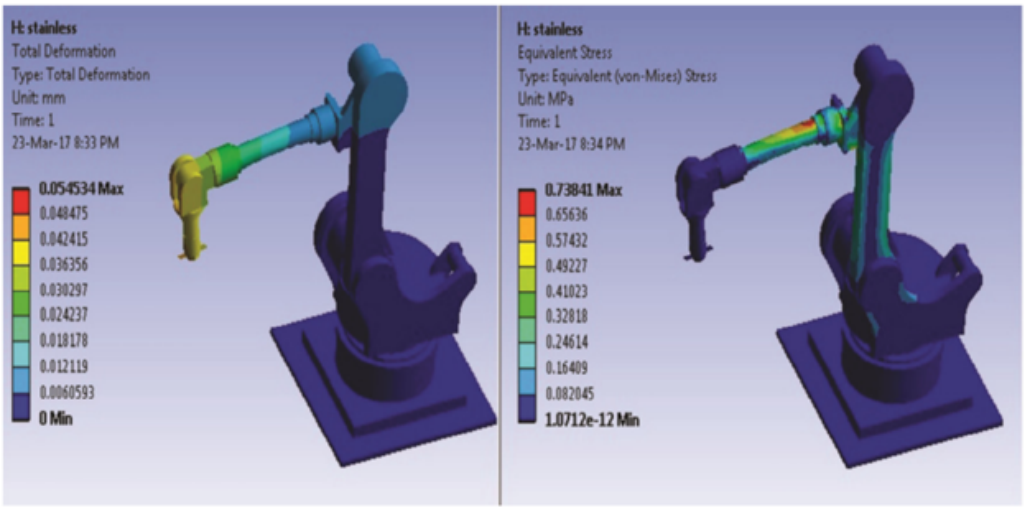


Figure H.13: An example of a FEA on a 5 DOF robot arm, which shows the total deformation and stress produced in using the robot when it is made of stainless steel. The color is an indicator of the amount of deflection, with red being the highest and dark blue being the lowest deflection (2019).

Non-kinematic errors includes the following errors:

- *Environmental factors*

  Environmental factors such as temperature and humidity variations can change the configuration of the robot manipulator in its different parts or geometry (Lubrano and Clavel, 2010). This could lead to a difference in calculated or estimated spatial position and actual spatial position. This could differ from day to day or even throughout a registration process. It is stated that thermal expansion does not influence the rotation measurements of the robot end-effector but will affect the translation position measurements (2010). The amount of thermal expansion is highly dependent on the types of materials chosen.

During the registration itself, it is important to keep the environmental temperature stable and relatively cold, which is already the case in ORs. However, the drilling could add temperature to the robotic system and cause local thermal expansion. Since this occurs after registration, this will not be elaborated further. It is stated by Vocetka et al. (2021) that high accurate robotic systems assembly is sensitive to temperature change, thus thermal expansion can already play part in the assembly phase.

- *Compliance or bending of robot manipulator*

  Compliance or bending in the system can occur due to deflection in the robot manipulator structure due to the application of a certain force. Some form of compliance could be the bending of the arm links. This leads to positioning errors at the end-effector. The stiffness of mechanical components and friction means the robot links and gearboxes. The stiffer the robot mechanical components, the less bending or compliance will occur.

- *Singularities*

  When registration has to be performed near a singularity point, the robot control can be limited and thus the position can have an error. In this particular case, it is not an issue due to the design of the robot to work in a singularity-free region.

- *Mechanical backlash*

  Backlash is the play between gear tooth in the gearboxes in a robot joint, which introduces a spatial error between the commanded position, and the actual position as shown in Figure H.14. Giovannitti et al. (2022) describes a method to measure mechanical backlash with-

Figure H.14: A schematic example of the mechanical backlash gap between mating teeth of two gears (2022).

out the addition of an extra device. Their method observes the vibration patterns that arise on the motor speed signal when backlash affects the joints. It is stated that when backlash occurs, the excessive space causes oscillations that can be measures by the motor speed signal, the encoder (2022). The type of oscillation that occurs is specific to backlash and therefore can be distinguished from other oscillations. The amplitude of the oscillation is directly related to the amount of backlash in the gear (2022). In Figure H.15 the disturbance pattern for backlash at different amounts is shown. Since the maximum backlash is the difference in teeth and gaps of two gears this can be calculated. However, this backlash can occur at each joint and since the robot contains five joints, the total amount of backlash could become larger. The paper by Slamani et al. (2012) describes that in bad cases the amount of backlash can reach up to 134 μm, this is not added directly to the end-effector localisation error but it is stated to be related to the robot configuration by a polynomial model of degree two in a 6 DOF industrial robot (2012). The error introduced by mechanical backlash can either be a systemic error that is the same throughout the registration process, or can change in time.

In this project it is assumed there is not mechanical backlash due to the usage of brush less stepper motor.

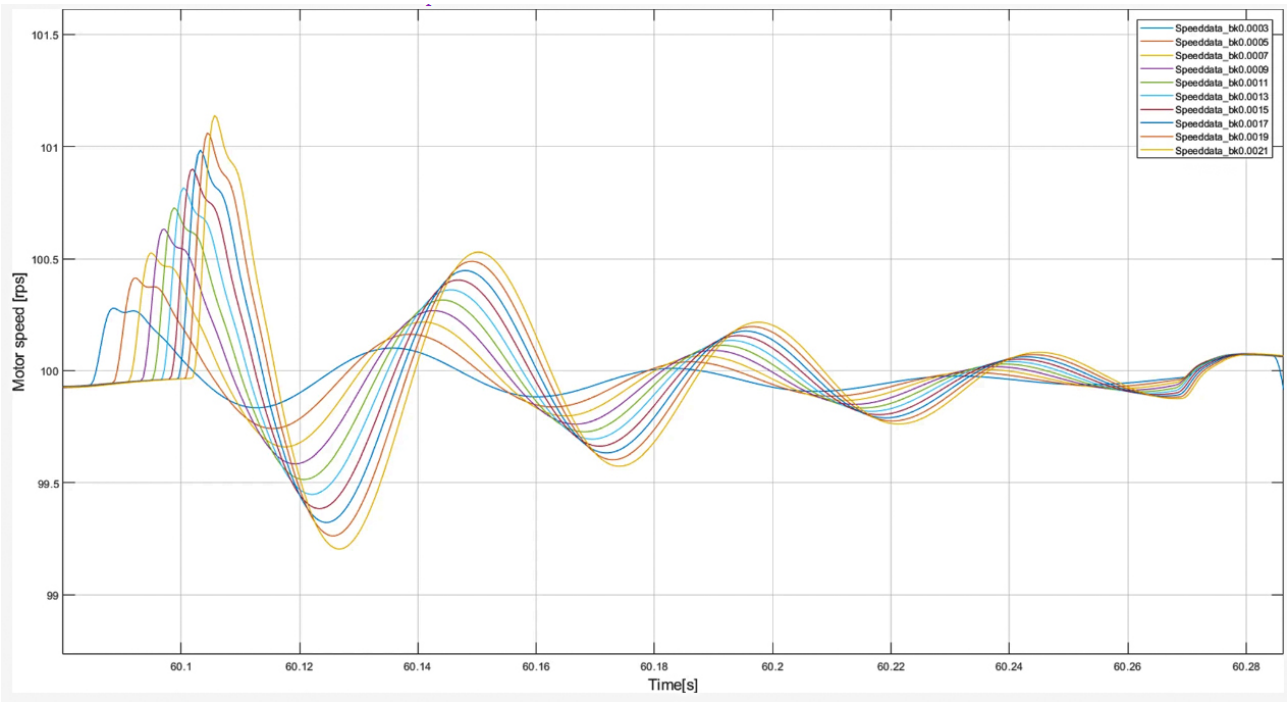


Figure H.15: Different disturbance patterns for different amounts of backlash. The 10 signals correspond to the 10 different backlash values considered in this simulation (2022).

- *Errors induced by joint angle measurement sensors*

  These errors can lead to a difference between the reported movement as registered by the sensors/encoders and the actual movement made by the joint gives rise to this error. These sensors can also be sensitive to low frequency noise or vibrations. Also non-linearities and eccentricity are included in this categorisation.

Patient state

The patient state in the registration process refers to changes in the patients positioning during registration. These errors related to the patient state are categorised in the following three categories:

- *Head fixation*

  This error is introduced due to the head movement of the patient during registration. If the head is moved before registration, it should not influence the FLE. If the patients head moves randomly throughout the registration, this could add a FLE in different directions for each fiducial and influence the TRE.

  However, if the patients head is moved after registration, this could lead to a systematic localisation error during surgery and should be considered as well. The movement of the patient after registration but before surgery could be treated as a systematic FLE for each fiducial in the same direction. This effect will be analysed more in the error simulation section.

  Head fixation in surgery can be done in different ways, either invasively or non-invasively and the best method is outside the scope of this project. In general, the more invasive, the less fluctuation is likely to occur.

  One way to estimate head fixation errors is the placement of reflective markers on a stable anatomical reference point of the patient (Carminucci et al., 2018). This marker could be tracked by an infrared camera in time from a fixed point at the table so that a relative

movement with respect to the table is detected. An additional infrared camera that tracks the marker from a fixed point on the robot could also state what the relative movement is with respect to the robot.

- *Head position*

  The position of the patients head could enlarge or reduce the amount of influence the FLE could have on the TRE. This is more so dependent on the relative placement of the fiducials with respect to each other and with respect to the target point rather then the actual placement of the patients head. However, when positioning the patients head, it should be considered what the robot workspace is and where its singularity points lie in space. These points should be avoided at all time.

- *Table fixation*

  This error is introduced due to the movement of the patients table during registration. If the table moves, this could increase the FLE of one or more fiducials. This error is similar to the head fixation in its effects both during and after registration.

  Table fixation errors could be mitigated by placing the robot on the table so that if the table moves, the robot and patient do not move with respect to each other. However, this could be difficult and dangerous. Therefore, the same placement of reflective markers could be used on the table, that is tracked from a fixed point on the robot so that a relative movement with respect to the robot is detected.

Fiducial and tip design

It was shown by West et al. (2001) that fiducials do have a significant role in minimising the TRE. It is stated that using as many points as possible, avoiding near-collinear configurations and ensuring that the centroid of the fiducial points is as near as possible to the target, can minimise the TRE. This is also underlined by Zhi (2015). The errors that can contribute to the TRE from the fiducial and tip design are categorised in the following four categories:

- *Fiducial amount*

  Since the RMSE is used to defined the FLE and TRE the more fiducials used, the smaller each individual FLE contributes to the final error. For the calculation of the transformation matrix, the centroid of all measured fiducial locations is used, this centroid becomes less influential by one individual point if more points are being used.

  As described by Van Wyk and Marvel (2017) the minimum number of fiducials needed is 3, to limit the amount of DOF and only provide one option for registration, but the addition of more redundant fiducials can increase the registration accuracy. Zhi (2015) shows that the amount of fiducial points is correlated to the FLE and from 17 fiducial points on the accuracy does not improve much on the FLE. Additional accuracy can then be reached by combining different registration techniques (Eggers et al., 2006). This is also emphasized by Bao et al. (2014) who concludes that from six fiducial markers on the accuracy is already 86 % as also shown in Figure H.16.

  The more fiducials, the less influence a certain variance or error in localising that single fiducial has and thus the more accurate the TRE. It should also be considered that if a certain fiducial gets damaged, an abundant amount of fiducials could mitigate problems that can occur (Shamir et al., 2011). Additional points can also be used to obtain multiple transformation matrices. For instance, four fiducial points can be used in different combinations of three pair points and yield different matrices. From these matrices, their average

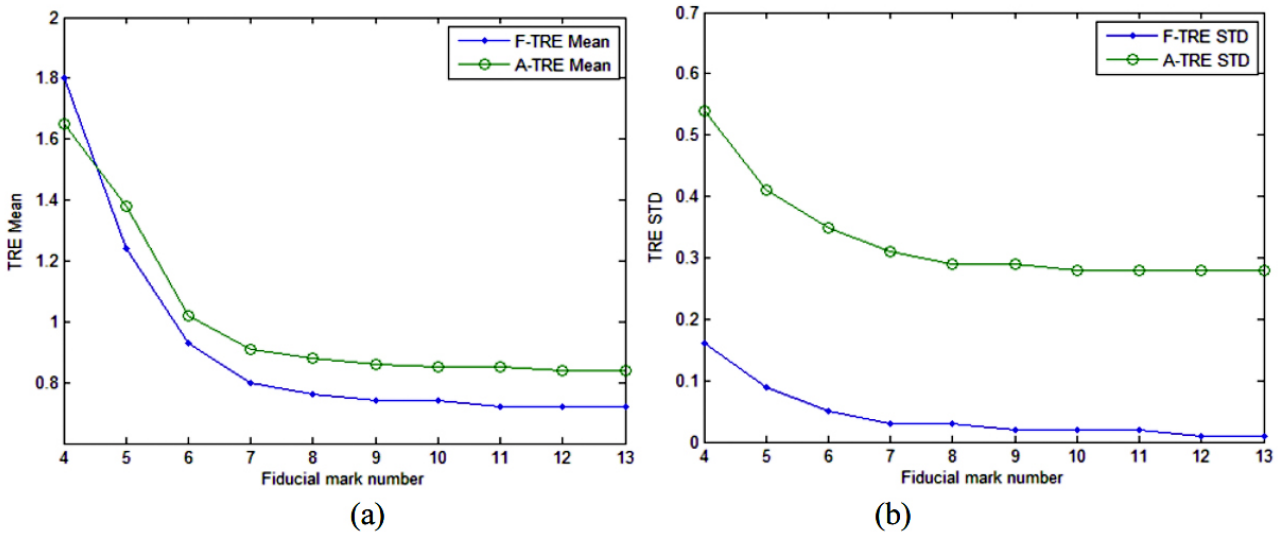


Figure H.16: The effect of the number of fiducial markers on the TRE mean **(a)** and TRE STD **(b)**. A distinguishment is made between the actual TRE (A-TRE) or FRE (A-FRE) and the measures TRE or FRE with the method of Fitzpatrick and West (2001) (F-TRE and F-FRE). It can be seen that both TRE and FRE decrease with increasing number of fiducials (Bao et al., 2014).

homogeneous transformation matrix can be used to get a more accurate result (Van Wyk and Marvel, 2017). Selecting which subset of fiducials should be used to minimise the TRE is described by Shamir et al. (2011) which minimises the estimated TRE.

$$(A^*, B^*) = \arg \min_{TRE_{estimated}} (a_{target}, A^{'}, B^{'}, FLE) \tag{H.27}$$

One method to find the optimal pairing is to enumerate all possible pairings and compute the TRE for each of them and then select the subset that yields the smallest TRE value. This is computationally heavy, but for a small amount of fiducials, smaller then 12, it is feasible within a short amount of time (2011).

Additionally, placing an additional marker somewhere at a known position, either with a reflective component and infrared camera or measuring with another measurement equipment, can be used as a validation point for the robot localisation position.

- *Fiducial placement*

  As stated before, the placement of the fiducials does influence the TRE. West et al. (2001) states four guidelines with respect to fiducial placement:

  1) avoid linear and almost linear configurations. In full symmetrical placement of fiducials, the algorithm can not make a distinction between different fiducials in space and interchanging can occur. Therefore the results will be unreliable. The results will also be unreliable with collinear locating the fiducials, which means the placement of the fiducials on one line and should also be avoided (Shamir et al., 2011).

  2) Arrange the centroid of the fiducial configuration as close as possible to defined important targets. It is suggested to place the fiducials with a wide spacing between them and arranging them such that their centroid approximates the target (Labadie et al., 2004). It is stated that the TRE increases when the distance from the centroid of all fiducial markers increases (Eggers et al., 2006). Moreover, if the fiducials are already biased towards a certain space around the target, the TRE will also be biased in that direction. This is

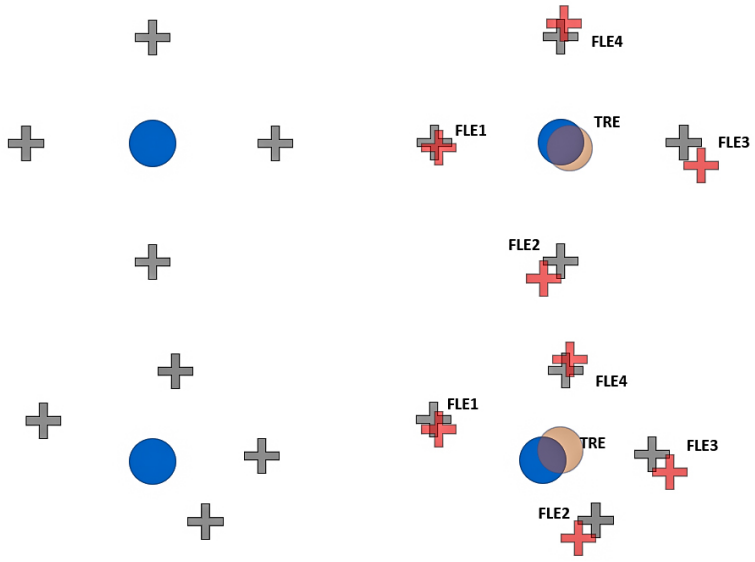


Figure H.17: A schematic example of the influence of fiducial placement on the TRE. **A)** A non-biased placement of the four fiducials, as shown in grey, surrounding the target, as shown in blue, leads to a non-biased TRE. The measured fiducial positions are shown in red and are randomly distributed around the actual fiducial positions. The measured target position after transformation is shown in orange. **B)** The biased placement of fiducials, as shown in grey, to the right side of the target, as shown in blue, leads to a biased TRE. The fiducial measured positions are shown in red and are randomly distributed around the actual fiducial positions the same as in A. However, since the fiducials are all placed more to the right of the target and placed more to the top of the target, the TRE, as shown in orange, is biased to the right top side.

sketched in 2D in Figure H.17. Placing the fiducial markers further away from each other also lowers the spatial correlation between the TRE and the distance from the fiducials to the coordinate system origin, as shown in Figure H.18 (Van Wyk and Marvel, 2017). Here a large spread of 3 points has a smaller spatial correlation with respect to a small spread of 3 points. This means that the relation between the TRE and the distance of the target point with respect to the origin is smaller when the fiducials are placed further away from each other.

3) Arrange the markers as far from each other as possible. Spreading the fiducial markers lead to a 30 percent reduction in the mean TRE as stated by Bao et al. (2014). This is also reflected in Figure H.18

4) Use as many markers as feasible. As stated in the fiducial amount section above.

Thus for minimising the TRE, the target point should be close to the center of all the fiducials and the fiducials should be placed as far as possible from each other. The fiducial placement and selection can be optimised according to Shamir et al. (2011) to minimise the TRE. The steps they propose to improve are shown in bold in Figure H.19. In their paper different studies are mentioned that optimise the fiducial placement which leads to TRE improvement ranging from 1.5 mm to 0.5 mm (2011). However, these methods are mostly tested on phantoms or simulations and therefore do not provide information about intraoperative TRE and use the TRE estimation calculation by Fitzpatrick and West (2001) which is not always a good estimator for the actual TRE (Shamir et al., 2011). Their method also provides a so-called TRE map of a head, which shows depending on the target position, where fiducial placements are good or bad. An example of this TRE map is shown in Figure H.20.

Registration Method	r	p
Small 3-point	0.855	0.000*
Large 3-point	0.687	0.000*
Large 5-point	0.141	0.214
SA Small 3-point	0.546	0.000*
SA Large 3-pomt	0.513	0.000*
SA Large S-point	−0.109	0.337
Clustering	0.316	0.004*

Figure H.18: Correlation Coefficient (r) and the statistical p-Value of the relation between the TRE and the fiducial placement distance from the coordinate system origin. * is an indication for statistical significance. It can be seen that the larger the distance, the smaller the correlation between TRE and distance from fiducial placement to the orgin. Moreover, it can be seen that the more fiducials are used, the smaller the correlation. Small 3-Point: The 3-point registration method trained using data collected over a 250 × 175 mm^2 area. Large 3-Point: The 3-point registration method trained using data collected over a 750 × 525 mm^2 area. Large 5-Point: The 5-point registration method trained using data collected over a 750 × 525 mm^2 area. SA Small 3-Point: Small 3-point, above, optimised using a SA method. SA Large 3-Point: Large 3-point, above, optimised using a SA method. SA Large 5-Point: Large 5-point, above, optimised using a SA method. Clustering: A clustering method described, with two clusters based on four registration data points taken over a 750 × 525 mm^2 area (2017) .

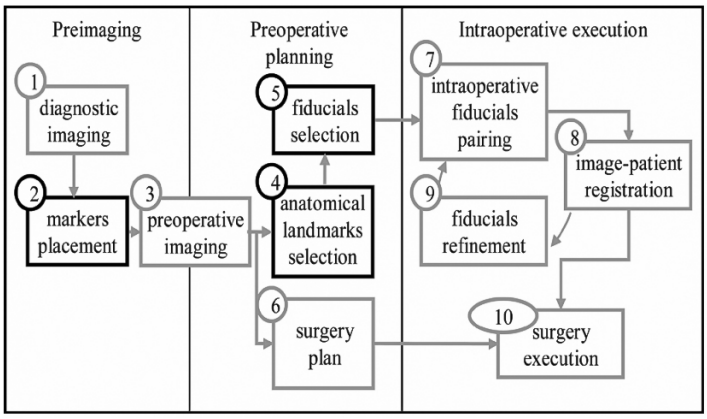


Figure H.19: A workflow overview for pair-point-based registration divided into the three main planning phases in surgery planning: pre-imaging, pre-operative planning and intra-operative planning. The bold blocks are the proposed protocol steps to improve the accuracy for pair-point-based registration (2011) .

The map shows zones instead of exact locations since placing it on an exact location is difficult. The TRE map is calculated based on diagnostic images obtained well before surgery takes place. The fiducial markers subset that minimises the TRE can be different at the time of surgery due to changes in image quality, errors in marker placement, and morphological changes that can affect the target location (2011). A way of defining the marker placement is done by evenly sampling points on the head surface extracted from the CT scan to obtain a set of potential fiducial marker locations. Then for each possible subset of 3-12 markers, the estimated TRE is calculated (2011). The optimal landmark set on a small set of potential marker locations sampled on the surface is calculated. Then for each fiducial marker its four descendants are found and the estimated TRE is calculated for all possible combinations. The one with the smallest estimated TRE is selected.

- *Fiducial dimensions*

  Although nothing can be found in the literature with respect to the fiducial dimensions it is expected to play a part in the TRE. With respect to the fiducial placement, it seems logical that the fiducial dimensions should avoid a bias in a certain direction while not being symmetrical.

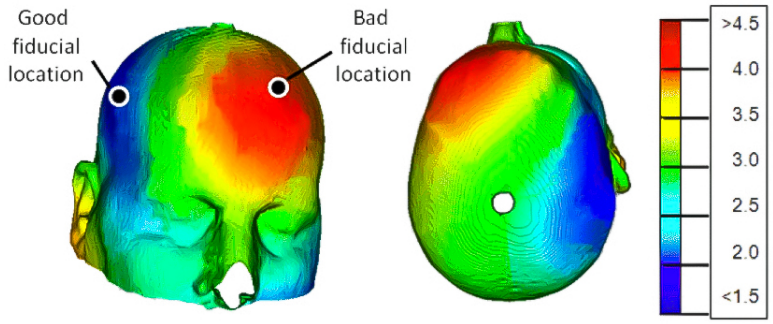


Figure H.20: A 3D colourmap of the human head showing the expected TRE by placing fiducial markers at two locations on the head surface area. The different colors are an indicator of how well the performance is of placing fiducial markers in those areas. Red means a high expected TRE in mm and blue indicates a low expected TRE in mm. This is based on the chosen target position as shown in white. (2011).

Moreover, the less possible variance of the fiducial markers, the less the FLE will be, and thus the less the TRE will be. How the max FLE is estimated from a fiducial design is shown in Figure H.21. Therefore, it seems that smaller markers will give a lower TRE. Because when physically touching the fiducials, the human-robot admittance control will be unlikely to touch next to the marker and thus the FLE will be limited to the dimensions of the fiducial. So for a fiducial with only a possible FLE either above or below the middle of the marker will give the following max FLE: However, a cross-like fiducial introduces another dimension of possible FLE and will lead to other max FLE. However, FLE can occur anywhere within the dimensions of the marker.

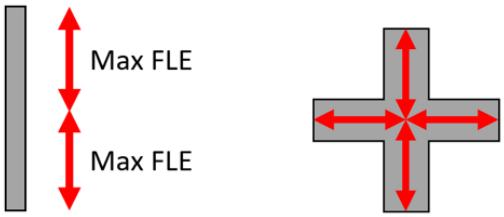


Figure H.21: A schematic example of the max FLE depending on the fiducial marker dimensions. The maximum deviation from the fiducial center point depends on the length and width of the fiducial marker. On the left side the FLE is limited to one direction since the fiducial marker is very small. However, on the right side the FLE can be large in either top-bottom direction or left-right direction.

- *Tip dimensions*

  The fit of the tip within the fiducial marker is an important configuration to minimise the FLE. If the fiducial and tip fit perfectly only in one way, no FLE would occur as shown in Figure . However, this would have implications for the easiness of placing the tip inside the fiducial. Moreover, if the tip contains a ball it means that no matter what position the tip is held to touch the middle, the same point will be reached. However, if that is not the case, the rotation can play a major part in the TRE as shown in the figure below. In addition, if the fiducial height is increased this would imply that the fiducial and tip fit is more difficult to reach and if a certain rotation is taking place, this would lead to a FLE.

Operator state

Despite a lack of research found on this subject, the operator plays a role in adding FLE to the measurements and thus influences the TRE. Therefore, this project will include the human factor of the operator performing the robot registration. In order to notice what errors contribute to the FLE and

Figure H.22: A schematic example of the influence of the tip design with respect to the screw for the maximum amount of FLE to take place. Here the fit is perfect thus no FLE is mechanically possible to take place. However, a perfect fit limits the ability to rotate the tip with respect to the fiducial screw.

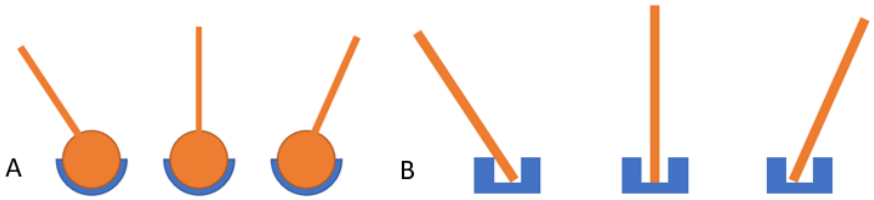


Figure H.23: A schematic example of the influence of the tip design with respect to the screw for the maximum amount of FLE to take place. **A)** An example of a round tip and fiducial design that fits as a ball-in-socket principle. Here it can be seen that the orientation of the tip does not matter for the fit inside the screw and the position measurement. **B)** An example of a square shaped fiducial and tip with a non-perfect fit such that orientation is possible. It can be seen that here the orientation leads to a different position of the tip end inside the fiducial screw.

TRE the level of failures as mentioned in the HFACS should be considered. For this project only the failure level of unsafe acts of operator will be considered. This includes both errors and violations for robot localisation. The possible errors and violations found in this project are:

- *Omitted step in the procedure*

  If a certain step in the procedure is not followed well, it could mean that a certain fiducial could be not well registered. This could occur when: a certain fiducial is skipped, a certain fiducial is registered twice, switching to registration bit does not occur, failing to put robot in registration mode, reading the x, y, z by hand etc.

- *Incorrect use of admittance control*

  Not using or misusing the admittance control and setting the robot to a position in another way. The operator could for instance position the robot through the GUI with x, y and z, adjust the gains to other values etc.

- *Failed to prioritise attention*

  Being attentive to other things and therefore not paying enough attention to complete a highly accurate localisation task. Examples of attention failures could be: not realising the tip is off, marking the wrong fiducial, hitting the patient, using the wrong registration bit, pressing the wrong button.

- *Incorrect use of workspace*

  The operator could go outside of the workspace and hit a singularity point this could damage the robot and might influence the position measures of the robot.

- *Poor localisation technique*

Having a poor technique to perform localisation with the robot. Humans are shown to have different reaching strategies depending on target size and distance as shown in Figure (Peternel and Babič, 2019). In addition, it is also shown that different humans can have different FLEs (W. Liu et al., 2009). It was found that the larger the target, the more

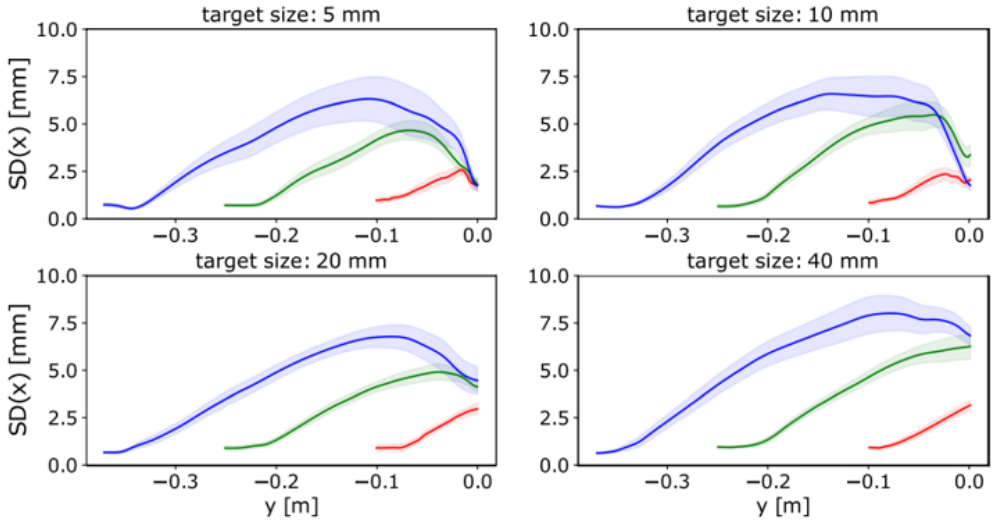


Figure H.24: An example of human localisation techniques in space when a certain target at a certain distance with a certain dimension has to be reached. Different trajectories and their SD along one axis that is perpendicular to the main movement as a function of the position in the movement axis. The solid line shows the mean value whereas the shaded area surrounding it shows the deviation from that mean value. The different colors show different distances between the starting position and the target position: blue is long distance, green is medium distance and red is short distance. This is shown for four different target sizes: 5mm, 10mm, 20mm and 40mm (2019).

deviation takes place from the optimal path this is also the case for the larger the distance to the target. However, if the target size is small, the path will deviate back to a small trajectory near the target (2019). Moreover, it was found that the smaller the target size the more time it took for the operator to complete the task. It can be concluded that for a target size of 5mm and a target distance of 10cm the maximum deviation was around 2.5mm.

Examples of poor localisation technique errors or violations are for instance: only touching the outside of the marker, only touching half the marker, large possibilities in localisation paths, using the wrong angle of attack, putting too much force, moving too fast, etc.

- *Communication error*

  When another human is needed to perform registration, the communication has to be good. If communication is lacking, it could be that the FLE will increase due to misplacement of the final tip position. Communication errors that can occur could be: selecting the wrong fiducial, not timing correctly, etc.

- *Human tremor*

  Physiological tremor exists in all human with amplitude lying in the frequency range of 8–12 Hz (Naik and Rube, 2014). If not filtered, human tremor can have an influence on the FLE.

- *Timing of final registration*

  If the final registration point is not timed correctly, this could influence the FLE. For safety reasons, the robot is programmed to go back to a setpoint when no force is acted on it by a human operator. Therefore, if the final registration is not timed correctly, the tip could already be moved away from the fiducial.

- *Perceptual errors*

  Perceptual errors are errors due to failing to see or feel the fiducial markers by the operator. If the human cannot see or feel the middle of the fiducial clearly, it can lead to an increase in FLE.

- *Incorrect working with materials*

  The operator could fail to work with proper materials, like not replacing a damaged registration bit, using too much force on the tip, not properly attaching the registration bit, damaging the registration bit or a fiducial, etc.

- *Incorrect response to robot error*

  When a robot error occurs, like it stops due to high torque or singularity points, the operator should handle accordingly. If the operator does not do that, this could lead to additional error. For instance: failing to restart the robot, failing to restart the registration procedure etc.

- *Incorrect response to patient errors*

  In the workflow of getting the patient ready for surgery, some errors can occur previously that the operator has to respond to. This could be errors like not having sufficient fiducials, wrong location of fiducials, wrong placement of fiducials, patients head having dimensions outside of normal head fixation, patient location outside of workspace, etc.

H.4. TRE simulation

Based on a matlab simulation, it was estimated how TRE increases with an increase in fiducial localisation variance for four fiducials placed around the surgical area based on real CT data. These results are shown in Figures H.25 and H.26

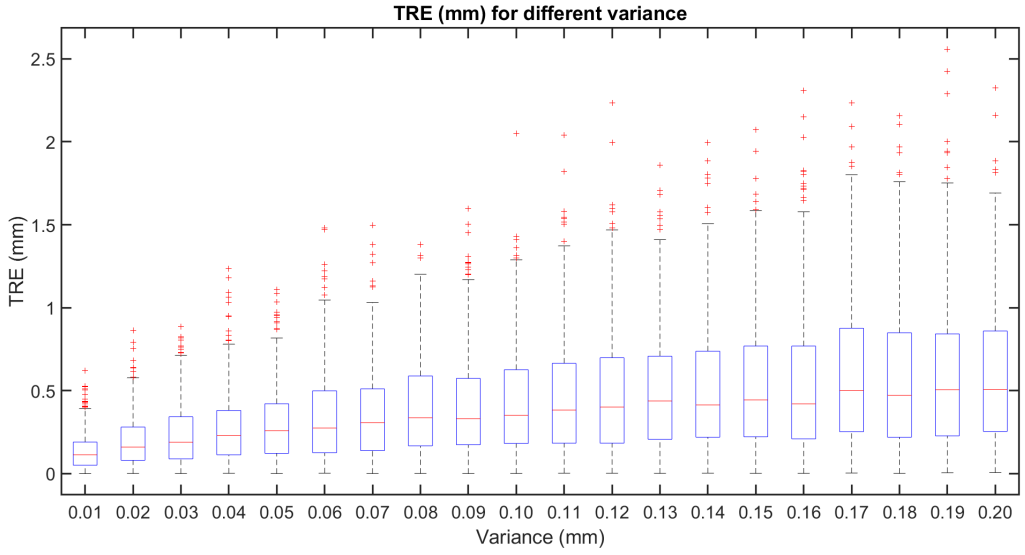


Figure H.25: Mulitple boxplots showing the distribution of TRE values in mm depending on the variance surrounding fiducial localisation based on a Matlab simulation (hence the variance of normal distributed noise surrounding one fiducial). The x-axis shows the different variance values ranging from 0.05 to 0.20. The y-axis provides the TRE value, ranging from 0 to 2.5mm. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers.

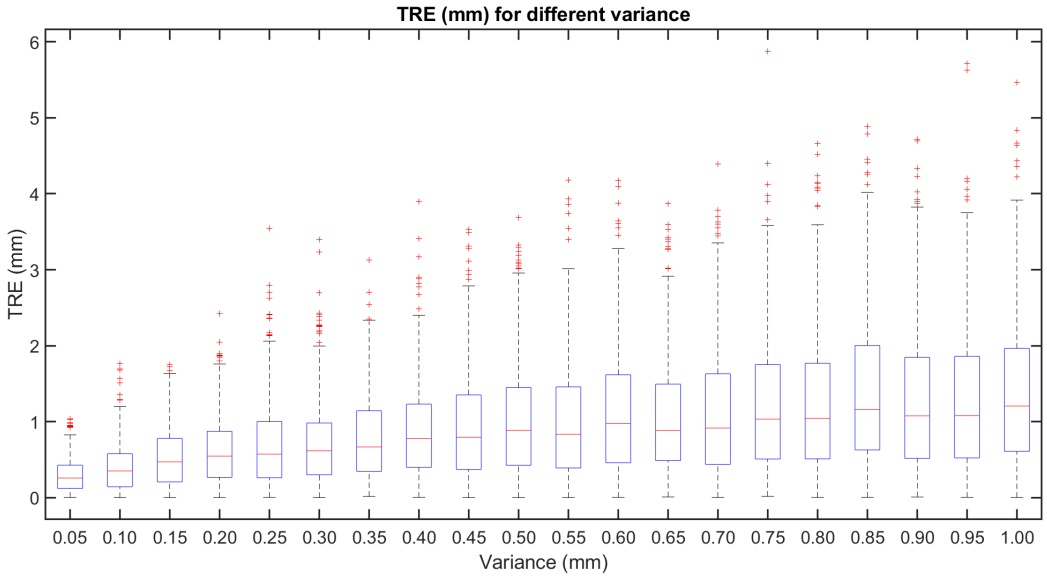


Figure H.26: Mulitple boxplots showing the distribution of TRE values in mm depending on the variance surrounding fiducial localisation based on a Matlab simulation (hence the variance of normal distributed noise surrounding one fiducial). The x-axis shows the different variance values ranging from 0.05 to 1.00. The y-axis provides the TRE value, ranging from 0 to 6.0mm. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers.

I. Concepts

This Appendix provides information on the drafted concepts and their scoring.

This chapter starts with an overview of the different concepts followed by the scoring details.

I.1. Concept selection

Three methods did not fail any criteria: touch-based, camera-based and head-fixation frame. From these criteria, four concepts were drafted and scored along the scoring criteria: touch-based bone-anchored fiducial pair-point-based with admittance control, touch-based bone-anchored fiducial pair-point-based with tele-operation, camera-based bone-anchored fiducial pair-point-based with calibration balls and head-fixation frame-based registration through ball in grooves. Each concept is described and provided with their disadvantages, advantages and innovativeness.

I.1.1. Touch-based bone-anchored fiducial pair-point-based with admittance control

Since BFR has proven to be the gold standard for registration accuracy this concept uses these fiducials. To localise these fiducials, the user has to move the tip of a tool attached to the end-effector of the robot towards the fiducials screws and tap them for their location as sketched in Figure I.1. The movement of the tool is achieved with a force sensor in the end-effector and admittance control that enables the human to move the robot through the sensed interaction force at the robot's end-effector. Consequently, if the operator exerted a forwarded force on the robot, it moved in that direction proportionally with the preset admittance parameters.

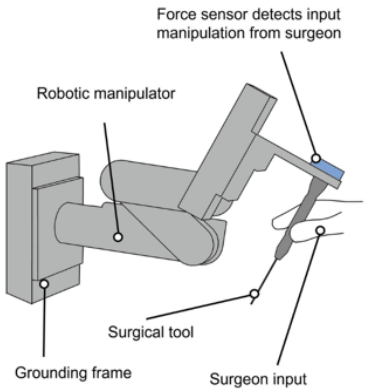


Figure I.1: Example of bone-anchored fiducial registration with admittance control where the surgeon its input is used through admittance control to guide a surgical tool (2021).

The disadvantages of this concept are:

1. User has to move the robot, can lead to damage and hazardous situations
2. Invasive
3. Difficult to automate
4. Human errors

The advantages of this concept are:

1. High accuracy
2. HITL provides more trust
3. HITL provides error detection

The innovativeness or scientific value of this concept is the inclusion of a HFA for both the designing as well as testing.

I.1.2. Touch-based bone-anchored fiducial pair-point-based with tele-operation

Since BFR has proven to be the gold standard for registration accuracy this concept also uses these fiducials. However, in this concept the localisation takes place through tele-operation as sketched in Figure I.2. In tele-operation a camera will be attached to the robot end-effector and a tool has to be placed inside the fiducials by steering it with a joystick. A force sensor in the tip will provide haptic feedback to the human operator.

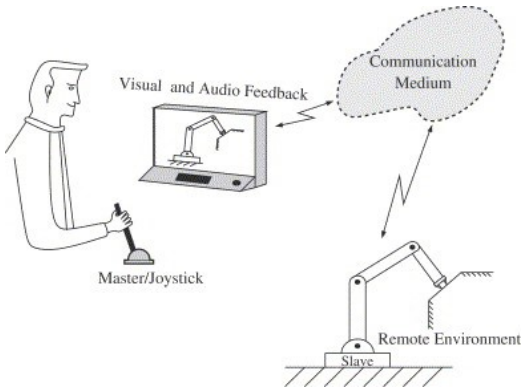


Figure I.2: Example of bone-anchored fiducial registration with tele-operation where the surgeon directly controls the robot with a remote joy stick (2012).

The disadvantages of this concept are:

1. Haptic feedback needed for operator
2. Camera calibration
3. Camera occlusion problems
4. Invasive
5. Human errors

The advantages of this concept are:

1. Distance between operator and robot limits hazardous situations
2. Scalability of movements
3. HITL provides trust
4. HITL provides error detection

The innovativeness and scientific value of this concept lies in the use of clear haptic feedback without delay in image-to-patient registration. Moreover, the scalability could also be changed accordingly in a shared control manner, where the camera is used to provide feedback on distance to patient and adjust the scalability accordingly.

I.1.3. Camera-based bone-anchored fiducial pair-point-based with calibration balls

Another suggested concept is a camera-based fiducial frame with calibration balls. Stereo-vision and marker balls with a fixed distance from each other are used. The balls are attached to a fiducial screw on the patient with one ball in a fixed position straight above that screw. With different orientations of those balls, that are being recorded by the camera, the position of the bottom ball can be determined with high accuracy, which is a fixed distance from the screw.
The disadvantages of this concept are:

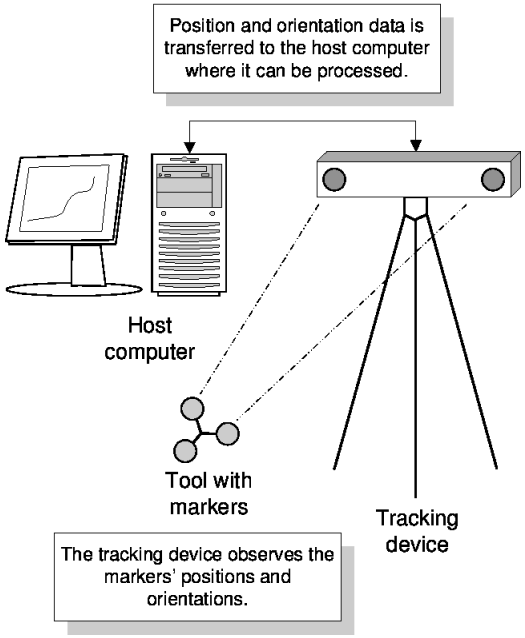


Figure I.3: Example of camera-based fiducial frame with calibration balls. A tracking device tracks markers on the calibration ball with high accuracy (2002).

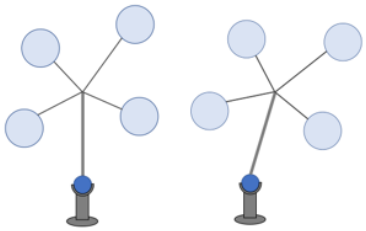


Figure I.4: Example of the marker balls screwed on top of fiducial screws inside the patient.

1. Additional, separate parts
2. Camera calibration
3. Camera occlusion problems
4. Invasive
5. Unknown accuracy, but Feng and Max (2014) mentions an static tracking accuracy of 0.15mm ± 0.11mm

The advantages of this concept are:

1. Minimisation of hazardous situation
2. Automation possible
3. Minimisation of human error

The innovativeness and scientific value in this concept lies in the fact that no non-invasive registration method alike was found in the literature.

I.1.4. Head-fixation frame-based registration through ball in grooves

The last method uses two objects, one fixed to the base of the robot and the other one fixed to the patient's head. These two objects can only fit on top of each other in one way. It is a ball in grooves principle with three different grooves in a triangle. The object on the base plate has to be clicked onto the object on the patient's head, where the only variable is the rotation.

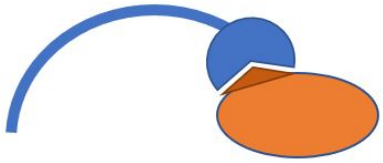


Figure I.5: Example of the ball in grooves prnciple for the head-fixation frame-based registration. The blue part is the object fixed to the robot and the dark orange object is fixed to the patient. They fit perfectly.

The disadvantages of this concept are:

1. Material dependency
2. Not patient friendly
3. Invasive
4. Unknown accuracy, but likely similar to frame-based registration.

The advantages of this concept are:

1. Minimisation of hazardous situation
2. Automation possible
3. No transformation matrix needed
4. Minimisation of human error

The innovativeness and scientific value is in the fact that no method alike has been reported in the literature.

I.2. Scoring

Scoring the concepts along the scoring criteria in Figure 4.3 on a scale from 0 to 5. Each criteria was scaled from 0 to 4 in importance providing an inverse weight to multiple the given score. The scores are thus multiplied by their weight and summed over all four categories of scoring criteria. From four different stakeholders the mean value is calculated according to Equation 2.3, which yields an average score of 221.25, 218, 179.5 and 212.33 for concepts 1 to 4, given in Figure I.6. Concept 1 scores best, whereas concept 3 scores the lowest. The different scores from each stakeholder are provided below in Figures I.7, I.8, I.9, I.10.

	C1	C2	C3	C4
Stakeholder1	218	173	183	178
Stakeholder2	203	251	152	269
Stakeholder3	211	178	158	190
Stakeholder4	253	270	225	0
Avg	221.25	218	179.5	212.333

Figure I.6: Concept scoring summary

Criteria													
Weight	Importance	Technical	Meaning	C1		C2		C3		C4	
4	0	Accuracy	Depends on the errors, this has to be elaborated in a detailed error-analysis	3	12	2	8	3	12	1	4
2	2	Error detection	How easily can the different crucial errors be detected and measured	2	4	3	6	3	6	1	2
1	3	Efficiency	Computational efficiency achieves the result with little to no additional effort or energy	4	4	2	2	2	2	4	4
1	3	Level of automation	How autonomous is the method	2	2	2	2	3	3	2	2
2	2	Robustness	Small pertubations have little effect on the outcome	3	6	2	4	3	6	2	4
2	2	Robot workspace	How the method affects	4	8	4	8	3	6	3	6
2	2	Registration time	Time it takes to finish the registration procedure	4	8	3	6	2	4	3	6
3	1	Repeatability	Results reproduced from the same procedure lie in a small range from each other	3	9	2	6	3	9	3	9
					53		42		48	19	37
		Human factors	Meaning	C1		C2		C3		C4	
3	1	Ease of use	How the user perceives how easily the procedure can be finished	4	12	2	6	3	9	3	9
2	2	Training needed	How fast the user can learn to use the technique	3	6	2	4	2	4	3	6
3	1	Human error	How human errors influence the end-result	2	6	2	6	2	6	2	6
1	3	Influence of operator st	How much the user state can influence the end-result	2	2	2	2	3	3	2	2
3	1	Reproducibility	How easy it is for the user to reproduce the same procedure	2	6	2	6	2	6	3	9
2	2	Adaptation in workflow	How does the procedure influence the current workflow	3	6	2	4	2	4	3	6
2	2	Workload	How focused the user must be to provide good end-result	3	6	2	4	3	6	3	6
					97		74		86	38	81
		Clinical	Meaning	C1		C2		C3		C4	
3	1	Invasiveness	The amount of additional stress for the patient by additional incisions etc.	2	6	2	6	2	6	1	3
1	3	Footprint	The amount of space it takes in the OR	4	4	2	2	2	2	3	3
2	2	Patient-friendly	If the patient has to wear attribute between the pre-operative and intraoperative phases	3	6	3	6	3	6	2	4
					16		14		14		10
		General	Meaning	C1		C2		C3		C4	
1	3	Costs	Additional costs for purchase, production and maintenance	4	4	3	3	2	2	3	3
3	1	Complexity	The number of interacting	3	9	2	6	1	3	4	12
4	0	Safety procedure	The ease of providing a safety procedure that is not relying on a single point of failure	3	12	4	16	3	12	2	8
2	2	Re-registration possible	The ease of re-registration when needed	3	6	3	6	3	6	3	6
1	3	Installation time	The time it takes for the first installation and initialisation	3	3	2	2	2	2	3	3
2	2	Additional parts	How much additional parts are needed	4	8	1	2	1	2	2	4
2	2	IP	Intellectual property, is this method already patented	2	4	2	4	2	4	4	8
2	2	Maintenance	The amount of maintenance needed	3	6	2	4	2	4	3	6
					52		43		35		50
				Sum:	218		173		183		178

Figure I.7: Scores from stakeholder 1

Criteria													
Weight	Importance	Technical	Meaning	C1		C2		C3		C4	
4	0	Accuracy	Depends on the errors, this has to be elaborated in a detailed error-analysis	3	12	3	12	3	12	3	12
2	2	Error detection	How easily can the different crucial errors be detected and measured	2	4	5	10	1	2	2	4
1	3	Efficiency	Computational efficiency achieves the result with little to no additional effort or energy	3	3	5	5	1	1	5	5
1	3	Level of automation	How autonomous is the method	2	2	2	2	2	2	5	5
2	2	Robustness	Small pertubations have little effect on the outcome	2	4	2	4	2	4	2	4
2	2	Robot workspace	How the method affects	3	6	3	6	3	6	4	8
2	2	Registration time	Time it takes to finish the registration procedure	3	6	3	6	2	4	5	10
3	1	Repeatability	Results reproduced from the same procedure lie in a small range from each other	3	9	5	15	2	6	5	15
					46		60		37		63
		Human factors	Meaning	C1		C2		C3		C4	
3	1	Ease of use	How the user perceives how easily the procedure can be finished	3	9	4	12	2	6	3	9
2	2	Training needed	How fast the user can learn to use the technique	4	8	4	8	3	6	4	8
3	1	Human error	How human errors influence the end-result	2	6	4	12	2	6	4	12
1	3	Influence of operator st	How much the user state can influence the end-result	3	3	3	3	3	3	3	3
3	1	Reproducibility	How easy it is for the user to reproduce the same procedure	3	9	5	15	2	6	4	12
2	2	Adaptation in workflow	How does the procedure influence the current workflow	3	6	3	6	2	4	4	8
2	2	Workload	How focused the user must be to provide good end-result	2	4	4	8	2	4	5	10
					91		124		72		125
		Clinical	Meaning	C1		C2		C3		C4	
3	1	Invasiveness	The amount of additional stress for the patient by additional incisions etc.	1	3	1	3	1	3	1	3
1	3	Footprint	The amount of space it takes in the OR	4	4	3	3	2	2	3	3
2	2	Patient-friendly	If the patient has to wear attribute between the pre-operative and intraoperative phases	1	2	1	2	1	2	1	2
					9		8		7		8
		General	Meaning	C1		C2		C3		C4	
1	3	Costs	Additional costs for purchase, production and maintenance	5	5	4	4	2	2	3	3
3	1	Complexity	The number of interacting	4	12	3	9	2	6	5	15
4	0	Safety procedure	The ease of providing a safety procedure that is not relying on a single point of failure	2	8	3	12	2	8	3	12
2	2	Re-registration possible	The ease of re-registration when needed	1	2	4	8	1	2	5	10
1	3	Installation time	The time it takes for the first installation and initialisation	4	4	4	4	2	2	5	5
2	2	Additional parts	How much additional parts are needed	3	6	2	4	1	2	5	10
2	2	IP	Intellectual property, is this method already patented	5	10	5	10	5	10	5	10
2	2	Maintenance	The amount of maintenance needed	5	10	4	8	2	4	4	8
					57		59		36		73
				Sum:	203		251		152		269

Figure I.8: Scores from stakeholder 2

	Criteria									
Weight	Importance	Technical	Meaning	C1		C2		C3		C4	
4	0	Accuracy	Depends on the errors, this has to be elaborated in a detailed error-analysis	3	12	2	8	2	8	2	8
2	2	Error detection	How easily can the different crucial errors be detected and measured	2	4	3	6	2	4	2	4
1	3	Efficiency	Computational efficiency achieves the result with little to no additional effort or energy	3	3	2	2	2	2	3	3
1	3	Level of automation	How autonomous is the method	2	2	3	3	2	2	2	2
2	2	Robustness	Small pertubations have little effect on the outcome	3	6	3	6	2	4	3	6
2	2	Robot workspace	How the method affects	4	8	4	8	3	6	2	4
2	2	Registration time	Time it takes to finish the registration procedure	3	6	3	6	2	4	3	6
3	1	Repeatability	Results reproduced from the same procedure lie in a small range from each other	3	9	3	9	2	6	3	9
					50		48		36		42
		Human factors	Meaning	C1		C2		C3		C4	
3	1	Ease of use	How the user perceives how easily the procedure can be finished	4	12	2	6	3	9	3	9
2	2	Training needed	How fast the user can learn to use the technique	3	6	2	4	2	4	3	6
3	1	Human error	How human errors influence the end-result	2	6	2	6	3	9	2	6
1	3	Influence of operator st	How much the user state can influence the end-result	2	2	2	2	3	3	3	3
3	1	Reproducibility	How easy it is for the user to reproduce the same procedure	3	9	2	6	2	6	3	9
2	2	Adaptation in workflow	How does the procedure influence the current workflow	3	6	2	4	2	4	3	6
2	2	Workload	How focused the user must be to provide good end-result	2	4	1	2	3	6	3	6
					95		78		77		87
		Clinical	Meaning	C1		C2		C3		C4	
3	1	Invasiveness	The amount of additional stress for the patient by additional incisions etc.	1	3	1	3	1	3	3	9
1	3	Footprint	The amount of space it takes in the OR	3	3	2	2	2	2	2	2
2	2	Patient-friendly	If the patient has to wear attribute between the pre-operative and intraoperative phases	2	4	2	4	2	4	3	6
					10		9		9		17
		General	Meaning	C1		C2		C3		C4	
1	3	Costs	Additional costs for purchase, production and maintenance	4	4	2	2	2	2	3	3
3	1	Complexity	The number of interacting	4	12	3	9	2	6	2	6
4	0	Safety procedure	The ease of providing a safety procedure that is not relying on a single point of failure	3	12	3	12	2	8	2	8
2	2	Re-registration possible	The ease of re-registration when needed	3	6	3	6	3	6	3	6
1	3	Installation time	The time it takes for the first installation and initialisation	2	2	2	2	2	2	3	3
2	2	Additional parts	How much additional parts are needed	4	8	2	4	2	4	2	4
2	2	IP	Intellectual property, is this method already patented	3	6	2	4	2	4	4	8
2	2	Maintenance	The amount of maintenance needed	3	6	2	4	2	4	3	6
					56		43		36		44
				Sum	211		178		158		190

Figure I.9: Scores from stakeholder 3

	Criteria									
Weight	Importa	Technical	Meaning	C1		C2		C3		C4	
4	0	Accuracy	Depends on the errors, this has to be elaborated in a detailed error-analysis	4	16	5	20	3	12		0
2	2	Error detection	How easily can the different crucial errors be detected and measured	5	10	5	10	4	8		0
1	3	Efficiency	Computational efficiency achieves the result with little to no additional effort or energy	4	4	4	4	4	4		0
1	3	Level of automation	How autonomous is the method	2	2	2	2	3	3		0
2	2	Robustness	Small pertubations have little effect on the outcome	4	8	5	10	3	6		0
2	2	Robot workspace	How the method affects	3	6	3	6	3	6		0
2	2	Registration time	Time it takes to finish the registration procedure	2	4	2	4	3	6		0
3	1	Repeatability	Results reproduced from the same procedure lie in a small range from each other	4	12	5	15	3	9		0
					62		71		54		0
		Human factors	Meaning	C1		C2		C3		C4	
3	1	Ease of use	How the user perceives how easily the procedure can be finished	2	6	2	6	3	9		0
2	2	Training needed	How fast the user can learn to use the technique	3	6	3	6	2	4		0
3	1	Human error	How human errors influence the end-result	3	9	3	9	4	12		0
1	3	Influence of operator st	How much the user state can influence the end-result	3	3	3	3	4	4		0
3	1	Reproducibility	How easy it is for the user to reproduce the same procedure	3	9	3	9	4	12		0
2	2	Adaptation in workflow	How does the procedure influence the current workflow	4	8	2	4	1	2		0
2	2	Workload	How focused the user must be to provide good end-result	3	6	3	6	3	6		0
					109		114		103		0
		Clinical	Meaning	C1		C2		C3		C4	
3	1	Invasiveness	The amount of additional stress for the patient by additional incisions etc.	4	12	4	12	2	6		0
1	3	Footprint	The amount of space it takes in the OR	2	2	2	2	4	4		0
2	2	Patient-friendly	If the patient has to wear attribute between the pre-operative and intraoperative phases	4	8	4	8	2	4		0
					22		22		14		0
		General	Meaning	C1		C2		C3		C4	
1	3	Costs	Additional costs for purchase, production and maintenance	4	4	5	5	3	3		0
3	1	Complexity	The number of interacting	4	12	3	9	2	6		0
4	0	Safety procedure	The ease of providing a safety procedure that is not relying on a single point of failure	2	8	4	16	5	20		0
2	2	Re-registration possible	The ease of re-registration when needed	5	10	5	10	2	4		0
1	3	Installation time	The time it takes for the first installation and initialisation	4	4	3	3	3	3		0
2	2	Additional parts	How much additional parts are needed	4	8	3	6	2	4		0
2	2	IP	Intellectual property, is this method already patented	4	8	4	8	4	8		0
2	2	Maintenance	The amount of maintenance needed	3	6	3	6	3	6		0
					60		63		54		0
				Sum	253		270		225		0

Figure I.10: Scores from stakeholder 4

I.3. Conclusion

Robots have weaknesses and strengths that are complementary to human strengths and weaknesses, and automation is known to lead to additional problems such as overreliance or under reliance on a system (Crouser et al., 2013; Fitts, 1951). Since the medical field requires minimisation of errors and failures, a human-in-the-loop (HITL) method is considered best. Then the skills of both the robot and the human can be combined accordingly. This also mitigates problems related to liability. This also follows from the concept scoring.

The concept scoring uncovers touch-based bone-anchored fiducial pair-point-based registration with admittance control as the best method that satisfies the scoring and exclusion criteria related to the contextual background with an HRI focus. Admittance control will take place by translating the forces and torques applied by the human operator to the velocity movements of the robot. In this design, the human operator and the robot will collaboratively execute the localisation task.

J. Registration selection

This Appendix provides information on the selection procedure in this thesis to select the best specific registration methods.

This chapter starts with details about the general selection leading to Figure 4.4. Then the details about the specific selection leading to Figure 4.5 are given.

J.1. General registration selection

The first four exclusion criteria are applied to the general methods to select the best general registration method for the application. Figure 4.4 summarises the scoring results and concludes invasive fiducial screw markers the best method with no criteria failed. Red and green indicate the exclusion and inclusion of the method based on the criterion, respectively. Orange denotes a criterion that is neither met nor failed, thus not including nor including the method. This research excludes voxel-based and deformable model surface-based registration methods. It was found that non-invasive extrinsic registration methods scored the worst by failing two or more criteria with an exception for dental adapters.

J.1.1. Extrinsic invasive registration methods

Overall, the extrinsic invasive registration methods scored better than the non-invasive ones.

Stereotactic frame

The research excludes the stereotactic frame-based registration method

Sub-millimetric accuracy

Following the STOTA methods accuracy, this research finds that the TRE varies between 0.7 ± 0.5mm and 1.5 ± 0.8mm. Thus the TRE exceeds 1.0mm; therefore, failing this criterion.

No limit to robot workspace

As frame-based registration mostly covers large parts of the head, this limits the robot workspace. Therefore excluding the method based on this criterion.

High robustness

As frame-based registration includes a frame fixated on the patient's head. Research finds this method fairly robust. Alam et al. (2016) indicates this robustness in Figure 3.7. Therefore including the method based on this criterion.

Low system complexity

Frame-based registration incorporated one central part attached to the operating table and fixated on the patient's head. The operator does not need to perform additional events or acquire additional technical knowledge. Therefore, meeting this criterion.

Fiducial screw markers

This research concludes that the fiducial screw markers-based registration method is the best general method for this project since it failed no criterion, met three criteria, and neither failed nor met one criterion.

Sub-millimetric accuracy

According to STOTA methods, fiducial screw markers have promising results with TRE values smaller than 1.0mm or in the range of 1.0mm. Alam et al. (2016) underlines this and multiple works of literature refer to this method as the golden standard for accuracy, thus satisfying this criterion.

No limit to robot workspace

Fiducial screws attached to the patient's head must be placed outside the surgical area and, therefore, will not obstruct the robot workspace, fulfilling this criterion.

High robustness
Attaching fiducials to human bone limits perturbation occurrence. Using multiple screws allows for easy mitigation of possible damages. Thus considering this a robust method, even according to the evaluation of Alam et al. (2016). Thus this criterion is met.

Low system complexity
As this registration method requires matching pair points, these must be localised, possibly delivering system complexity. It likewise requires high skills, according to Alam et al. (2016). Regardless, it does not require any complicated optimisation and does not add any additional parts to the current system. Since the complexity of this method highly depends on the specific registration method, the criterion is neither met nor failed.

J.1.2. Extrinsic non-invasive registration methods

The extrinsic non-invasive registration methods scored worst compared to all other general methods.

Mould, frame and dental adapters
The analysis separates the three types of non-invasive registration.

Sub-millimetric accuracy
From the STOTA and the evaluation of Alam et al. (2016), it follows that the accuracy of moulds and frames is larger than 1.0mm, therefore, failing this criterion. For dental adapters, some results are promising and surround the 1.0mm accuracy. J. Wang et al. (2020) mentions an average marker displacement error of 0.03mm with an angular deviation of 0.07 degrees. The STOTA literature on the neuroclate reaches a TRE value of 0.67 ± 0.29mm. Therefore dental adapters neither meet nor fail this criterion, whereas non-invasive moulds and frames fail.

No limit to robot workspace
Non-invasive registration methods, most of the time, are removed during the surgery. Therefore they will not pose any limit to the robot workspace. Nevertheless, non-invasive frames limiting the robot workspace can occur. Therefore non-invasive moulds and dental adapters meet this criterion, whereas non-invasive frames neither fail nor meet this criterion.

High robustness
All three have low robustness; they are all created in the pre-operative phase and re-applied in the intraoperative phase. However, the patient's anatomy can change in this time frame, leading to large perturbations that can influence the outcome. Thus all methods are excluded based on this criterion.

Low system complexity
Moulds are highly complex as they have to be personalised, which requires additional skills and resources, thus failing this criterion for non-invasive moulds. Dental adapters add more parts to the system, and their placement requires additional skill. Nevertheless, they do not need additional technical optimisation, knowledge or other add-ons to the current robot, thus neither meeting nor failing this criterion. Non-invasive frames meet this criterion as their complexity is similar to invasive frame-based registration.

Fiducial skin markers
This research excludes fiducial skin marker registration as it fails two criteria, meets one criterion, and neither satisfies nor fails one criterion.

Sub-millimetric accuracy
The STOTA literature finds TRE values in the 2.0mm or higher range. Alam and Rahman (2016) underlines this, thus failing this criterion.

No limit to robot workspace

Fiducial skin markers are similar to fiducial screw markers; they do not limit the robot workspace, thus meeting this criterion.

High robustness
This research considers this method not robust since skin markers can move due to skin elasticity. Alam and Rahman (2016) stresses this, thus failing this criterion.

Low system complexity
Just like screw markers, the complexity differs per specific method. In general, it requires high skills and careful placement but does not require additional robot parts, thus neither meeting nor failing this criterion.

J.1.3. Intrinsic point-based registration methods

This research uncovers that these methods only fail the accuracy criterion.

Anatomical
This research excludes the anatomical intrinsic registration method as it fails one criterion, meets two criteria, and neither meets nor fails one criterion.

Sub-millimetric accuracy
The STOTA concludes TRE values ranging from 3.0 to 5.0mm. Alam and Rahman (2016) underlines this, thus failing this criterion.

No limit to robot workspace
Since anatomical structures can be chosen, it will not limit the robot workspace, thus meeting this criterion. However, not choosing the optimal placements can affect the accuracy significantly.

High robustness
Alam and Rahman (2016) evaluates the robustness as high. However, this highly depends on the specific type and the chosen landmark positions., thus neither failing nor meeting this criterion.

Low system complexity
Just like screw markers, the complexity differs per specific method. In general, it requires high skills and careful choosing of landmarks. Even though Alam and Rahman (2016) notes the demand for multiple landmarks, which will increase the complexity a little, the method still meets this criterion.

Geometrical
This research excludes the geometrical intrinsic registration method as it fails one criterion, meets one criterion, and neither meets nor fails two criteria.

Sub-millimetric accuracy
The STOTA finds TRE values ranging from 0.9 to 5.5mm. Alam and Rahman (2016) mentions this too, thus failing this criterion.

No limit to robot workspace
This method provokes no limits to the robot workspace, thus meeting the criterion.

High robustness
According to Alam and Rahman (2016), the robustness is high. However, this highly depends on the specific type, the chosen landmark positions and the number of landmarks; thus, neither failing nor meeting this criterion.

Low system complexity
The complexity significantly differs per specific method. In general, acquiring geometrical landmarks requires more skills and knowledge of geometrical structures, thus neither meeting nor failing this criterion.

J.1.4. Intrinsic surface-based registration methods

This research finds that intrinsic surface-based registration methods only fail the accuracy criterion.

Rigid models
This research excludes the rigid model's intrinsic registration method as it fails one criterion, meets two criteria, and neither fails nor meets one criterion.

Sub-millimetric accuracy
The STOTA concludes TRE values ranging from 1.0 to 1.6mm. Alam and Rahman (2016) states that accuracy is only possible if the pre-segmentation setup is performed precisely, thus failing this criterion.

No limit to robot workspace
This method does not limit the robot workspace and thus meets this criterion.

High robustness
Alam and Rahman (2016) considers this method robust. It is robust as no points are the same, and the number of gathered data points is large, thus meeting this criterion.

Low system complexity
The complexity significantly differs per specific method. In general, acquiring surface landmarks requires more skills, knowledge and time, thus neither meeting nor failing this criterion.

J.2. Specific registration selection

The specific bone-anchored pair-point-based fiducial registration methods are summarised in Figure J.1. Touch-based is performed by physically touching the fiducials through an exterior probe or with the robot end-effector. Camera-based uses a camera to determine the fiducial locations. This is also the case for active tracking, but here the locations are tracked in real-time. Intraoperative utilises an imaging modalitity to determine fiducial locations during surgery. The electromagnetic method uses electromagnetic forces to localise the fiducials. Lastly, the head-fixation frame uses two objects, one fixed to the base of the robot and the other one fixed to the patient's head. These two objects can only fit on top of each other in one way.

Scoring each method led to the results shown in Figure 4.5, concluding the three best methods to be touch-based, camera-based and head-fixation frame since they do not fail any criteria. Red and green indicate the exclusion and inclusion of the method based on the criterion, respectively. Orange denotes a criterion that is neither met nor failed, thus not including nor including the method.

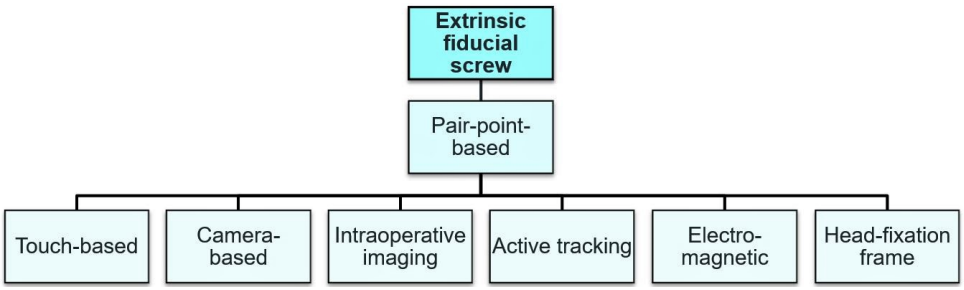


Figure J.1: The defined specific registration methods for one-anchored pair-point-based fiducial registration.

J.2.1. Touch-based

Touch-based bone-anchored pair-point-based fiducial registration can be performed through an exterior touch-based probe or with the robot end-effector. Touch-based registration can be executed manually with a human operator or automatically. This research concludes that this method is the best as it fails no criterion, fails eight criteria, and neither fails nor meets two criteria.

Sub-millimetric accuracy
The STOTA shows that touch-based methods, such as the paper of Gerber et al. (2013), score well in

accuracy, thus meeting this criterion.

No limit to robot workspace
Fiducials do not limit the robot workspace and can be placed accordingly, satisfying this criterion.

High robustness
Attaching fiducials to human bone limits perturbation occurrence. Using multiple screws allows for easy mitigation of possible damages. Manual touching involves a human operator, which mitigates perturbations very well (Crouser et al., 2013; Fitts, 1951). Thus meeting this criterion.

Low system complexity
The system complexity depends on the design of the fiducials and touching probe. However, localising an object by touching it is a natural human skill. It does not require any complicated technical systems, optimisation or parts to the current system, thus meeting this criterion (Crouser et al., 2013).

Low system maintenance

The system maintenance depends on the stability of the closed-loop system involved in localising the fiducials. Final design implementations influence this stability and differ per design. Nevertheless, this design is generally not demanding large amounts of parts. Thus, this criterion is neither met nor failed.

High patient safety
Attaching fiducials to the patient and touching them can harm the patient. Automated localisation can be dangerous, but manual localisation is less dangerous as operators can prevent damage to the patient, thus neither meeting nor failing this criterion.

Limited training
The method requires little training for manually performing localisation since humans can intuitively perform this task. The automated versions can be programmed accordingly and also requires low training, hence meeting this criterion.

No IP limits
This research finds no IP limits for this specific method, thus meeting this criterion.

Only CT imaging modality
This technique can be performed with all types of imaging, hence meeting this criterion.

Restricted development time
The EMR robot is already equipped with a touch-based method, meeting this criterion.

J.2.2. Camera-based

Camera-based bone-anchored pair-point-based fiducial registration can be performed by fiducial object detection through camera sensing and recognition, directly finding the fiducials or finding an object placed on top of the fiducials. This research concludes that this method is one of the three best methods as it failed no criterion, meets five criteria, and neither meets nor fails five criteria.

Sub-millimetric accuracy
High-accurate situations increasingly use cameras. For instance, Feng and Max (2014) mentions an static tracking accuracy of $0.15\text{mm} \pm 0.11\text{mm}$. This can differ per distance to the tracking object, but will meet the criterion (A. Wang and Gollakota, 2019).

No limit to robot workspace
Fiducials do not limit the robot workspace and can be placed accordingly, meeting this criterion.

High robustness
Attaching fiducials to human bone limits perturbation occurrence. Using multiple screws allows for easy mitigation of possible damages. Nonetheless, cameras require re-calibration after perturbations and are not very robust; hence they neither fail nor meet this criterion.

Low system complexity

Camera-based methods require additional parts, calibration and setup instructions; this increases the system's complexity. However, the criterion is not completely failed since camera-based methods are quickly improving and becoming easier to use. Thus this criterion is neither met nor failed.

Low system maintenance
Camera-based methods require additional maintenance as cameras are easily damaged. Nevertheless, they do not require very high maintenance and become increasingly easy to use and maintain; consequently, they neither fail nor meet this criterion.

High patient safety
The camera recognises fiducials at a distance, requiring no direct contact. However, wrongly calibrating can lead to dangerous situations, thus neither meeting nor failing this criterion.

Limited training
The training required to perform camera-based localisation is neither low nor high, as cameras can perform the registration automatically. Regardless, camera calibration and usage require additional training; accordingly, neither meeting nor failing this criterion.

No IP limits
This research finds no IP limits for this specific method, thus meeting this criterion.

Only CT imaging modality
This method can be performed with all types of imaging, hence meeting this criterion.

Restricted development time
Adding a camera to the system does not require much development time, thus meeting this criterion.

J.2.3. Intraoperative imaging

Intraoperative imaging uses an additional CT or another imaging device inside the OR to localise the fiducials and perform registration. This research excludes this method as it fails four criteria, meets two criteria, and neither fails nor meets four criteria.

Sub-millimetric accuracy
Intraoperative imaging accuracy has a wide range from 0.3 - 2.2mm for CT scans as the difference between the CT display and actual anatomic location in the operative field (Zinreich et al., 1993). Thus this criterion is neither met nor failed.

No limit to robot workspace
The localisation takes place from a distance, just like a camera-based method hence, meeting this criterion. *High robustness*

Attaching fiducials to human bone limits perturbation occurrence. Using multiple screws allows for easy mitigation of possible damages. Intraoperative imaging is fairly robust since imaging modalities are robust. However, if the modality is damaged, additional resources are demanded to re-calibrate, consequently, neither failing nor meeting this criterion.

Low system complexity
Additional imaging in the OR adds a lot of complexity, such as the need for an additional fairly complex system, therefore failing this criterion.

Low system maintenance
Additional imaging in the OR requires additional maintenance as it is a complex system that needs to be maintained. However, the workflow already uses imaging modalities that require maintenance. Thus this criterion is considered neither met nor failed.

High patient safety
Attaching the fiducials to the patient and localising these at a distance demands no direct contact. Nonetheless, the additional imaging exposes the patients to extra radiation, which can cause damage to the patient, thus neither meeting nor failing this criterion.

Limited training

This method requires additional training in intraoperative imaging system usage in registration, failing this criterion.

No IP limits
This research finds no IP limits for this specific method, thus meeting this criterion.

Only CT imaging modality
This intraoperative imaging method adds imaging modality for which the system is not optimised and demands hospitals to acquire this modality, which is not the case in most hospitals, hence failing this criterion(Mitchell and Labadie, 2020).

Restricted development time
As the researcher and companies involved in this project do not possess an intraoperative imaging device, executing this method will require significant development time, consequently failing this criterion.

J.2.4. Active tracking

Active tracking is a system that performs localisation of the fiducial markers in real-time with cameras. This research excludes active tracking from further study as it fails three criteria, meets three criteria, and neither fails nor meets four criteria.

Sub-millimetric accuracy
Active tracking can acquire a high accuracy, ranging from 0.010 - 0.050mm, as shown by Gerber et al. (2013), thus meeting this criterion.

No limit to robot workspace
Active tracking requires occlusion prevention, as this can lead to dangerous situations. This might limit the workspace, therefore neither meeting nor failing this criterion.

High robustness
Attaching fiducials to human bone limits perturbation occurrence. Using multiple screws allows for easy mitigation of possible damages. Just like the camera-based method, active tracking is not very robust and requires re-calibration, thus neither meeting nor failing this criterion.

Low system complexity
Active tracking adds a lot of complexity, such as the need for an additional complex system and requires additional tracking systems and software. It can have delay issues, high acquisition costs, and OR workflow disruption, thus failing this criterion (Übelhör et al., 2020; Wong et al., 2014).

Low system maintenance
An active tracking system requires much additional maintenance as it needs to be calibrated and optimised to prevent errors and dangerous situations from occurring, failing this criterion.

High patient safety
This method requires no direct patient contact. Nevertheless, the miscalibration of cameras can lead to dangerous situations, consequently neither meeting nor failing this criterion.

Limited training
Same as for camera-based registration, the training required to perform camera-based localisation is neither low nor high, therefore neither meeting nor failing this criterion.

No IP limits
This research finds no IP limits for this specific method, thus meeting this criterion.

Only CT imaging modality
Active tracking can be performed with CT modality, meeting this criterion.

Restricted development time
An active tracking system increases the number of parts added to the robot, hence requiring large amounts of development time and thus failing this criterion.

J.2.5. Electromagnetic

This method uses an electromagnetic way to localise the fiducials, for instance, by magnetically finding and attaching the end-effector to them. This is not a known method. This research excludes this method as it fails two criteria, meets three criteria, neither fails nor meets two criteria and entails three unknown criteria.

Sub-millimetric accuracy
As no such method is known, the accuracy is unknown.

No limit to robot workspace
As this method uses electromagnetic forces, there is no need for additional parts that can block the robot workspace, meeting this criterion.

High robustness
As no method is known, it is unclear what the method's robustness would be.

Low system complexity
As the method includes different technical phenomena, namely electromagnetic forces, this would greatly add to the complexity, thus failing this criterion.

Low system maintenance
The amount of maintenance is unknown for this method.

High patient safety
The method includes electromagnetic forces, which could harm the robotic system or interfere with other equipment in the OR. If an electric circuit is closed through the patient, this can be very dangerous, hence failing this criterion.

Limited training
Using electromagnetic forces to localise the fiducials, humans do not need much effort to execute registration. Regardless, they need additional training on electromagnetic forces and system usage; accordingly, they neither fail nor meet this criterion.

No IP limits
This research finds no IP limits for this specific method, thus meeting this criterion.

Only CT imaging modality
This method would be useable with CT imaging modality, thus meeting this criterion.

Restricted development time
This method demands development from scratch and many additional parts, implying significant development time; nonetheless, electromagnetic forces are well known. Hence, this criterion is neither met nor failed.

J.2.6. Head-fixation frame

This method has a fixed object on the patient's head that can be attached to a separate frame attached to the robot. This mechanism would click fit, directly localising the fiducials of the fixed object. This is not a known method. This research considers this method one of the best methods further investigated, as it fails no criteria, meets six criteria, and neither meets nor fails four criteria.

Sub-millimetric accuracy
As fiducial localisation has a high accuracy, it should correlate to this method. Nevertheless, frame-based registration retains lower accuracy. It is unsure how that affects the accuracy of this method, but it is expected to yield accuracy, meeting this criterion.

No limit to robot workspace
Attaching a frame could obstruct the robot workspace, but it depends on the final implementations, therefore neither meeting nor failing this criterion.

High robustness

As the objects will fit perfectly and are fixated, it is anticipated to be very robust, thus satisfying this criterion.

Low system complexity
Attaching an additional frame to the robot slightly complicates the system, just as the need to develop an object for attachment to the patient's head. Once the frame and object are developed, the system will contain low complexity. Therefore neither meeting nor failing this criterion.

Low system maintenance
No additional techniques are added to the system; it is expected to have low system maintenance, thus meeting this criterion.

High patient safety
This method requires no direct patient contact, additional forces or radiation. Thus providing a safe system for the patient and meeting this criterion.

Limited training
This method requires limited training, including positioning the object to the patient's head and clicking it to the frame, meeting this criterion.

No IP limits
This research finds no IP limits for this specific method, thus satisfying this criterion.

Only CT imaging modality
This method would be useable with CT imaging modality, consequently meeting this criterion.

Restricted development time
This system requires much research into the best objects to fix to the patient's head and the addition of a robot frame, thus expecting considerable development time, neither meeting nor failing this criterion.

K. Mitigation strategies

This Appendix provides the drafted mitigation strategies related to the robot localisation operator state.

This chapter provides all the mitigation strategies drafted and scored leading to Figure 5.1.

K.1. Mitigation strategies

Mitigation strategies were conducted and scored from the design requirements and failures associated with the robot localisation operator state, as highlighted in Figure 5.6. Numerous mitigation strategies were drafted and scored according to their advantages and risk. Their advantages and risks are also related to other errors, such as the robot state. Based on which the decision was made to either implement or not implement a mitigation strategy. When it is decided to be implemented, it is described what steps will be implemented. In the case of rejection to implement, the reasoning is summarised.

Table K.1: Mitigation strategies: omitted step in procedure

Strategy	Description	Advantage	Risk	Incl	Functionalities/reasons
GUI with step-by-step checklist	Make a GUI with a clear checklist so that the operator is more engaged and skipping steps becomes more difficult.	The operator is required to go through the checklist and thus responsibility lies at the operator. It adds to the robustness, repeatability and safety. Moreover, it lowers the training needed and human error.	It adds additional time, workload, parts and maintenance (if checklist needs updating). It could also decrease the ease of use.	Yes	1. When the registration is started make a step-by-step procedure for registration of four fiducials 2. Add a checklist that should be filled in by the operator
Track operator time	Save the time that the operator uses to perform registration. This data can be saved and later compared to give key performance indicators of different operators.	The tracker can be used to compare different operators and analyse when mistakes occur if this is linked to the time spend. Moreover, it forces the operator to take more time, knowing it is being tracked. It increases error detection, robustness and safety procedure. It will not affect the maintenance or additional parts since it is very easy to implement.	It could lower efficiency and it could increase the registration time and workload.	Yes	1. Timer starts when registration process is initiated 2. Timer ends when registration process is ended 3. Save time to raw data registration
Add supervisor checklist	Make an additional GUI or app for the supervisor to sign off on the same or another checklist.	When both supervisor and operator have to perform a checklist changes of omitting a step become smaller. Moreover, both operator and supervisor will feel more responsible in filling in the checklist correctly since it has to overlap between both. It will increase robustness, safety procedure and error detection.	It can increase registration time, workload, maintenance if checklist has to be updated, training needed. Moreover, it could make the adoption in workflow more complicated and might affect the ease of use. Moreover, it means the addition of another part	No	This will not be added to the GUI or registration process since it will take more time and clutters the GUI which can lead to more distractions. Therefore, only one operator will perform the checklist in the GUI. However, it is advised to perform registration with an additional supervisor who can check if the full checklist is filled incorrectly.

Table K.2: Mitigation strategies: incorrect use of admittance control

Strategy	Description	Advantage	Risk	Incl	Functionalities/reasons
Better tuning of admittance control	The admittance control should be tuned with respect to its high / low-pass filters and cut-off values such that the operator feels that the use is not requiring too much force but small tremors or noise is still filtered out. This means adjusting the gains and might adjusting the type of filter	Make the adittance control pleasant increases the ease of use, reduces the workload, training needed and does not lead to any additional parts.	The risk of better tuning is very small, however it should be considered that tuning will always be user specific and therefore tuning it in perfection is not possible since every operator might prefer other tuning parameters.	Yes	1. Provide tuning parameters such that amount of force needed is limited and response to lower forces is available 2. Low cut-off value 3. Better filtering parameters 4. Adjusting of filter type
Track if admittance control is used	A functionality where the use of admittance control is tracked. Thus when admittance control is turned on, this will be recorded to the raw data.	This ability to track the use of admittance control can increase the safety procedure.	No risks are related to this tracking.	No	The problem of not using admittance control can be mitigated more easily by limiting the use of positional control during registration. Moreover, the time of registration procedure is already tracked, therefore this functionality does not any additional value any more. Another additional part or function without the addition of actual functionality is not recommended.
Surgery only possible if admittance control is used	Only allow surgery when admittance control is used during registration. Thus tracking of admittance control use is needed for this functionality. Then if admittance control is not used during the registration procedure, the robot is not allowed to perform its surgery.	The addition of this functionality will increase the safety procedures.	This functionality can limit the response of the operator to unknown errors and might hinder the adaption in workflow as well as ease of use.	No	The implementation of this functionality can have unforeseen implications and limiting positional control during registration already mitigates the risk of not using admittance control. Another additional part or function without the addition of actual functionality is not recommended.
Limit use of positional control during registration	During registration procedure,limit the use of positional control in the left-side of the GUI. Thus the position of the end-effector of the robot can only be changed by admittance control of the operator with the robot.	This limitation of positional control forces the operator to use admittance control and it increases the safety procedure.	The limitation of positional control could limit the ability to respond to unknown situations.	Yes	1. Limit the use of positional control during registration mode by blurring out that part of the GUI during registration.

Table K.3: Mitigation strategies: failed to prioritise attention

Strategy	Description	Advantage	Risk	Incl	Functionalities/reasons
Clear and constant communication	A communicative GUI that lets the operator know what is happening and what the implications of its actions are to the process.	The GUI can increase ease of use, efficiency, keep the operator more engaged, limit the training needed.	It could increase the registration time, workload and maintenance if the registration procedure changes in time.	Yes	1. GUI that keeps operator engaged 2. Process bar
Average registration	Instead of performing registration on each fiducial marker only one time, let the operator repeat multiple measurements for each fiducial so that the average position can be taken over all measurements.	The averaging can increase the accuracy, error detection, robustness and limit the human error and training needed.	The averaging can lead to additional registration times, workload, efficiency and might affect the ease of use.	Yes	1. Provide an option for registering multiple times 2. Do not allow registering one point two times in a row without movement 3. Find clustering data points 4. Averaging of data points in each cluster
Find outliers	In performing multiple registration points for each fiducial marker, find the measurements that lie outside of a perceived range. These outliers could be removed from the measurements or repeated.	This could increase the accuracy, robustness, error detection, safety procedures and limit the human error and training needed	This could lead to additional registration times, workload, efficiency and might affect the ease of use.	Yes	1. Separate outliers with a certain threshold 2. Make outlier threshold adaptable 3. Ask operator to repeat for outliers 4. Re-calculate for added data points
Limit buttons	When performing registration, limit the amount of buttons and controls that the operator has to use.	Limiting the amount of controls can reduce the operator workload and thus give the operator more resources to focus on the registration task at hand. Moreover, it reduces the additional parts to a minimum and increases the ease of use and efficiency. It limits the costs and maintenance.	When multiple functionalities are performed by the same button or control, it requires more training and can affect the ease of use maintenance as well as system robustness.	Yes	1. Provide minimum amount of buttons or controls possible
Limit operator actions	When performing registration,limit the amount of actions the operator has to perform to a minimum. These are actions such as mouse-clicks, positional changes, hand movements, upper body movements, lower body movements, talking, etc.	This reduces the addition of workload to a minimum and increases the system efficiency. It should decrease the human error, training needed and registration time. It should affect the operators ease of use.	If not performed correctly, the operator could be less engaged.	Yes	1. Limit the amount of actions needed for the operator to a minimum
Limit operator movement	During the registration process limit the movements that the operator needs to perform. These movements include movements of their arms, hands, upper body most importantly but also their lower body and head movements.	As registration is a high-accuracy task, the limitation of movements will increase the accuracy and limits the human error. Moreover, it will reduce the workload and increase the efficiency.	The limitation of operator movements could decrease their attention and can affect the training needed.	Yes	1. Provide buttons near operator 2. Keep operator focused on registration point when pressing buttons 3. Prevent movement of arms when registration is performed
Use eye-tracking	Tracking the eye movement ofthe operator to ensure its focus on the registration. When the operator is gazing away from the system for a prolonged time, it could provide warnings to the operator.	The use of eye tracking could provide more information about the operators KPIs and prevent attention failures to have large influence on the system. It could increase the safety procedure, error detection and robustness.	The use of eye tracking can lead to additional costs, maintenance, training times, workload, parts.Moreover it could lower the efficiency and ease of use. When the system fails it could lead to annoying warnings that distract the operator more.	No	The implementation of an eye-tracking system is costly, takes a lot of time and can interfere with the users experience without adding significant value to the accuracy.

Table K.4: Mitigation strategies: incorrect use of workspace

Strategy	Description	Advantage	Risk	Incl	Functionalities/reasons
Workspace boundary feedback	The distance between the end-effector location and the workspace boundaries can be shown to the operator and warnings can be given when this distance becomes within a certain threshold.	The addition of workspace boundary feedback could give the operator more knowledge on why robot errors occur and prevent these from happening. Moreover, it could limit potential robot damage. It increases the safety procedure and robustness.	If the workspace of the patient is near a boundary in a certain situation, it could lead to additional distraction for the operator. In addition the GUI can become to cluttered and keep the operator from focusing on what is truly important, namely the registration.It also leads to the increase of maintenance, and training needed.	No	The implementation of workspace boundary feedback could clutter the GUI and distract the operator unnecessary.
Limit workspace	When performing registration limit the workspace of the robot to its boundaries so that no damage to the system can occur.	This limitation increases the safety procedure and limits the occurrence of damage to the system and thus additional cost and maintenance	When the workspace is taken very small, the adaption to certain exceptional situations can be difficult for the operator. In addition, it can be annoying for the operator when the robot keeps stopping near a certain point in space. It can add to the registration time and human training.	Yes	1. Limit movement of the robot outside expected workspace with some margin 2. If robot hits the limiting movement vectors,stop movement 3. Provide error warning: “Outside robot workspace”to the operator

Table K.5: Mitigation strategies: poor localisation technique - 1

Strategy	Description	Advantage	Risk	Incl	Functionalities/reasons
Operator training	Provide an additional GUI or option within the GUI that can be used to train an operator based on different situations.For example first the operator is trained to perform registration normally, then an error occurs and it is explained to the operator what the error means and how to respond. In this training the most likely not standard situations can be trained.	An advantage of having the training as an additional option in the GUI is that it is prevented that bad training by the supervisor or management leads to errors since the operator can still get a quality training. It increases the robustness or the system, decreases the additional training needed, and provides an additional safety procedure.	The additional GUI comes with an additional set in costs, maintenance and can clutter the GUI.	No	Due to time limitations this will not be implemented in the design. However, a protocol will be made totrain different operators in using the registration procedure and thus this information can be used when a training GUI will be made in the future.
Feedback on operator performance	After the first round of registration, find the localisation errors in different x, y, and z directions and then show the operator its error so that it can be improved in time.	It could give the operator more knowledge and insight in its performance and thus improve over time, it could also measure operators KPI and compare with other operators. This could increase error detection, accuracy, robustness. It could also lower human error and training needed since part of the training takes place while performing registration.	The real time feedback could clutter the GUI and distract the operator from performing its task. Thus it can increase workload and decrease ease of use.	No	The implementation of real-time feedback is time consuming and will add to more distractions on the human operator. However, raw data of the registration will be saved so that this feedback can still be given to the operator after registration takes place.
Feedback on localisation technique	During registration keep track of the localisation technique used such as how fast the operator is moving the robot arm and in what direction. How much force is applied and for how long. Then give feedback to the operator when a technique is used that is unpreferable.	An advantage of this feedback is that the localisation technique of different operators can be compared and the best localisation technique can be filtered out. This could make the system more robust and increase the accuracy, reproducibility, lower the human error and training needed	Even though some localisation techniques seem weird, it could be that the operator still prefers that technique and gets good results with it. Therefore limiting the operator to a certain localisation technique can reduce the ease of use and increase the workload and might even the registration time. Moreover the additional feedbackcan clutter the GUI and add to the distractions of the operator to its main task.	No	Due to the time limit and the many risks with respect to this functionality, it will not be implemented. However raw data such as position, speed, time, force, rotation, etc. will be saved so that the localisation technique can still be reviewed after registration and compared with other operators.
Provide multi modal reaching strategy	It was found from the literature that humans use different strategies depending on the size of the target and the distance to target. Moreover, the lack of rotational control could limit the human in its ability to respond to unexpected situations. The current design only has one mode of control with one speed and one angle. Multi-modal means that different modes like fast, slow, rotation can be used by the human to adapt its strategy. The modes and their advantages, risk and implementation will be discussed in more detail later.	The usage of different modes gives the operators more ability to adjust its strategy where needed. This could prevent large mistakes to lead to large errors. Moreover, it could make it easier to use the admittance control and can lead to better accuracy, efficiency and ease of use.	If the modal control is complicated as perceived by the operator it could increase its workload and training needed. In addition, it requires more maintenance.	**Yes**	1. Rotational control 2. Slow control 3. Fast control 4. Switching ability between the modes
Use averaging	Instead of performing registration on each fiducial marker only one time, let the operator repeat multiple measurements for each fiducial so that the average position can betaken over all measurements.	The averaging can increase the accuracy, error detection, robustness and limit the human error and training needed.	The averaging can lead to additional registration times, workload, efficiency and might affect the ease of use.	**Yes**	1. Provide an option for registering multiple times 2. Do not allow registering one point two times in a row without movement 3. Find clustering data points 4. Averaging of data points in each cluster

Table K.6: Mitigation strategies: poor localisation technique - 2

Strategy	Description	Advantage	Risk	Incl	Functionalities/reasons
Leave the other human out	Adjust the registration process so that it can be used by one operator without the use of a second operator.This means the buttons should be easily available for the operator.	Communication errors will be limited and it leads to more focus being possible for the single operator. Moreover, it could reduce the registration time, training needed and human error.	When one operator performs the registration without the help of another operator, this could mean that when this operator is not feeling well or is not being focused, there is not operator to correct for these mistakes.	Yes	1. GUI that can be used by one operator 2. Hardware interface that can be used by one operator
Provide clear task division between both operators	When the registration is going to be performed by two operators, make a clear task and responsibility division between both operators	This will reduce the amount of communication errors and limit the human error. In addition it will make more clear who is liable for what if errors do occur. It will increase efficiency and safety procedure. It could decrease the workload when implemented correctly	It could lead to an increase in training needed and might limit the ease of use since both operators feel more responsible for specific tasks. It should also be avoided that the operators get supine when the task lies outside of their scope and that they do not solve any problems outside the scope. It could also increase the workload when not implemented correctly	Yes	1. For the case where two operators are used write down clear tasks for both operators 2. Write down responsibility for each operator
Provide training to operators	Provide a training procedure, either with GUI or without GUI, to train both operators in their tasks. It could be that both operators are trained on both tasks or just one task	The training procedure could lower the human error and increase the accuracy, ease of use and safety procedures. The operator will be trained how to respond to different scenarios and therefore will be more aware when problems occur	The addition of an extensive training procedure adds time, resources, costs and needs to be imbedded into the workflow	Yes	1. Provide a step by step training procedure 2. Provide different case studies 3. Provide a training procedure for both operator roles
Automatic registration when standing still	When the tip is held still at the same position for a prolonged time, automatically perform registration	The communication error between the operators is limited and when the operator is unable to press a button to register, registration is still possible. This could increase the efficiency and lower the workload and registration time	When the operator takes a break or is not moving for another reason, a new registration point is added without intent. This could decrease the accuracy and ease of use and could increase the training needed. It is possible to add a verification button to prevent registering points that should not be included	No	Automatic registration seems to come with additional risks and buttons or actions to prevent it from adding to the error. Therefore it will not be implemented. It could be implemented that the system provides a warning if the tip is hold still at an exact position for a prolonged time and asks the operator to register the point
Movement only possible after registration	When the tip is held still at the same position for a prolonged time, movement is disabled until registration took place	It prevents additional movement of the operator that is unwanted. This could add an additional safety procedure, and lower the human error	Again, if the operator is taking a break or another reason occurs to hold the tip at the same time for a prolonged period, then the operator is forced to register this point or press a button to ignore the point. It could interrupt the workflow of the operator, increase the registration time and training needed and decrease the ease of use	No	Implementing this functionality adds to the risks and is likely to make the system more difficult to use without having much added advantages. In addition, other methodologies can be used to prevent this error from occurring that require less radical interruptions of the operator workflow

Table K.7: Mitigation strategies: communication error between operators

Strategy	Description	Advantage	Risk	Incl	Functionalities/reasons
Leave the other human out	Adjust the registration process so that it can be used by one operator without the use of a second operator.This means the buttons should be easily available for the operator.	Communication errors will be limited and it leads to more focus being possible for the single operator. Moreover, it could reduce the registration time, training needed and human error.	When one operator performs the registration without the help of another operator, this could mean that when this operator is not feeling well or is not being focused, there is not operator to correct for these mistakes.	Yes	1. GUI that can be used by one operator 2. Hardware interface that can be used by one operator
Provide clear task division between both operators	When the registration is going to be performed by two operators, make a clear task and responsibility division between both operators	This will reduce the amount of communication errors and limit the human error. In addition it will make more clear who is liable for what if errors do occur. It will increase efficiency and safety procedure. It could decrease the workload when implemented correctly	It could lead to an increase in training needed and might limit the ease of use since both operators feel more responsible for specific tasks. It should also be avoided that the operators get supine when the task lies outside of their scope and that they do not solve any problems outside the scope. It could also increase the workload when not implemented correctly	Yes	1. For the case where two operators are used write down clear tasks for both operators 2. Write down responsibility for each operator
Provide training to operators	Provide a training procedure, either with GUI or without GUI, to train both operators in their tasks. It could be that both operators are trained on both tasks or just one task	The training procedure could lower the human error and increase the accuracy, ease of use and safety procedures. The operator will be trained how to respond to different scenarios and therefore will be more aware when problems occur	The addition of an extensive training procedure adds time, resources, costs and needs to be imbedded into the workflow	Yes	1. Provide a step by step training procedure 2. Provide different case studies 3. Provide a training procedure for both operator roles
Automatic registration when standing still	When the tip is held still at the same position for a prolonged time, automatically perform registration	The communication error between the operators is limited and when the operator is unable to press a button to register, registration is still possible. This could increase the efficiency and lower the workload and registration time	When the operator takes a break or is not moving for another reason, a new registration point is added without intent. This could decrease the accuracy and ease of use and could increase the training needed. It is possible to add a verification button to prevent registering points that should not be included	No	Automatic registration seems to come with additional risks and buttons or actions to prevent it from adding to the error. Therefore it will not be implemented. It could be implemented that the system provides a warning if the tip is hold still at an exact position for a prolonged time and asks the operator to register the point
Movement only possible after registration	When the tip is held still at the same position for a prolonged time, movement is disabled until registration took place	It prevents additional movement of the operator that is unwanted. This could add an additional safety procedure, and lower the human error	Again, if the operator is taking a break or another reason occurs to hold the tip at the same time for a prolonged period, then the operator is forced to register this point or press a button to ignore the point. It could interrupt the workflow of the operator, increase the registration time and training needed and decrease the ease of use	No	Implementing this functionality adds to the risks and is likely to make the system more difficult to use without having much added advantages. In addition, other methodologies can be used to prevent this error from occurring that require less radical interruptions of the operator workflow

Table K.8: Mitigation strategies: human tremor

Strategy	Description	Advantage	Risk	Incl	Functionalities/reasons
Filter out tremor	Add filters to the control loop that accounts for human tremor so that these are not acted upon by the robots end-effector	With filtering out the tremor, the accuracy, efficiency as well as the ease of use can be increased. It namely makes sure the tip is held at the same position. Additionally it could decrease the registration time, training needed and workload since the operator does not have to put in a lot of effort to maintain still	When the filtering is not implemented well, it could make it more difficult for the operator to use the admittance control and therefore decrease the ease of use. It should also be considered that when the filter fails it leads to additional maintenance and costs	Yes	1. Provide a low or high-pass filter that filters out tremor 2. Optimise filter parameters for human tremor 3. Optimise filter parameters for usability
Fixed mode	Provide a modus wherein the tip is held at the exact same position no matter what noise or disturbance is added to the end-effector. Then during this fixed mode the operator can perform the final registration to make sure no tremor influences this	The fixed mode has the advantage of making it more easy for the operator to perform the final registration without changes of interrupting. It increases the accuracy, ease of use, safety procedure. It could lower the registration time, human error, workload and training needed	The risk of implementing the fixed mode is that it could be harmful to the patient if something happens since the system is not compliant in this case. Moreover, it could require more training if the fixed mode is difficult to use	Yes	1. A mode where the tip stays in the exact same position 2. When a threshold of force is applied, cancel the fixed mode 3. An easy way to switch fixed mode on and off

Table K.9: Mitigation strategies: timing error

Strategy	Description	Advantage	Risk	Incl	Functionalities/reasons
Fixed mode	Provide a mode wherein the tip is held at the exact same position no matter what noise or disturbance is added to the end-effector. Then during this fixed mode the operator can perform the final registration to make sure no tremor influences this	The fixed mode has the advantage of making it more easy for the operator to perform the final registration without changes of interrupting. It increases the accuracy, ease of use, safety procedure. It could lower the registration time, human error, workload and training needed	The risk of implementing the fixed mode is that it could be harmful to the patient if something happens since the system is not compliant in this case. Moreover, it could require more training if the fixed mode is difficult to use	Yes	1. A mode where the tip stays in the exact same position 2. When a threshold of force is applied, cancel the fixed mode 3. An easy way to switch fixed mode on and off
Automatic registration	When the tip is held still at the same position for a prolonged time, automatically perform registration	The communication error between the operators is limited and when the operator is unable to press a button to register, registration is still possible. This could increase the efficiency and lower the workload and registration time	When the operator takes a break or is not moving for another reason, a new registration point is added without intent. This could decrease the accuracy and ease of use and could increase the training needed. It is possible to add a verification button to prevent registering points that should not be included	No	Automatic registration seems to come with additional risks and buttons or actions to prevent it from adding to the error. Thus, it will not be implemented. It could be implemented that the system provides a warning if the tip is hold still at an exact position for a prolonged time and asks the operator to register the point

Table K.10: Mitigation strategies: incorrect response to robot errors

Strategy	Description	Advantage	Risk	Incl	Functionalities/reasons
Guidelines	Provide guidelines on how to respond to robot errors. These guidelines can be given either before or during the registration. When provided before, they can also be included in the training procedure	The advantage of clear guidelines is that it provides an additional safety procedure and lowers the human error. It makes the system more robust	The risk of providing guidelines for robot error is that it could limit the operator to respond to the errors when they do not fall exactly in the provided guidelines. It could also increase the registration time	No	Due to time limitations this will not be implemented in the design. However, a protocol will be made to on how to respond to robot errors along the way as testing takes place so that this could be used in the future
Training	Provide an additional GUI or option within the GUI that can be used to train an operator based on different situations. For example first the operator is trained to perform registration normally, then an error occurs and it is explained to the operator what the error means and how to respond. In this training the most likely not-standard situations can be trained	An advantage of having the training as an additional option in the GUI is that it is prevented that bad training by the supervisor or management leads to errors since the operator can still get a quality training. It increases the robustness or the system, decreases the additional training needed, and provides an additional safety procedure	The additional GUI comes with an additional set in costs, maintenance and can clutter the GUI	No	Due to time limitations this will not be implemented in the design. However, a protocol will be made to train different operators in using the registration procedure and thus this information can be used when a training GUI will be made in the future
Possible to recalibrate	During registration provide an option for re-calibration when the operator finds this needed. For instance after bumping into the table, patient or robot during registration	This will give the operator more options to prevent large errors. It will be an additional safety procedure and increase the reproducibility	Performing calibration is a different task as compared to registration and therefore additional training is needed. Even so it could increase the workload and registration time	No	Since calibration is very much different from registration it is unlikely that this process will be carried out by the same operator as the registration. Different human resources and knowledge are needed

Table K.11: Mitigation strategies: perceptual error

Strategy	Description	Advantage	Risk	Incl	Functionalities/reasons
Sound when touching	When the operator touches something, a sound plays so that the operator is aware	This could increase the alertness of the operator when performing registration. Itis also an additional safety procedure for touching the patient	The implementation of this sound could distract the operator from performing its task and could add to the maintenance needed, since the sound could break	Yes	1. Play sound on first touch 2. Make sound stop after touching for a prolonged time 3. Use a sound that draws attention but does not distract or annoy operator
Vibrational feedback when touching	When the operator touches something, a vibration is given to the end-effector so that the operator is aware	The vibration could increase the alertness and function as an additional safety procedure for touching the patient	A vibrational feedback could change the position of the tip and thus affect the accuracy of the registration. It could also harm the patient or fiducial and tip when the vibration is too large. It adds to the costs, maintenance and amount of parts to the system, whereas it lowers the accuracy, ease of use as it can be perceived as annoying	No	The increased awareness can also be achieved with other functionalities that do not impact the accuracy or impose the amount of risks that vibration does to the system
Camera for visual feedback	Add a camera to the tip of the end-effector to provide visual feedback to the operator of its positioning with respect to the middle of the fiducial	Having a better visual feedback could increase the accuracy of registration since it can be With a zoomed in view that provides better feedback then the operators visual system	An addition of a camera leads to additional costs, maintenance and parts. The camera could break or malfunction and can become expensive. When implementing, it can also be difficult to provide the feedback needed since the view is 2D whereas the operator movements are in 3D. This could therefore be difficult for the operator to use and adds to the workload, training needed and registration time	No	The usage of a camera adds a lot of risks to the system such as the added maintenance and costs while it is not sure to lead to a better accuracy and performance for the operator and system
Prevent impaired operator usage	Prevent an operator with impaired senses to use the robot system by providing clear requirements to the operator or by testing its senses in a training situation	This could decrease the human error and provide an additional safety procedure	The addition of this check or requirements could lead to additional training needed and more maintenance when they have to be adjusted	No	Due to time limitations of this project and the limited advantage of this function it was not implemented. It was assumed that only operators with non-impaired senses were used to perform the registration procedure
Electrical/ circuit	Use an electrical circuit to find the exact middle of the fiducial screw. Here the circuit is between the fiducial and the tip	When the exact middle can be found this could increase the accuracy and efficiency since the operator does not have to perceive the middle by itself. It could make it more repeatable	The risk of an electrical circuit near the patient is that it could harm the patient. Moreover, if damaged is caused to the circuit it could harm both the robot, patient or operator. Since the patient has to carry the fiducials around for a prolonged time, it could be very dangerous	No	Implementing an electrical circuit near the patient is too dangerous
Magnetic attraction	Creating a small magnetic attraction between the tip and fiducial to increase the awareness of the operator of when both are touching	This could increase the feeling of touching the fiducial and therefore increase the accuracy, ease of use, reproducibility and decrease the workload	Since imaging is needed and even though CT is the most chosen method, adding a magnetic field to the fiducial can be dangerous when the CT is not available and MRI is the chosen method. Added it could lead to a decrease in patient friendliness. Also it should be considered that when the tip is attracted to the fiducial by a magnetic field it could make it more difficult to find the center point of the fiducial due to the added resistance in its movement	No	The addition of a magnetic field could increase a lot of hazards and might not add to the increase in accuracy that is intended
Add supervisor	During the localisation task at a supervisor that checks the operator performance and only when both agree on the result the surgery can take place	The addition of a supervisor adds to the robustness of the system and reduces the human error as well as adds to the safety procedure	The addition of another supervisor could lead to more costs, might increase the registration time and can affect the ease of use	Yes	1. Clear description of supervisor role 2. Clear description of operators role

Table K.12: Mitigation strategies: incorrect workings with materials and material damage

Strategy	Description	Advantage	Risk	Incl	Functionalities/reasons
Step-by-step GUI	Make a GUI with a clear checklist so that the operator is more engaged and skipping steps becomes more difficult. Add a checklist for materials that has to be checked before registration can take place	The operator is required to go through the checklist and thus responsibility lies at the operator. It adds to the robustness, repeatability and safety. Moreover, it lowers the training needed and human error	It adds additional time, workload, parts and maintenance (if checklist needs updating). It could also decrease the ease of use	Yes	1. When the registration is started make a step-by-step procedure for registration of four fiducials 2. Add a checklist that should be filled in by the operator before the registration
Operator training	Provide an additional GUI or option within the GUI that can be used to train an operator based on different situations. For example first the operator is trained to perform registration normally, then an error occurs and it is explained to the operator what the error means and how to respond. In this training the most likely not-standard situations can be trained	An advantage of having the training as an additional option in the GUI is that it is prevented that bad training by the supervisor or management leads to errors since the operator can still get a quality training. It increases the robustness or the system, decreases the additional training needed, and provides an additional safety procedure	The additional GUI comes with an additional set in costs, maintenance and can clutter the GUI	No	Due to time limitations this will not be implemented in the design. However, a protocol will be made to train different operators in using the registration procedure and thus this information can be used when a training GUI will be made in the future
Compulsory change	Keep track of material changes trough checklists and set time limits where after materials should be changed compulsory and the procedure can only continue of change takes place. Also provide warnings if materials have to be changed soon	This adds to the safety measures and makes it less likely for large errors to occur. It could increase the accuracy and lower the operator workload since the operator does not have to consider these changes themselves	A compulsory change could mean that in case of exceptions the procedure cannot continue which could be costly and make the usability low. It could also add to the costs since change takes place more often then maybe would have without these warnings	No	Due to time limitations of this project, this was not included. However, the functionality of adding material change to the checklist will be implemented for the registration tip and tip holder
Track materials	Provide a system that tracks material wear through calculations such as force used,time since last replacement, the amount of movements and type of material	This could prevent material damage to influence the registration procedure and thus increase the accuracy. It also increases the safety procedures, ease of use andlimits the amount of needed training since the computer tells the operator the status rather then the operator needing to check this themselves	The addition of material wear updates could distract the operator from its task and therefore should only be provided before the registration procedure. It also forms a risk where the operator only relies on the information provided and if this information malfunctions it could lead to large errors	No	Due to time limits of this project, a procedure of tracking material wear will not be implemented
Force limit	Provide a limit to the amount of force the operator can execute on the end-effector	This could decrease the amount of deflection and material damage since less force is exerted to the system. This could thus decrease the amount of maintenance and costs needed for the system and make it more robust	The risk in implementing a hard limit is that it could distract the operator and make it more difficult to use the system	Yes	1. Provide a warning when force exceeds certain limit 2. Exerting more force does not translate to faster movement above a threshold

Table K.13: Mitigation strategies: incorrect response to patient errors

Strategy	Description	Advantage	Risk	Incl	Functionalities/reasons
Add anatomical markers	When registration is performed, give the operator the option to use additional markers from anatomical structures. In this functionality it should be optional to add more markers as well as tell the system what those markers are	Increasing the number of markers could increase the accuracy as shown in the literature. It could also provide the operator with more options when a fiducial is missing, damaged or misplaced. Therefore, it adds as a safety procedure and can increase the accuracy and ease of use	The option to add more fiducials asks for more training of the operator and knowledge of the anatomical structures. It also requires additional functionalities to the GUI that could clutter it	No	Due to time limitations and the complexity of this functionality,it was not implemented. However, in the averaging functionality it will be considered how to give space for extra markers and label them
Correct wrong placement	When registration takes place and the fiducials are placed wrongly inside the patient, such as different orientation or difficult to reach location, give the operator options to still reach the fiducial by providing rotational control	The added rotational control gives the operator the capability to respond to misplacement and still find the middle of the fiducial. Therefore it could increase the accuracy, efficiency, ease of use and safety procedure while decreasing the registration time and workload	The added control options require more training, maintenance, costs and might influence the ease of use and registration time	**Yes**	1. Rotational control to adjust for orientation misplacement of fiducials 2. Switching mechanism to get rotational control
Possible to recalibrate	During registration provide an option for re-calibration when the operator finds this needed. For instance after bumping into the table, patient or robot during registration	This will give the operator more options to prevent large errors. It will be an additional safety procedure and increase the reproducibility	Performing calibration is a different task as compared to registration and therefore additional training is needed. Even so it could increase the workload and registration time	No	Since calibration is very much different from registration it is unlikely that this process will be carried out by the same operator as the registration. Different human resources and knowledge are needed
Connect table and robot	Connect table to the robot or vice versa so that movement of the table can not happen separately. This would mean that the position of a certain point on the table always has a fixed distance to the robot	When the table and robot are fixed it means that no misalignment between different registration points can take place. For instance, if you register fiducial 1 and then someone bumps in the table, the position of fiducial 2 with respect to fiducial 1 will be very different. When the table and robot are fixed, this would not occur. An advantage here is an increase in accuracy and safety procedure while decreasing the amount of parts	In case of usage in hospitals, it might be unrealistic to combine the table and the robot together. The robot will become very large and difficult to transfer from room to room. It would need an adjustment in workflow and can lead to additional costs	**Yes** /No	In the final test, both groups will perform registration on an object fixed to the robot table. This is done so that the external factors influencing the tests are limited throughout. However, for real-life it could be very difficult to design a table that is fixed to the robot and functions well in every hospital, it would likely be too costly to implement. 1. Have an object fixed to the robot base where registration takes place
Measure movement	A camera-marker system that could measure the movement of the table of patient with respect to the robot base and alarm the operator when such movement is detected	This measure could increase the error detection, accuracy and safety procedure of the system. The operator can respond better when such accidents occur and prevent large errors from happening	This could increase the cost, maintenance, parts, training time, and workload. The operator should know how to respond to movement and this could take a lot of time. Additionally, the camera could break and needs calibration from time to time to perform at high-accuracy	No	Due to time limitations and the added costs and maintenance to the system, this was not added in the design

L. Additional control information

This Appendix provides more information on the general admittance control loop improvements.

This chapter provides more detailed information on the general admittance control loop improvements as well as the parameter optimisation off the final admittance control loop in Figure 5.7.

L.1. General admittance control loop

As mentioned in 3.6, the admittance control loop from the robot looked like Figure 3.28. This project elaborates this control loop with better noise filtering and safety thresholds as summarised in Figure 5.8 and provided in this chapter.

L.1.1. Noise Filtering

The force sensor equips data about the force and torque applied to the robot end-effector. That data drives the robot arm into motion. Noise to this sensory input can lead to hazardous situations, such as the robot responding to a sudden force spike by moving at high speed towards the patient. Filtering noise to this sensor is, thus, critical. Filters can limit signals of a certain frequency from passing through. A low-pass filter filters high-frequency noise (Rajiv, 2022). Low-frequency noise is filtered by a high-pass filter (2022). Both are combined in a band-pass filter (2022). However, testing different filter types concluded that high-pass filters introduce a delay in the admittance control. Exclusively low-pass filters can be implemented without compromising the admittance control. The higher order of the filter, the better the cut-off, as shown in Figure L.1 and the more pleasing the filter (Raja, 2019). Hence implementing a second-order low-pass filter with a pole of 10Hz and a damping pole of 0.7. An example of the functioning of the second-order low-pass filter in response to force sensor reading is illustrated in Figure L.2.

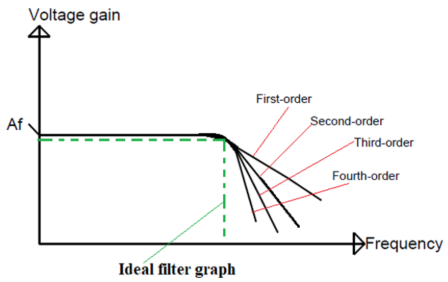


Figure L.1: A schematic example of a well-known low-pass filter with different orders. The filter shows that it passes through low-frequency signals, and the higher the order, the steeper the cut-off, and the more ideal the filter (2019).

L.1.2. Safety thresholds

In HRI, safety is crucial, thus enclosing safety thresholds in the admittance control loop design. These safety thresholds include force threshold, torque threshold and a positional error. Due to time restrictions, no workspace threshold was implemented.

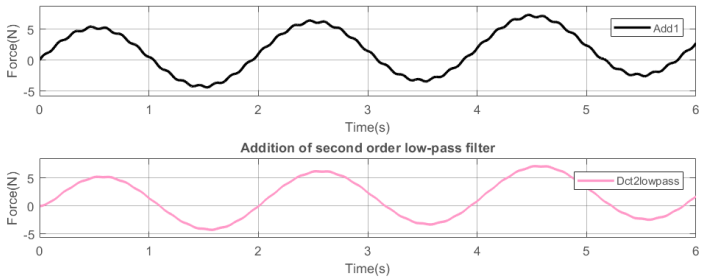


Figure L.2: An illustration of a force sensor reading with high- and low-frequency noise disturbance and using a second-order low-pass filter. With force in N on the y-axis and time in s on the x-axis. The top figure shows an example of force reading input from the operator with the addition of both high- and low-frequency noise. The bottom figure shows the signal after filtering it with a second-order low-pass filter with a pole of 5Hz and a frequency damping of 1.

Force threshold

The force vector limit is 8N. If the force vector applied by the operator exceeds this limit, the vector employed to control the robot remains at 8N according to the formula

$$F_{vec} = |\vec{F_{vec}}| \tag{L.1}$$

to find the normalised value of the force sensor force vector. According to the condition

if $F_{vec} > 8.0$ **then**
  Set $\vec{F_{vec}}$ to 8.0 times the normalised value $\frac{\vec{F_{vec}}}{F_{vec}}$.
else
  Leave $\vec{F_{vec}}$ unchanged.
end if

This threshold discourages high force sensor bias from causing sudden end-effector movement and thus the possibility of damaging or injuring the operator, patient or robot. Besides, the force threshold limits the operator's actions, which the operator will adjust. The operator will learn that exerting higher forces will not affect the robot's control and thus restrict its force. This lower force can prevent non-kinematic errors, such as bending the end-effector or registration tip.

Torque threshold

The torque threshold sets the torque limit to 0.2N. If the torque vector applied by the operator surpasses this limit, the vector used to calculate the cross product for the rotational output will stay 0.2N according to the formula

$$T_{vec} = |\vec{T_{vec}}| \tag{L.2}$$

to find the normalised value of the force sensor torque vector. Following the condition

if $T_{vec} > 0.2$ **then**
  Set $\vec{T_{vec}}$ to 0.2 times the normalised value $\frac{\vec{T_{vec}}}{T_{vec}}$.
else
  Leave $\vec{T_{vec}}$ unchanged.
end if

The torque vector usage in changing the heading angle is explained in more detail in the rotational mode. The torque threshold prevents high torque sensor bias from causing sudden robot joint rotations, which can cause singularities and internal collisions. Additionally, the threshold prevents the robot from injuring the operator or patient and limits the operator from applying a high torque to the system, thereby causing damage such as bending the end-effector or registration tip.

Speed threshold

The last threshold limits the moving speed of the robot towards the reference position by limiting the positional error. The further away from the reference position, the larger the positional error. The

moving speed is proportional to the positional error. The positional error is calculated with

$$Pos_{error} = K_{set_p} * (Pos_{ref} - Pos) \tag{L.3}$$

where K_{setp} is the positional gain set point, Pos_{ref} the end-effector reference position in [x y z] and Pos the current position of the end-effector in [x y z]. Normalising the positional error to a value of 1 with the condition

if $|Pos_{error}| > 1.0$ **then**
  Set Pos_{error} to $\frac{Pos_{error}}{|Pos_{error}|}$.
else
  Leave Pos_{error} unchanged.
end if

This prevents the robot from moving with excessive speed to its reference position in case the end-effector is far away. For instance, when the robot is positioned on the wrong side of the patient and the admittance control is turned on, the robot will move towards its reference position and not react to external forces unless the foot pedal is pressed. That can provoke significant harm to the patient. Limiting the movement speed permits the operator to react and limits the potential damage.

L.2. Parameter optimisation

The parameters are optimised accordingly. The most critical parameters are control gains, force sensor cut-off values, filter pole and damping, positional and heading reference and the safety thresholds. The results are summarised in Figure 5.11.

L.2.1. Control gains

The control gains were optimised with iterative tests, including human testers. A translational gain higher than 4.0 can lead to system instability in the contact dynamics. Lower values decrease the perceived control smoothness. The smoothness differs in all three translational directions and depends on the motor module orientations. For rotational control, a gain of 0.7 is optimal in discouraging operators from exercising high torques and overshooting the desired heading angle. Gains of 0.2 were found for the robot's movement to the safety position. Thus the robot moves slowly yet noticeable to the operator and supervisor.

L.2.2. Force sensor cut-off values

As mentioned, the force sensor is equipped with a low-pass filter. Optimising this filter by cut-off values [-1 1] for the force sensor and [-0.3 and 0.3] for the torque sensor. Higher values require high non-intuitive forces to initiate the robot movement. Lower values do not prevent accidental touches from initiating robot movement. Setting cut-off values normalises the force and torque measurements. Consequently, a safety threshold of 0.2N means a torque applied by the operator of 0.5N.

L.2.3. Filter pole and damping

Parameter optimisation established a low-pass filter with a pole of 10Hz and a damping pole of 0.7.

L.2.4. Positional and heading reference

The optimal positional reference is [315 -120 50] from the robot origin, assuming an adult-size head of 58cm; head placement at 100mm above the base plate on a pillow; average mastoid bone position; and an angle of 45 degrees from the centre axis of the head. A position closer to the patient can cause more hazards. A position at a greater distance from the patient requires more operator effort. The further away from the target, the more operators are inclined to exert higher forces that can yield dam-

age. This is in line with the localisation techniques found as operators tend to move faster with larger movements further away from a target.

A heading angle of [0 0.8829 -0.4695] or 62 degrees is best to match a human head lying with an angle of 45 degrees from the centre axis of the head. For the specific surgical area, this gives an angle of 28 degrees to the centre axis of the head and hence 62 degrees best matches this angle. This is explained in Figure L.3. Only the cosine has to be converted to -0.4695 to permit the vector to face the correct direction. The heading angle should always be a normalised value.

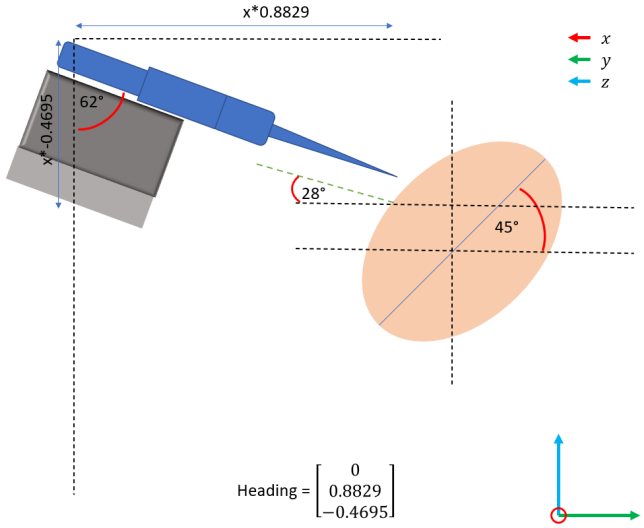


Figure L.3: A schematic overview of the optimal heading angle. Assuming that the patient lies with an angle of 45 degrees to the centre axis of the head, depicted with a dark blue line, it is given that the specific surgical path can be found with an angle of 28 degrees. This requires an angle of 62 degrees for the end-effector to match this angle and perform the best surgery. The coordinate frame is shown with x in red, y in green and z in blue. The registration bit length is given in a y-value of x*0.8829 and a z-value of x*-0.4695, with x being a scaling factor.

L.2.5. Safety thresholds

As mentioned, the safety thresholds for the force, torque and positional error are 8N, 0.2N and 1.0cm, respectively. Due to the positional reference gain in calculating the positional error as shown in L.2, it concludes a positional error of 5mm with a gain of 0.2.

M. GUI workflow

This Appendix describes the GUI and its workflow in more detail through description and a flowchart created with Draw.io.

M.1. GUI description

The GUI implementation was coded through Matlab AppDesigner. The GUI consists of the following steps, depicted in Figure 5.12: (1) Description, (2) Checklist before, (3) Localisation, (4) Results, (5) Checklist after, and (6) Validation. These steps and their functionalities are discussed below.

M.1.1. Description

The description step explains registration and its action steps. After reading this description, the supervisor can press the Next button, as depicted in Figure M.2.

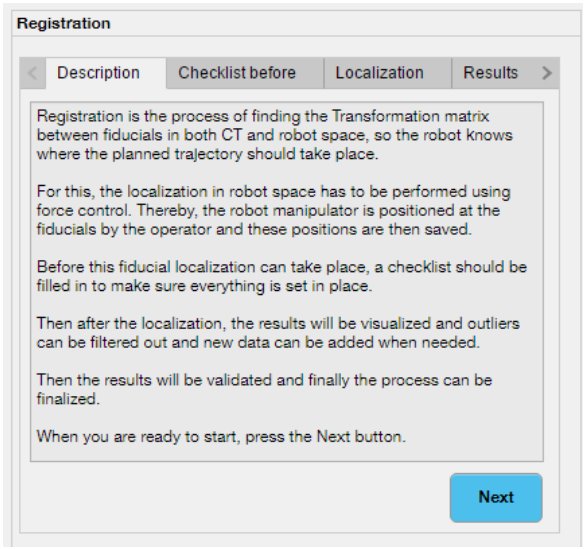


Figure M.1: The description step entailed in the description tab in the GUI explains registration and its action steps.

M.1.2. Checklist before

The checklist ensures the execution of checks so that registration can take place safely. The pedal is inspected for its functionality. If the pedal is not present or working accordingly, the GUI can control the robot, and the supervisor can help the operator perform the localisation. Next, the fiducials are examined in their presence and accessibility. The localisation can still proceed through adjusting calculations if there are only three fiducials. If less than three fiducials are functional, the operator and supervisor are urged to stop the procedure.

Subsequently, the registration bit is checked for presence and potential damage. Performing localisation with a drill bit instead of a registration bit affects the kinematic errors and influences the accuracy due to incorrect DH parameters.

The robot's beginning position should be near the patient and on the correct side. When the robot localisation starts, the end-effector moves towards the reference position. If the robot end-effector is located on the other side of the patient, it could harm the patient by executing a trajectory through the patient. Therefore it is crucial to check the starting position and place the robot accordingly.

Lastly, force bias introduced by the force sensor that limits the usability of the localisation or could harm the patient, operator or the robot itself should be avoided. Therefore, the supervisor should unbias the force sensor before starting the localisation step.

M.1.3. Localisation

The localisation process can start by pressing a ”Start” button in the GUI, initiating the safety mode automatically. Then the operator can perform the localisation using the foot pedal and the different modes as described. When a position is saved, its position is displayed to the supervisor, and the number of totally saved data points is updated. One important implementation is the ability to add fiducial locations in any order and multiple times. In this design, the localisation should be performed four times for each fiducial, leading to more accurate results and less influence of outliers or mistakes compared to performing it once. As the localisation is performed, the supervisor can press the ”Stop” button and then the Next button to move on to the Results step.

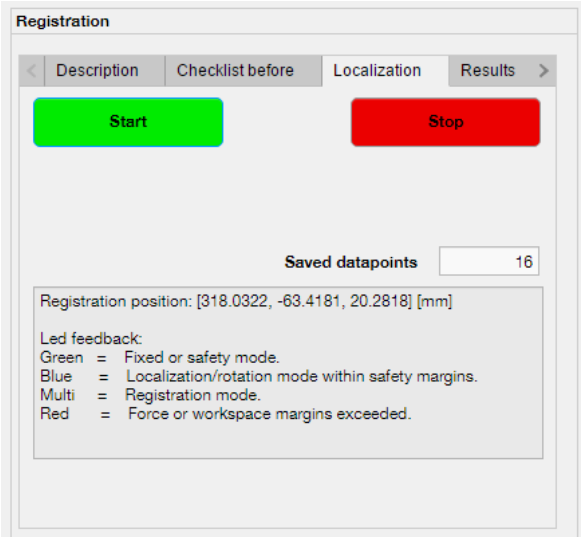


Figure M.2: The localisation step, which is entailed in the localisation tab in the GUI. During localisation, the operator can use admittance control and different modes to find the fiducial positions and save their data. When the data is saved, this is shown to the supervisor with the positional data. The number of saved data points is also updated upon saving. The supervisor needs to start and stop the admittance control by pressing the Start and Stop button, respectively.

M.1.4. Results

The results step provides visualised performance feedback of the operator with their measured positions and centroids shown in 3D. When outliers are found outside a certain threshold from these centroids, those are also visualised. Figure M.3 shows an example of this visualisation. The supervisor can manually adjust the threshold to carry out new calculations of outliers. For the outlier threshold, 1mm is the default setting. When outliers are detected, these can be removed from the dataset until satisfied. Additionally, the complete dataset can be removed if needed. The interface of the results step is portrayed in Figure M.4. When satisfied, the data can be saved, and the supervisor can continue to the next step.

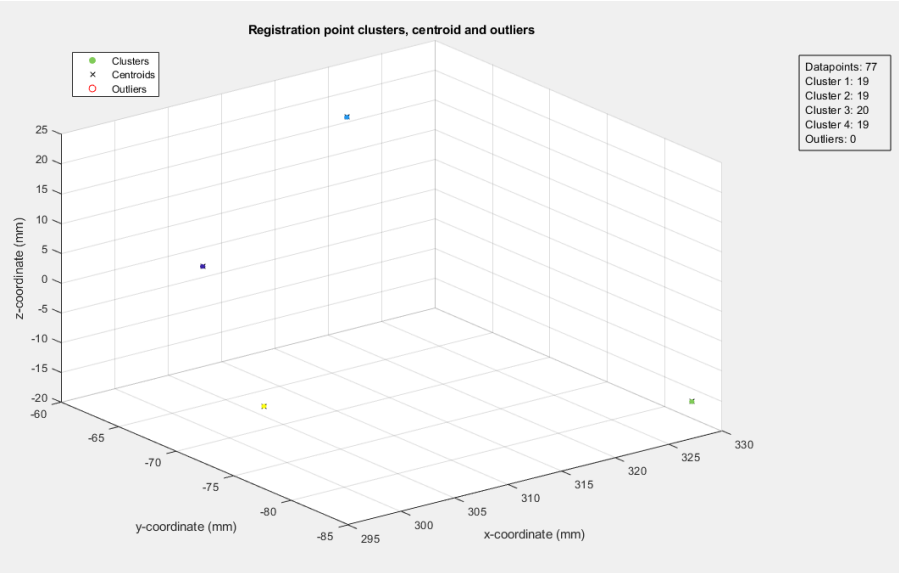


Figure M.3: An example of the visualised performance results of the operator localisation, showing four clusters and their centroids. If data points exceed the outlier threshold with respect to the centroids, they are highlighted.

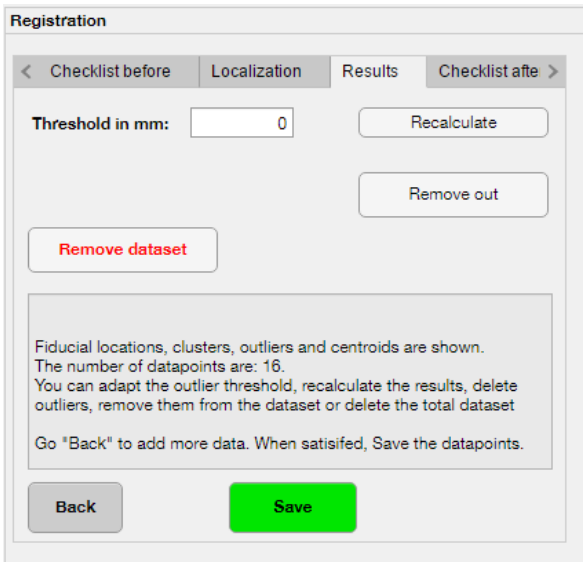


Figure M.4: The results step, which is entailed in the results tab in the GUI. Here the supervisor can adjust the outlier threshold to its liking and remove outliers when needed. The dataset can be entirely removed if it proves to be unsatisfactory. When the supervisor and operator are satisfied, the data can be saved, and the next step can be initiated.

M.1.5. Checklist after

This step prevents damages or changes in materials usage from going unnoticed. Therefore it is inspected if any damage has occurred in need of reporting. After that, the user has to check if admittance control was used. Conceivably admittance control was not working accordingly, and the robot was hard-coded to specific positions. This will significantly impact the accuracy and has to be notified. Additionally, it could be that four fiducials were initially recognised, but one of these four was unusable for localisation due to limited access, such as human hair blocking it. This should also be considered in calculations for the transformation matrix. All data is saved so that leading managers can check reported damages. Examples of the checklist after interface are pictured in Figure M.5.

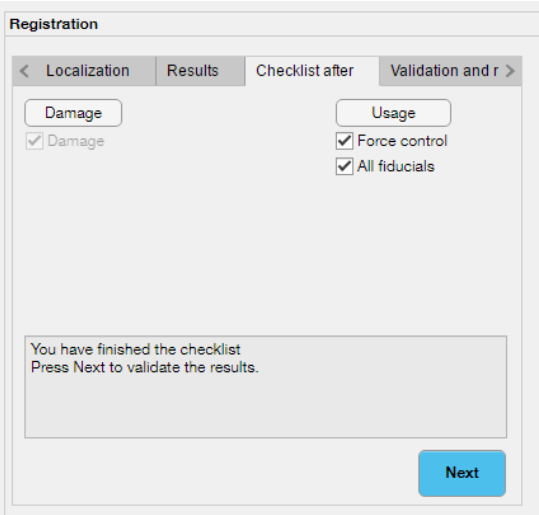


Figure M.5: An example of the checklist after step, which is entailed in the checklist after tab in the GUI. This step prevents damages or changes in materials usage from going unnoticed. First, the supervisor has to state the damage occurrence. After that, the usage of admittance control and all fiducials should be reported. In this figure, the checklist is filled in as an example. The interface can look different when boxes are unchecked.

M.1.6. Validation

Validation calculates the transformation matrix. Thereupon, the registration process is finished. The interface of this step is depicted in Figure M.6. Due to time limitations, no additional validation steps were implemented. For this project, it is assumed that the involvement of the operator and the supervisor terminates here. However, a surgery trajectory is generated directly after calculating the transformation matrix, after which the surgery starts.

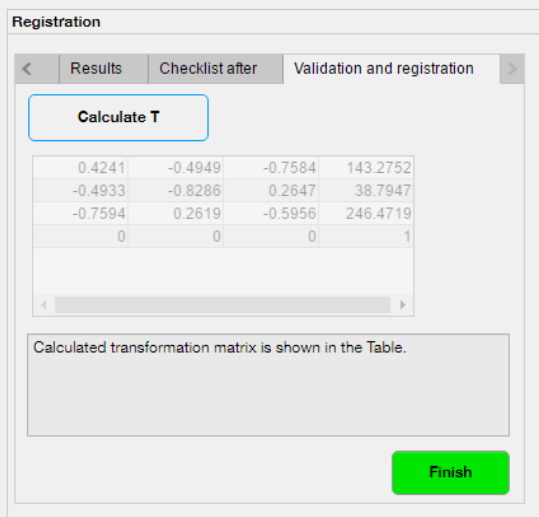


Figure M.6: An example of the validation step, which is entailed in the validation tab in the GUI. Here the transformation matrix is calculated. The Finish button finalises the registration procedure by initiating the saving process and resetting all data.

M.2. GUI workflow

This section describes the GUI workflow in detail for each seperate GUI section.

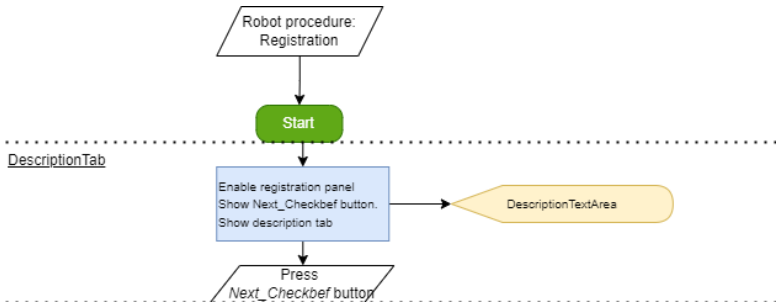


Figure M.7: A workflow associated with the registration GUI's first tab. An oval represents a start or end point; a line is a connector that indicates relationships between the representative shapes; a parallelogram represents input or output; a rectangle represents a process; and; a display symbol in yellow symbolises a step that displays information.

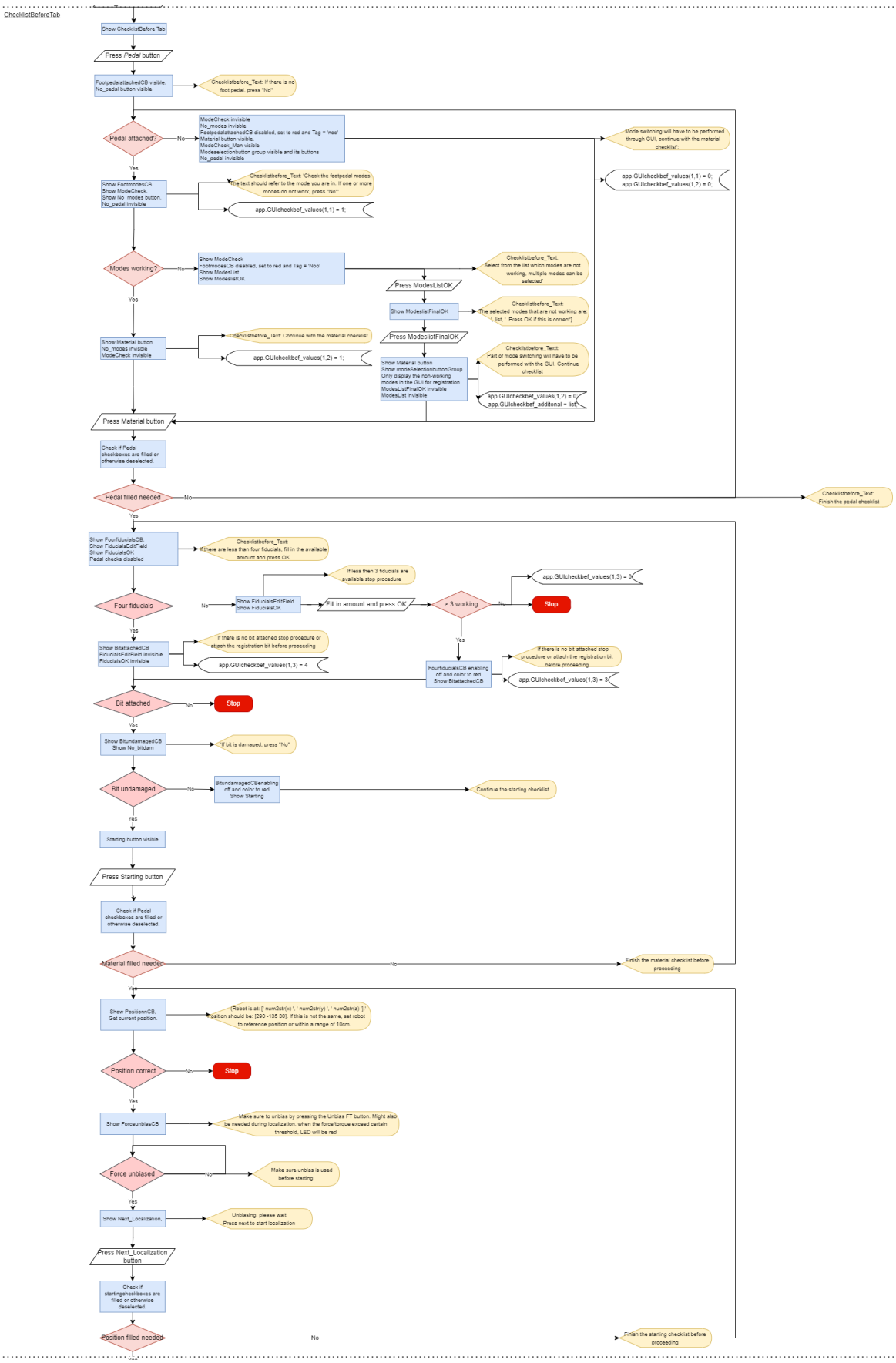


Figure M.8: GUI workflow related to the checklist before. An oval represents a start or end point; a line is a connector that displays relationships between the representative shapes; a parallelogram represents input or output; a rectangle symbolises a process; a display symbol in yellow symbolises a stage that displays information; and a diamond indicates a decision.

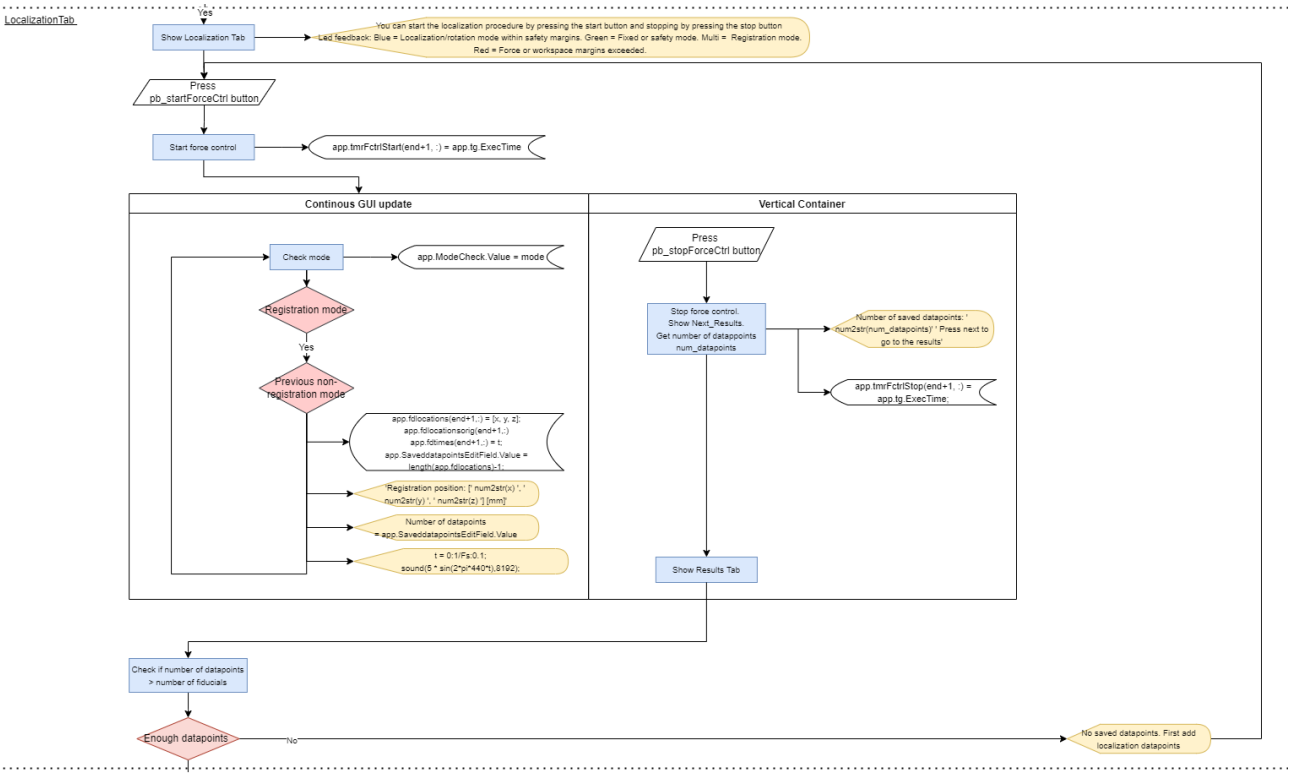


Figure M.9: GUI workflow related to the localisation task. An oval represents a start or end point; a line is a connector that shows relationships between the representative shapes; a parallelogram represents input or output; a rectangle represents a process; a display symbol in yellow represents a step that displays information; and a diamond indicates a decision.

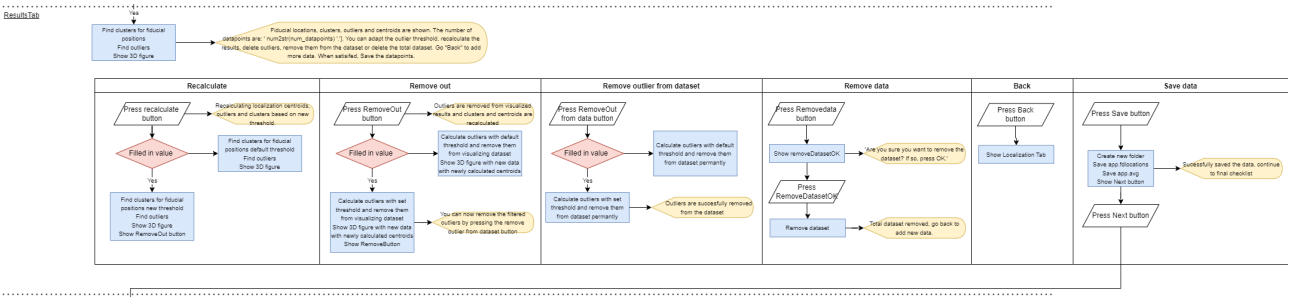


Figure M.10: GUI workflow related to the result visualisation, adaptation and feedback. An oval represents a start or end point; a line is a connector that shows relationships between the representative shapes; a parallelogram represents input or output; a rectangle represents a process; a display symbol in yellow represents a step that displays information; and a diamond indicates a decision.

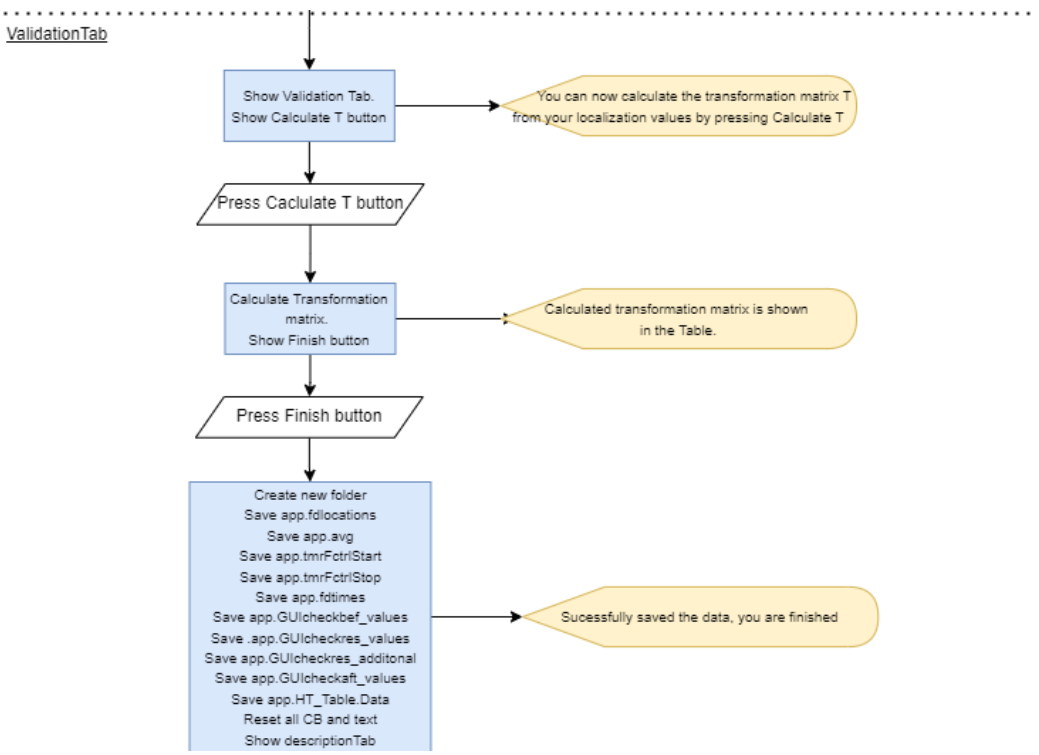


Figure M.11: GUI workflow related to a brief validation step. An oval represents a start or end point; a line is a connector that shows relationships between the representative shapes; a parallelogram represents input or output; a rectangle represents a process; a display symbol in yellow represents a step that displays information; and diamond indicates a decision.

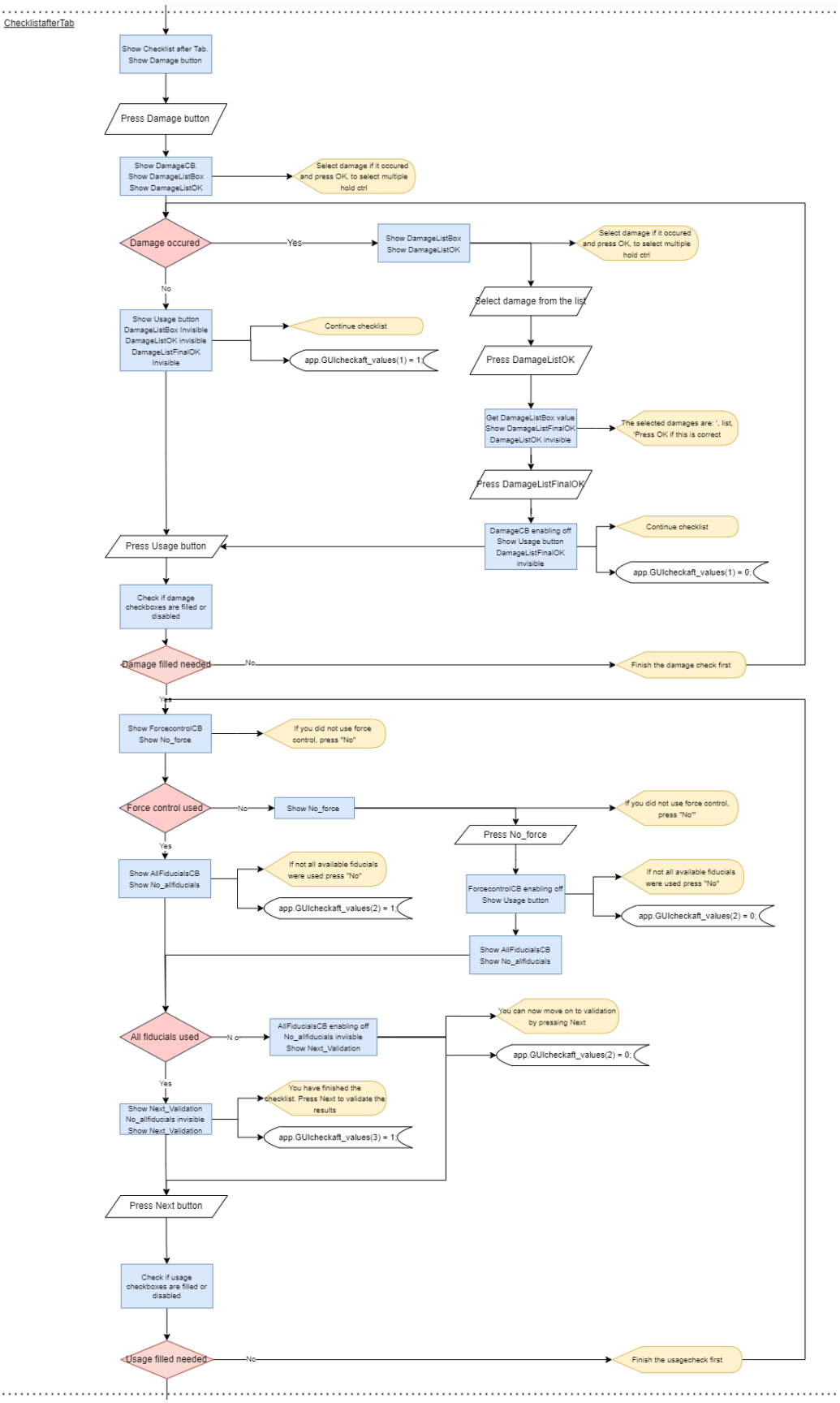


Figure M.12: GUI workflow related to the checklist after. An oval represents a start or end point; a line is a connector that shows relationships between the representative shapes; a parallelogram represents input or output; a rectangle represents a process; a display symbol in yellow represents a step that displays information; and a diamond indicates a decision.

N. Extended results

This Appendix provides additional results outside the most important results in Chapter 6.

N.1. Angle influence

Figure 6.10 is a summary of Tables N.1 and N.2.

Table N.1: The FLE value compared for different rotations of the heading angle for the new system. The FLE is compared based on its maximum (Max), median (Med), minimum (Min), IQR, the number of outliers, and the number of data points (Data).

Rot and axis	Max (mm)	Med (mm)	Min (mm)	IQR (mm)	Outliers	Data
0	0.3361	0.0582	0.0093	0.0507	7	224
5_x	0.2952	0.2273	0.0306	0.0405	2	160
5_y	0.4273	0.2231	0.1393	0.0558	8	160
5_z	0.5077	0.3051	0.2221	0.0627	2	160
-5_x	0.4023	0.2308	0.1418	0.0458	2	160
-5_y	0.2456	0.1154	0.0487	0.0342	3	160
-5_z	0.3225	0.2249	0.1440	0.0502	0	160
10_x	0.7308	0.4634	0.3726	0.0505	8	160
10_y	0.8247	0.3690	0.2850	0.0950	5	160
10_z	0.6445	0.4921	0.3778	0.0719	1	160
-10_x	0.4942	0.4194	0.3421	0.0717	0	160
-10_y	0.3285	0.2454	0.1642	0.0390	2	160
-10_z	0.5669	0.4890	0.4200	0.0385	1	160

Table N.2: The FLE value compared for different rotations of the heading angle for the old system. The FLE is compared based on its maximum (Max), median (Med), minimum (Min), IQR, the number of outliers and the number of data points (Data).

Rot and axis	Max (mm)	Med (mm)	Min (mm)	IQR (mm)	Outliers	Data
0	0.6371	0.1110	0.0228	0.0641	9	176
5_x	0.7781	0.2609	0.0508	0.0984	3	160
5_y	0.7178	0.1768	0.0498	0.0535	17	160
5_z	0.4095	0.2419	0.1564	0.0773	28	160
-5_x	0.9238	0.3406	0.1260	0.1491	4	160
-5_y	0.7659	0.2822	0.0513	0.1355	9	160
-5_z	1.9952	0.2943	0.0319	0.3890	11	160
10_x	0.5944	0.5205	0.3962	0.0402	4	160
10_y	0.4906	0.3140	0.2152	0.1094	34	160
10_z	0.7712	0.5658	0.4011	0.1331	0	160
-10_x	1.7361	0.5936	0.3067	0.3792	7	160
-10_y	1.8784	0.5296	0.0784	0.4134	5	160
-10_z	2.9782	0.4303	0.1748	0.3838	15	160

N.2. Skilled operator

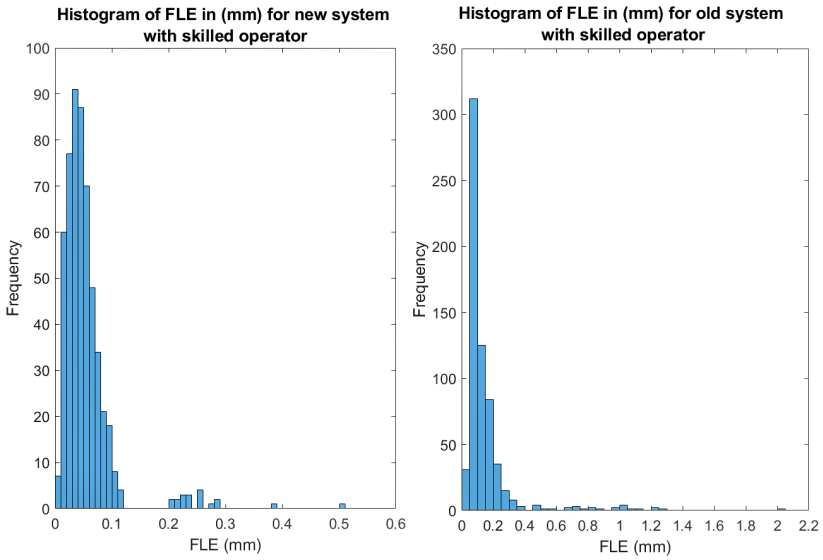


Figure N.1: A histogram of FLE skilled operator values, representing the frequency distribution of the FLE in mm for each system. The x-axis shows the FLE value in mm. The x-axis ranges from 0 to 0.6 for the new system and from 0 to 2.2 for the old system. The y-axis shows the related frequency, ranging from 0 to 100 for the new system and 0 to 360 for the old system.

N.3. TRE simulated

The TRE values were also simulated based on median FLE values from each participant, resulting in Figure N.2.

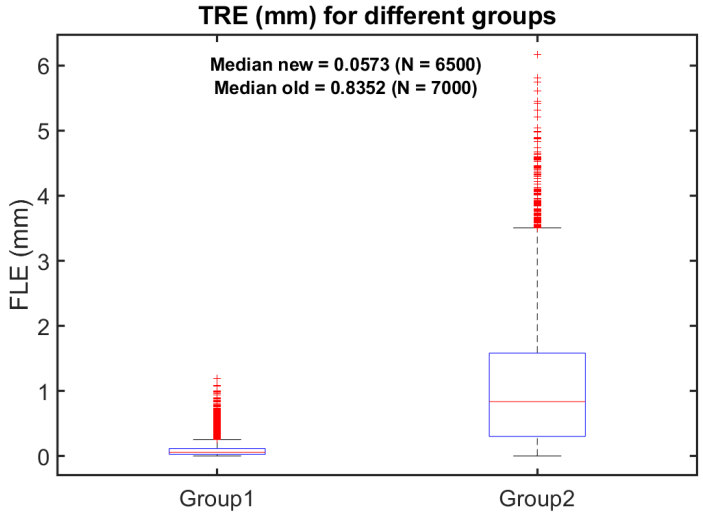


Figure N.2: Two boxplots showing the distribution of TRE values in mm over all participants for the new and old design based on simulating 500 different possible TRE values from each participant median FLE value. The x-axis shows the different systems with group1 being the new system and group2 the old system. The y-axis provides the TRE value, ranging from 0 to 2.5mm. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers.

N.4. Hold test

The system accuracies were also compared by holding the registration tip at one fiducial for 2 minutes for the new system in fixed mode, the new system without fixed mode and the old system. Their deviations were measured and showed in Figure N.3.

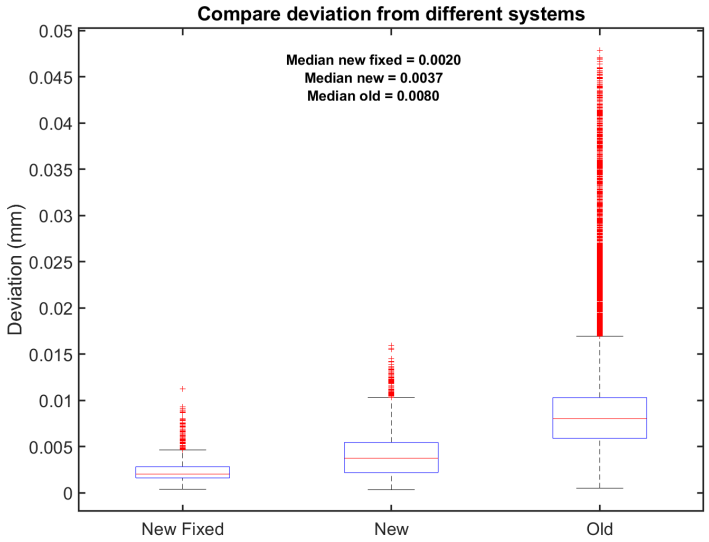


Figure N.3: Three boxplots showing the distribution of deviation in mm over 2 minutes of holding a fiducial tip at the same position for the new system with and without fixed mode, and the old system. The x-axis shows the different systems. The y-axis provides the deviation value, ranging from 0 to 0.05mm. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers.

N.5. Difference in x-, y- and z-direction

Table 6.2 is yielded from Table N.3.

Table N.3: The FLE value compared for different fiducials and directions for the new system. The FLE is compared based on its maximum (Max), median (Med), minimum (Min), IQR, the number of outliers and the number of data points (Data).

Fiducial and axis	Max (mm)	Med (mm)	Min (mm)	IQR (mm)	Outliers	Data
F1_x	0.2200	0.0013	-0.2200	0.0587	4	104
F1_y	0.6201	-0.0034	-0.3604	0.0477	14	104
F1_z	0.4120	-0.0010	-0.4740	0.0249	15	104
F2_x	0.1275	-0.0025	-0.1375	0.0406	6	104
F2_y	0.4493	-0.0008	-0.3251	0.0339	14	104
F2_z	0.3232	-0.0008	-0.2993	0.0169	13	104
F3_x	0.1850	-0.0075	-0.1950	0.0769	4	104
F3_y	0.2345	-0.0039	-0.2703	0.0493	8	104
F3_z	0.3798	-0.0036	-0.2512	0.0235	21	104
F4_x	0.2850	0.0025	-0.1575	0.0719	3	104
F4_y	0.5642	-0.0032	-0.3888	0.0516	7	104
F4_z	0.3977	-0.0013	-0.3539	0.0226	14	104

Table N.4: The FLE value compared for different fiducials and directions for the old system. The FLE is compared based on its maximum (Max), median (Med), minimum (Min), IQR, the number of outliers and the number of data points (Data).

Fiducial and axis	Max (mm)	Med (mm)	Min (mm)	IQR (mm)	Outliers	Data
F1_x	2.0338	0.1388	-0.3775	0.1813	24	112
F1_y	0.1132	-0.0154	-0.8593	0.4219	24	112
F1_z	0.0362	-0.8473	-2.1631	0.3423	16	112
F2_x	0.2962	0.0419	-0.5537	0.2781	2	112
F2_y	0.0724	-0.0592	-0.9210	0.1479	8	112
F2_z	0.0271	-0.7561	-1.4078	0.1822	24	112
F3_x	0.3325	0.0694	-0.5925	0.2350	6	112
F3_y	0.1369	-0.1234	-0.4019	0.1572	0	112
F3_z	0.0574	-0.7923	-1.0291	0.0757	32	112
F4_x	0.2725	0.0419	-0.3100	0.1694	0	112
F4_y	0.1140	-0.0876	-0.3451	0.1850	0	112
F4_z	0.0392	-0.7961	-1.2222	0.1028	32	112

N.6. NASA TLX

More details on the NASA TLX is displayed in Figures N.4, N.5, N.6, N.7, N.8 and N.9.

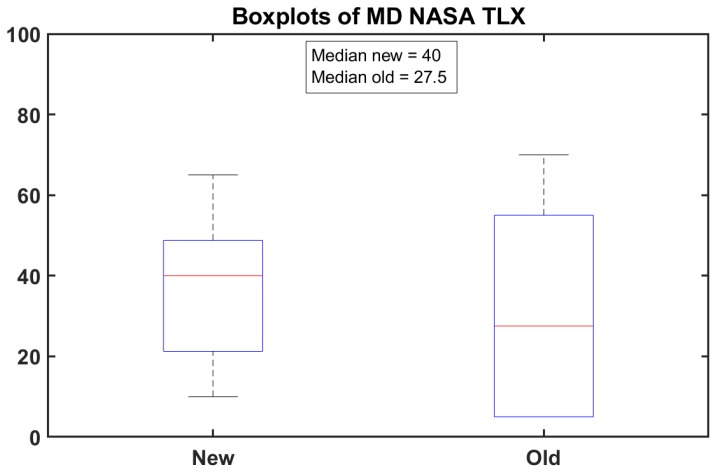


Figure N.4: Two boxplots showing the distribution of MD NASA TLX values over all participants for the new and old design, calculated from Equation 2.3 with their weights. The x-axis shows the different systems. The y-axis provides the NASA TLX value, ranging from 0 to 100. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers, with $P_1 = 13$ and $P_2 = 14$ the number of participants for the new and old system.

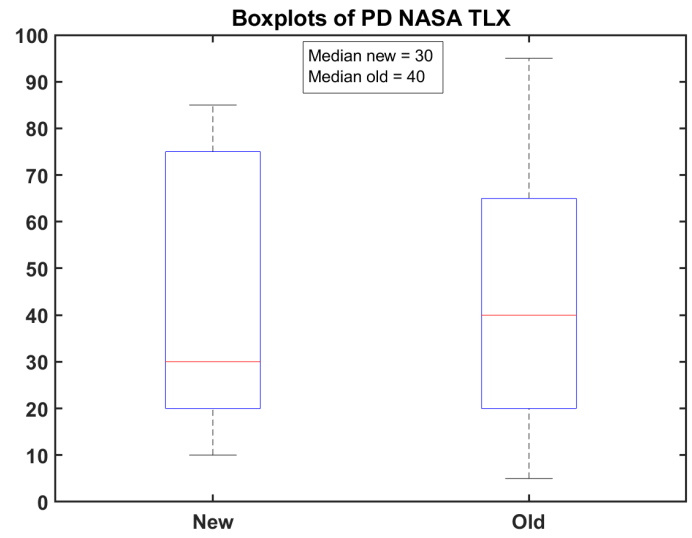


Figure N.5: Two boxplots showing the distribution of PD NASA TLX values over all participants for the new and old design, calculated from Equation 2.3 with their weights. The x-axis shows the different systems. The y-axis provides the NASA TLX value, ranging from 0 to 100. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers, with $P_1 = 13$ and $P_2 = 14$ the number of participants for the new and old system.

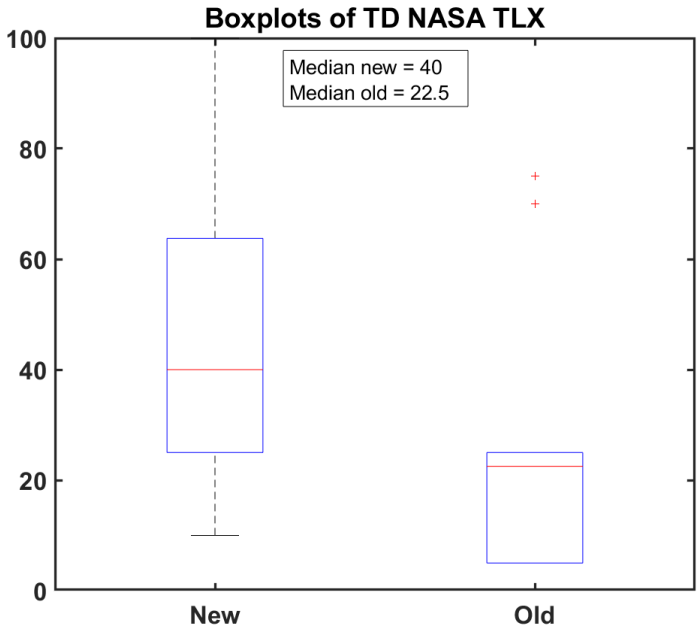


Figure N.6: Two boxplots showing the distribution of TD NASA TLX values over all participants for the new and old design, calculated from Equation 2.3 with their weights. The x-axis shows the different systems. The y-axis provides the NASA TLX value, ranging from 0 to 100. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers, with $P_1 = 13$ and $P_2 = 14$ the number of participants for the new and old system.

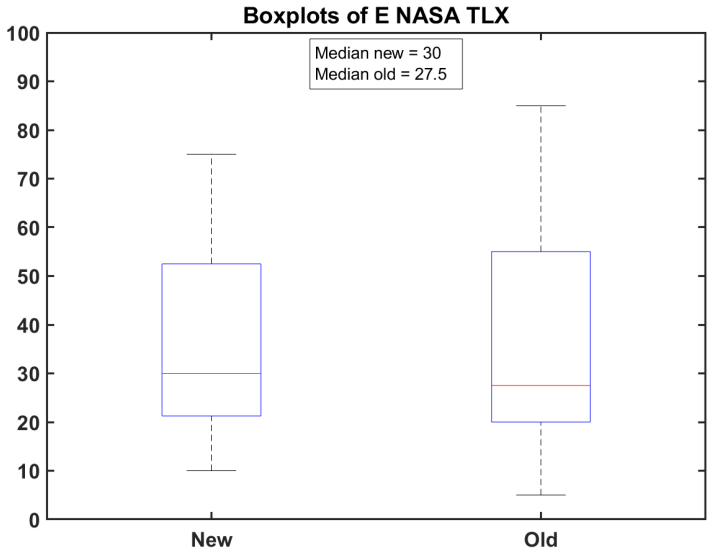


Figure N.7: Two boxplots showing the distribution of E NASA TLX values over all participants for the new and old design, calculated from Equation 2.3 with their weights. The x-axis shows the different systems. The y-axis provides the NASA TLX value, ranging from 0 to 100. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers, with $P_1 = 13$ and $P_2 = 14$ the number of participants for the new and old system.

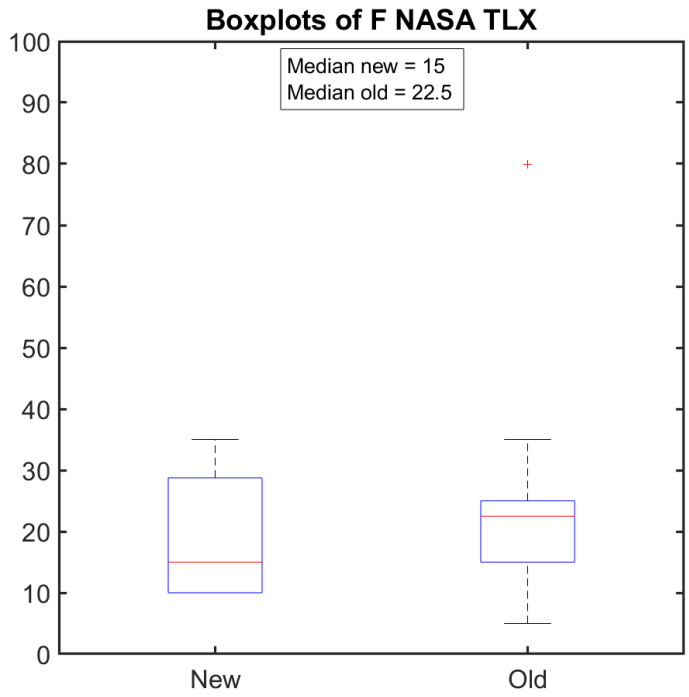


Figure N.8: Two boxplots showing the distribution of F NASA TLX values over all participants for the new and old design, calculated from Equation 2.3 with their weights. The x-axis shows the different systems. The y-axis provides the NASA TLX value, ranging from 0 to 100. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers, with $P_1 = 13$ and $P_2 = 14$ the number of participants for the new and old system.

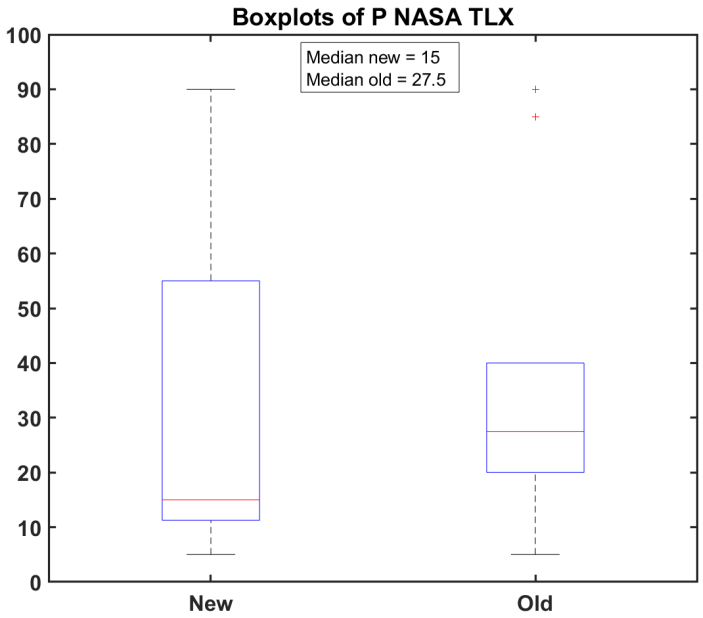


Figure N.9: Two boxplots showing the distribution of P NASA TLX values over all participants for the new and old design, calculated from Equation 2.3 with their weights. The x-axis shows the different systems. The y-axis provides the NASA TLX value, ranging from 0 to 100. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. The data with the red crosses are indicators of outliers, with $P_1 = 13$ and $P_2 = 14$ the number of participants for the new and old system.