Touchless, Direct Input Methods for Human-Computer Interaction to Support Image-Guided Interventions

Dissertation

zur Erlangung des akademischen Grades

Doktoringenieur (Dr.-Ing.)

angenommen durch die Fakultät für Informatik
der Otto-von-Guericke-Universität Magdeburg

von Benjamin Hatscher, M.A.
geb. am 13.10.1985 in Erlangen

Gutachter:
Prof. Dr. Christian Hansen
Prof. Dr. Rainer Malaka
Prof. Dr. Raimund Dachselt

Magdeburg, den 28.08.2020

Acknowledgements

This PhD thesis presents research that has been conducted at the Research Campus *STIMULATE* at the Otto-von-Guericke-University Magdeburg and was funded by the German Federal Ministry of Education and Research (BMBF) under grant number 13GW0095A and by the European Regional Development Fund (ERDF) under the grant numbers ZS/2016/04/78123 and ZS/2017/08/87591.

The author of this thesis further received the mobility grant of the German Society for Computer- and Robot-Assisted Surgery (CURAC).

Zusammenfassung

Bei minimal-invasiven Eingriffen werden Katheter oder Nadeln durch Inzisionen in den Körper des Patienten geführt um die Zielstruktur zu erreichen ohne dabei eine offene chirurgische Operation zu erfordern. Während eines solchen Eingriffs hat der Operateur weder direkte Sicht auf die Werkzeuge, noch auf Ziel- oder Risikostrukturen und ist auf bildgebende Verfahren wie Ultraschall, Röntgenbildgebung oder Magnetresonanztomographie angewiesen. Eine Anpassung der Echtzeit-Bildgebung, der Zugriff auf Planungsdaten oder auf bereits aufgenommene Bilder kann während einer Intervention erforderlich werden. Aufgrund der Gefahr die Sterilität zu verletzen, können herkömmliche Eingabegeräte wie Maus und Tastatur nicht ohne weiteres verwendet werden. Im der klinischen Praxis müssen Touchscreens, Tasten oder Joysticks mit steriler Folie abgedeckt werden. Alternativ werden Interaktionsaufgaben verbal oder gestisch an einen Assistenten delegiert.

Beide Ansätze sind nicht optimal: Stellvertreter-Interaktion ist anfällig für Missverständnisse und abhängig von der Erfahrung des Assistenten. Die Verwendung von steril abgedeckten, konventionellen Eingabegeräten ist umständlich und nicht möglich wenn beide Hände für die Durchführung der Intervention benötigt werden. Direkte Interaktion ist jedoch wichtig, um ein tieferes Verständnis für die medizinischen Bilddaten zu entwickeln.

Die vorgestellte Arbeit untersucht alternative Eingabemethoden für die direkte Interaktion mit Computersystemen in Situationen, in denen die Hände steril gehalten werden müssen und nicht jederzeit zur Verfügung stehen. Für MRT-geführten Nadelinterventionen wird eine berührungslose, einhändige Gestensteuerung vorgestellt und mit der Delegation an einen Assistenten verglichen. Für Situationen in denen die Hände nicht verfügbar sind werden handfreie Eingabemethoden zur grundlegenden Interaktion mit medizinischen Bilddaten untersucht. Als Eingabekanäle werden die Blickrichtung, die Füße, Sprachbefehle und Körperbewegungen eingesetzt. Sekundäre, zeitgleich ausgeführte Aufgaben wie die Interaktion mit Bilddaten beeinflussen möglicherweise die primäre, medizinische Aufgabe. Daher wird die Eignung und Auswirkung verschiedener handfreier Eingabemethoden während der Ausführung einer manuellen Aufgabe untersucht. Zudem werden passive Eingabemethoden für sekundäre Aufgaben aus natürlichem Nutzerverhalten abgeleitet und auf die resultierende subjektive Arbeitsbelastung untersucht.

Abstract

During minimally-invasive procedures, catheters or needles are inserted into the patient's body through small incisions and navigated to the desired structure without the need for open surgery. During such a procedure, the clinician does not have a direct view on the tools, the target or risk structures and therefore relies on imaging modalities such as ultrasound, X-ray or magnetic resonance imaging. Adjustment of live images, access to planning data or previously acquired images might become necessary during an intervention. Due to the risk of breaking asepsis, conventional input devices such as mouse and keyboards cannot be used without further measures. In clinical practice, touchscreens, buttons or joysticks need to be covered in sterile plastic sheeting. Alternatively, an assistant is instructed verbally or gesturally to act as a proxy user.

Both approaches are not optimal: Proxy-user interaction is prone to misunderstandings and relies on the experience of the assistant. Using plastic-draped, conventional input devices is cumbersome and not possible in case both hands are required to perform the intervention. However, direct interaction is essential to gain a deeper understanding of medical image data.

The presented work investigates alternative input methods for direct interaction with computer systems for situations where the hands are sterile and not available at all times. Touchless, one-handed gesture input is presented and compared to proxy-user interaction for MRI-guided needle interventions. Hands-free input methods for basic interaction with medical images are investigated for situations with the hands occupied. Gaze, feet, voice commands and body movements are employed as input channels. Direct, concurrently performed image manipulation might influence the primary, medical task. Therefore, the suitability and impact of different hands-free input methods while performing a manual task is investigated. Further, passive input methods for secondary tasks derived from natural user behavior are investigated regarding the resulting subjective workload.

Contents

Acronyms

HCI	human-computer interaction
NASA-TLX	NASA task load index
RTLX	Raw TLX
SEQ	Single Ease Question
SUS	System Usability Scale
OR	operating room
MRI	magnetic resonance imaging
CT	computed tomography
US	ultrasound
DoF	degree of freedom
FoV	field of view
TCT	task completion time
GUI	graphical user interface
UX	user experience
AR	augmented reality
VR	virtual reality
ISO	International Organization for Standardization

1 Introduction

Direct clinician-computer interaction during minimally-invasive procedures is crucial as it supports image interpretation [95] in a situation without a line of sight of the tools, the target or risk structures. The main challenges lie in maintaining asepsis in a sterile environment and the potentially occupied hands during interventions. Input devices that require direct contact, such as keyboards or touchscreens, bear the risk of bacterial contamination [54, 170]. Workarounds such as instructing an assistant, on the other hand, do not provide the required direct control over the medical image data [95].

To overcome the sterility issues during human-computer interaction in the medical domain, several approaches have been proposed. Input devices such as mouse [88] or Nintendo Wiimote controller [63] have been fitted into a sterile plastic bag, user interfaces have been optimized for proxy-user interaction [88], and touchless interaction methods were investigated [8, 36, 128]. The availability of commercial off-the-shelf gesture input devices such as the Microsoft Kinect 1 and the Leap Motion Controller led to increased numbers of publications on the topic of touchless interfaces with a focus on hand and arm gestures [128].

Gestural user interfaces solve the sterility problem but medical demands restrict their availability. Both hands might be occupied with holding instruments or catheter wires and cannot be used for human-computer interaction tasks [142]. Wrist-worn sensors that do not obstruct the fingers [94] or index finger gestures [131] aim to provide means of interaction without the need to put medical instruments down. As an alternative to the hands, voice input is deemed unsuitable for controlling continuous parameters but might be combined with other input modalities to allow hands-free interaction [142].

1.1 Contribution

This dissertation addresses the described issues by investigating methods for direct human-computer interaction when the hands are not available. For this purpose, a systematic investigation of hands-free alternative input modalities is conducted. The presented and evaluated approaches cover *foot steps, foot gestures, plantar pressure, voice commands, eye gaze, head* and *body movements*. They are used to perform basic interaction tasks such as activation of a system, navigation through medical image data and confirmation of selections. Further, *foot steps* and *eye gaze* are combined into a multimodal approach that allows discrete commands as well as coarse selection. An overview of the investigated input modalities and combinations for task types can be found in Table 1.1.

Foot pedals are a common, and therefore, an familiar input device in the operating room (OR).The foot-based approaches presented in Chapter 4 of this thesis contribute to the solution of two problems with conventional foot pedals. First, pedals can get lost under the operating table and are cumbersome to retrieve. Tactile floors and wearable input devices do not suffer from this disadvantage. Second, the limited interaction possibilities of foot pedals are extended by employing foot taps, pressure distribution and rotation on the heel. These gestures and movements allow performing discrete as well as continuous interaction tasks.

In an interventional scenario, the medical task is the top priority and must not suffer from direct interaction with medical images. This dissertation contributes by investigating how different input modalities for a secondary task influence the performance and accuracy of a primary one when executed concurrently. Further, natural user behavior is utilized as a passive input method to tackle the issue of additional workload through concurrent task performance.

An overview of the investigated input techniques regarding task type and input modality can be found in Table 1.1. The investigated tasks are categorized as discrete when a single input signal is required to switch between two states and as continuous input when a value in one or more dimensions needs to be adjusted or set.

1.2 Structure

This thesis is organized as follows:

- **Chapter 2** provides medical, technical and methodological background information for subsequent chapters. An overview of image-guided procedures outlines the today's issues and challenges of human-computer interaction in the medical domain. Further, touchless methods for human-computer interaction and research methods this thesis builds on are described.
- In **Chapter 3**, the fundamental question of whether direct control over the imaging modality is desired over clinically conventional interaction

methods is investigated. The current situation in clinical practice is assessed with an online survey among experienced radiologists. An approach for direct, touchless control of a magnetic resonance imaging (MRI) scanner is presented and compared against the proxy-user method.

- **Chapter 4** presents hands-free, direct interaction methods using the feet as they are a already established input channel in the form of foot pedals. Four foot-based methods for interaction with medical image data are described and evaluated. The chapter concludes with a summary of the advantages, disadvantages and suitability of the investigated approaches.
- **Chapter 5** investigates the use of multiple input modalities to compensate for the shortcomings of a single input method. Comparative user studies determine the suitability of three hands-free input modalities for image manipulation. A multimodal system is created and evaluated, producing insights in cross-modality influences.
- **Chapter 6** - As the presented direct input methods are meant to be used during medical interventions, minimal influence on the primary, manual task is crucial. This chapter investigates the mutual influence of primary and secondary interaction tasks and proposes an approach to find input methods with a low subjective workload.
- **Chapter 7** concludes the thesis by summing up the findings and limitations of this work. Open questions emerging from this work are discussed, and an outlook of how this work might contribute to the operating room of the future is given.

Table 1.1: Overview of investigated combinations of input modalities and task types. Corresponding sections of this thesis are shown in parenthesis.

Task Type	Hand	Foot	Voice	Eye Gaze	Head	Body
Discrete						
Activating	Dwell-time (5.1)	Triple-tap (4.1), Double-tap (4.2, 5.1), Toe/heel pressure (4.3)	Keyword (5.1, 6.2)	Dwell-time (5.2)	Nod/Shake (6.2)	
Triggering			Keyword (5.1)	Buttons (5.2)		
Confirming	Air-Tap (6.1), Bent Thumb (3.3)	Triple-tap (5.2), Tap (6.1)	Keyword (6.1, 6.2)	Dwell-time (5.2)		
Continuous						
Slicing	Lever-metaphor (5.1), Palm-down Gesture (3.3), Tap and hold (6.1)	Ball/heel pressure (4.3), Fixed-rate buttons (4.1, 5.2), Heel rotation (4.4, 5.1)		Fixed-rate edges (5.2), Head direction (6.1)	Fixed-rate left/right (6.2)	
Zooming					Lift/lower eyebrows (6.2)	Leaning (6.2)
Pointing	Extended index finger (3.3)			Eye-gaze (5.2)	Head direction (6.1, 6.2)	
Rotating		Fixed-rate buttons (4.1, 5.2)		Fixed-rate edges (5.2)		
Panning					Head direction (6.2)	

2

Background

This chapter provides fundamentals from the medical field as well as from human-computer interaction. It starts with outlining minimally-invasive methods, how image data is employed, and which kind of tasks are performed in sterile settings. The second part provides an overview on touchless human-computer interaction methods and multimodal approaches, which are investigated for touchless interaction in the subsequent chapters. At last, methodological definitions and tools that are relevant across the presented studies are described and explained.

2.1 Clinical Background

2.1.1 Minimally Invasive Procedures

The term ”minimally-invasive” describes approaches that aim to minimize the size of incisions needed to perform certain kinds of radiological or surgical interventions. This includes interventional radiology, laparoscopic surgery, percutaneous needle-based approaches and endovascular procedures.

In minimally-invasive percutaneous procedures, tubular devices such as needles, catheters or tissue ablation probes are inserted through the skin [2]. This can be used for the thermal ablation of tumors by delivering high temperatures (radiofrequency, microwaves) or low temperatures (cryoablation) into the tumor tissue [33] or for gathering tissue samples during biopsy [44].

In 1953, a technique for safe reproducible access to the vascular system via a catheter was presented [89]. This enabled a range of treatments for vascular lesions in a minimally-invasive fashion where a guidewire is inserted through small incisions at a remote site and navigated to the desired structure through the patients blood vessels [206]. Compared to open vascular surgery, the advantages of endovascular techniques are faster patient recovery, less morbidity and shorter hospital stays [31]. Endovascular techniques and interventional radiological procedures yield lower morbidity rates, faster recovery of patients, earlier hospital discharge [201, 31] and lower mortality rates [201] compared to open surgery. As a result, interventional radiology has increasingly replaced open surgical techniques over the years [201], and vascular surgeons used more and more endovascular treatment methods [206].

The advantage of minimally-invasive approaches comes at a cost: physicians have no direct line of sight on their tools and rely entirely on image-guidance using technologies such as fluoroscopy (continuous X-ray imaging), computed tomography (CT), ultrasound (US) or MRI [95]. The continuously produced images might be used for planning, guidance, reference, postprocedure assessment or documentation [168, 189, 95]. This dependency on images requires substantial interaction with imaging technology to capture, browse and manipulate the acquired images [95].

2.1.2 HCI in Clinical Practice

Conventional input devices such as a mouse, keyboard or touchscreens are prone to bacterial contamination [170, 54] and cannot be used as input devices in a sterile environment without further measures such as sterile plastic covers [78, 88]. To this end, a range of workarounds to overcome this limitation exists. Johnson et al. reported that a surgical gown was used to provide a boundary between the sterile gloved hand of a radiologist and the non-sterile computer mouse [95], even though concerns were expressed as such practices are not risk-free [95, 142] (see Figure 2.1). Another common approach is proxy-user interaction, where interaction tasks are delegated verbally or gesturally to an assistant next to the physician or in a non-sterile control room, which then operates the computer [211, 77, 58, 126]. Some of the various terms for this method are ”assistant controlled computer keyboard” [230], ”task delegation” [78], ”yell and click” [211] or ”assistant-in-the-middle” [59] (see Figure 2.2). This approach is not optimal for several reasons:

- Giving orders and verifying execution requires the performing physician to pay close attention [58].
- Team members may not be available immediately to act as proxy-user [141].
- Verbal task delegation might lead to misunderstandings [58, 59, 153].
- Recovering from input errors can be difficult [59].
- It may require the performing physician to perform the given input and resterilize afterward [59, 58], which is time-consuming [126].
- An assistant requires a certain level of experience and knowledge [153, 126]

Figure 2.1: A radiologist uses a non-sterile mouse with sterile gloved hands through a surgical gown. Image from Johnson et al. [95] © ACM 2011[1].

Figure 2.2: An assistant is instructed to operate the computer on behalf of the surgeon. Image from Grange, Fong, and Baur [59], reprinted with permission.

Tasks Depending on the medical field, the tasks to be performed vary. In the following, the input tasks which are investigated in subsequent chapters are categorized and connected to exemplary tasks from literature:

- **Discrete** input can only handle two states. This is required for system activation mechanisms which allow a transition from an inactive state into one that allows user input [141, 36]. In literature, this is referred to as ”clutching” [36] or ”sleep mode” [213]. Further, confirmation of a previously performed selection or triggering a specific function belongs to this kind of task.
- **Continuous** Input describes an unsegmented, ongoing signal that is used to control a value in the interactive system. Depending on the capabilities of the input modality, one or more values can be manipulated simultaneously.
  - **One Degree of Freedom** is required when a single value needs to be adjusted. This is the case for zooming [53] but is also common for navigating image series even though going to the next or previous image theoretically is a discrete interaction. The reason why it is treated as continuous in the context of this thesis lies

[1]Republished with permission of ACM, from Rose Johnson, Kenton O’Hara, Abigail Sellen, Claire Cousins, and Antonio Criminisi. 2011. Exploring the potential for touchless interaction in image-guided interventional radiology. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’11). Association for Computing Machinery, New York, NY, USA, 3323–3332. Fig. 3. DOI:https://doi.org/10.1145/1978942.1979436, permission conveyed through Copyright Clearance Center, Inc.

in its application: A series of medical images might be navigated back and forth multiple times at different paces to support image interpretation [95].

- **Two Degrees of Freedom** can be used for 3D rotation, panning [77] or pointing tasks. *Intensity windowing* describes the adjustment of two values of a linear transfer function, which maps grey values of an image onto the usually smaller range of a display [157, 93]. *Window width* determines the range to be displayed while the *window level* determines its center. Described more practically, it appears similar to adjusting the contrast (*width*) and brightness (*level*) of an image. This could be achieved by manipulating two single values independently. Professional radiological image viewer software often maps both values to the two axis of a mouse. Such a mapping has been adopted for touchless approaches [40, 53] or VR environments [228].

2.2 Touchless Human-Computer Interaction

Given the requirements for human-computer interaction in medical scenarios, alternatives to conventional input devices such as mouse and keyboard are required to allow direct, sterile input that does not interfere with the task at hand. Recent literature reviews reflect the importance of this issue: Mewes et al. report touchless image manipulation as the main objective for 34 of the 55 publications reviewed [128] and Alvarez-Lopez, Maina, and Saigí-Rubió found image manipulation to be the most common interaction (42/86) [8]. Cronin and Doherty found sterility the most common motivation for touchless control (27/41) and the OR (32/41) as well as interventional radiology (11/41) the most frequent context of use [36]. This section provides an overview of unimodal touchless input methods and multimodal approaches in general and for sterile environments.

2.2.1 Touchless Input Methods

Gestures

According to Mitra and Acharya, "gestures are expressive, meaningful body motions involving physical movements of the fingers, hands, arms, head, face, or body with the intent of: 1) conveying meaningful information or 2) interacting with the environment." [133]. Gestures are continuous motions, which consist of five phases: a *rest position*, a *preparation phase*, a *gesture stroke*, *holds* and a *retraction or recovery phase* [26].

As input modality, gestures bear several advantages: They are a *natural* form of communication that is easy to learn and enables *terse and powerful* interaction as a single gesture may encode the execution of a command along with additional parameters [19]. Further, hand gestures allow *direct* input, which does not require intermediate transducers while emulating other input devices such as mouse, keyboard or touchscreens is still possible [19].

For the medical domain, the primary motivation for using touchless control is in *sterility*, *three-dimensional applications*, *busy hands* and *removing barriers* in current practice [36]. For general purposes, gesture interaction has been found in the domain of distant displays, operating rooms, 3D spaces, ubiquitous environments, TVs, therapeutic assistance, accessibility, cultural heritage, text entry and data sharing among devices [105]. With a focus on interventional radiology and surgery, a literature review by Mewes et al. identified applications for gesture-based systems in medical image viewer control, laparoscopic assistance, telerobotic assistance, OR control, robotic OR assistance and intraoperative registration [128].

Gesture styles A basic classification of gestures is the separation into *discrete* and *continuous* gestures [160, p. 500]. Karam et al. proposed a taxonomy with the four categories *application domain*, *technologies*, *system response* and *gesture styles* [98]. Based on McNeill [123], Quek et al. [161] and Preim and Dachselt [160, p. 498-499], categories relevant in the context of gesture-based human-computer interaction (HCI) are described in the following.

- **Deictic Gestures** are pointing movements or poses. They are used to identify objects or locations [98, 123]. A typical example of deictic gestures in HCI is the extended index finger when used for pointing [145, 171]. Deictic gestures can be combined with other input methods such as voice commands [23] because gestures are more suitable for adjusting continuous parameters than voice [141].
- **Manipulative Gestures** apply a tight relationship between movements and the manipulated entity [161]. Interacting with elements that represent a continuous value, such as sliders, require this type of gesture. In the context of medical image manipulation, typical use cases for this gesture type are browsing through a dataset [39, 127, 171], panning, zooming [39, 167, 127, 171], rotating [167, 127, 171] and windowing [39, 167].
- **Semaphoric Gestures** are elements of a predefined set of gestures [161]. Gesture types are often combined [98], which means that semaphores can also be deictic or manipulative.

- **Iconic Gestures** accompany speech and bear a close relationship to the semantic content [123]. The context in which the gestures are interpreted is set by speech; for example, when the size or shape of an object is depicted using hand movements while talking [160, p. 491]. Manipulative gestures can be derived from iconic gestures to make them easier to remember. Examples are changing the distance between thumb and index finger for zooming [145] or performing a "grab" gesture to initiate manipulation of a virtual object [127].

Technical approaches Recent literature reviews found the Microsoft Kinect (Microsoft Corp., Redmond, WA, USA) and Leap Motion controller (Leap Motion Inc., San Francisco, CA, USA) as the most common gesture tracking devices for prototypical implementation [128, 36, 8]. Both devices use infrared cameras while the Microsoft Kinect 1 works on structured light, the Kinect 2 uses the time-of-flight principle [217] and the Leap Motion Controller employs three LEDs and a stereo camera which operates in the infrared spectrum [128]. The range of both devices differs as the Leap Motion Controller is intended for hand-tracking while the Microsoft Kinect tracks a person's skeleton in front of the device [8]. Other technical approaches include RGB cameras, time of flight cameras or inertial sensors [128].

Challenges Performing mid-air gestures for a more extended period, however, leads to muscle fatigue, the so-called gorilla-arm effect [24, 66]. To avoid this, interactive elements should be placed in a way so that they can be reached when the arm is bent and relative movements should be favored over absolute ones [79].

The overall goal of allowing the physician to access medical data directly when scrubbed in implies that the hands might be occupied by the primary task of performing the actual medical intervention. Therefore, the number of available hands needs to be considered when designing gesture interfaces [142].

Distinguishing intended gestures from movements unrelated to the task is called *gesture spotting*. It is technically difficult to find the right cues to determine when gestures start and end [214].

Foot Input

As the hands might not always be available for direct input during an intervention [142], using the feet to control input devices might be a reasonable approach. The lower limbs have been considered for human-computer interaction since the early days of HCI research, motivated by the upcoming need to interact with computer systems more efficiently [151]. Before the

Table 2.1: Range of motion in male subjects between 30-40 years of age [165] and corresponding interaction methods, based on Velloso et al. [208]. Methods are listed multiple times in case combinations of movements are required for execution.

	Range of Motion			Interaction Method
Joint	**Movement**	**M [°]**	**SD [°]**	
Ankle	Dorsiflexion	15.3	5.8	Ball tapping [41], Hanging foot switch [185], Pressure on the rear foot [12, 51, 192]
	Plantar flexion	30.7	7.5	Foot pedals [106], Pressure on the forefoot [12, 51, 192], Knee Lever [41]
	Inversion	27.7	6.9	Pressure on the outer foot [12, 51]
	Eversion	27.6	4.6	Pressure on the inner foot [12, 51]
Knee	Flexion	143.8	6.4	Kicks [147, 62, 7]
	Extension	1.6	2.8	
Hip	Flexion	120.3	8.3	Kicks [147, 62, 7]
	Extension	9.4	5.3	
	Medial rotation	32.6	8.2	Heel rotation [183, 232, 72, 162]
	Lateral rotation	33.6	6.8	
	Abduction	38.8	7.0	Knee Lever [41]
	Adduction	30.5	7.3	Knee Lever [41]

mouse was presented by English, Engelbart, and Berman, it was evaluated – among others – against knee input for text cursor control on a computer workplace [41]. In 1986, Pearson and Weiser proposed different topologies and designs for foot-controlled input devices to release the hands from the double-role of text entry and cursor-positioning [151]. Since then, a wide range of foot-based human-computer interfaces have been proposed. Velloso et al. give a comprehensive overview from the perspective of the user, the systems and the interactions between them [208].

It is important to understand the ability and restrictions that apply when the feet are used as an input modality. Foot and leg movement is mostly performed by the ankle joint, the knee joint and the hip. Table 2.1 shows the range of motion for each joint [165] and corresponding interaction methods and techniques [208] extended by relevant work from the medical domain. In an upright posture, foot movements are restricted to one foot to maintain a

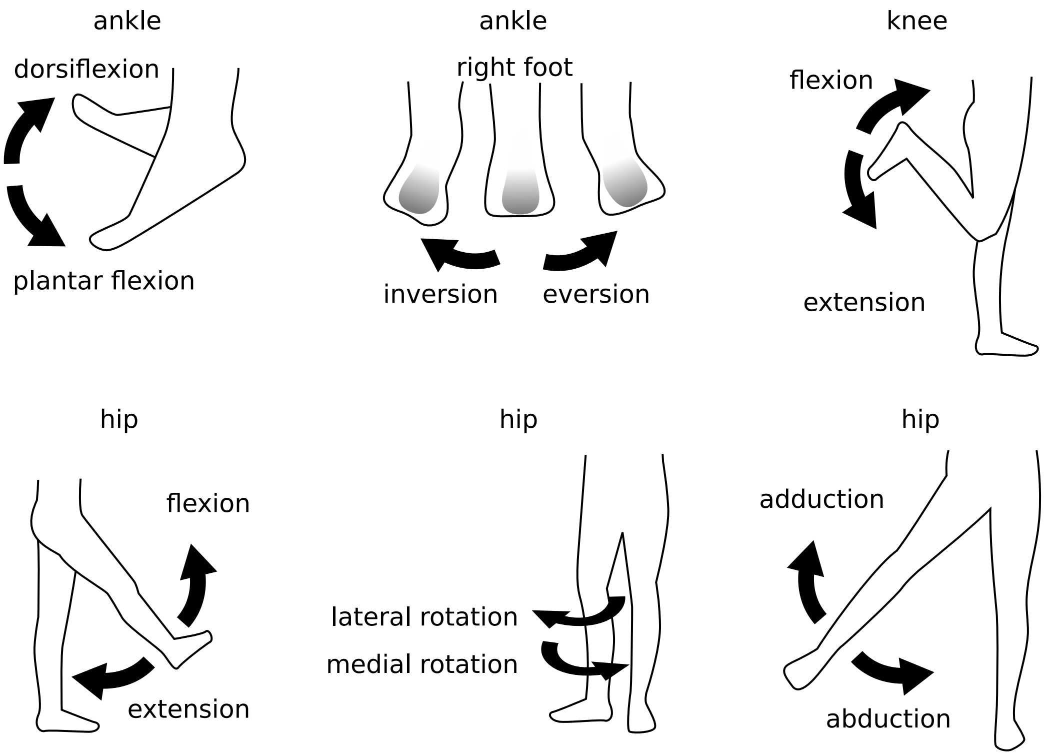


Figure 2.3: Ankle, knee and hip movements. Image adapted from Velloso et al. [208].

stable stance [208].

The feet are a suitable input modality when neither high accuracy nor short execution time is required [148]. Hoffmann investigated accurate movements of hands and feet and found that the feet are about two times slower than hands for visually-controlled movements and about 1.7 times slower for ballistic ones [80]. Pakkanen and Raisamo compared foot to hand input for non-accurate spatial tasks on a computer workplace and found the feet on average to be 1.6 times (2.6 s) slower than the hands and 1.2 times less accurate [148]. Scott et al. investigated the interaction space of four foot movements. For 10° to 40° dorsiflexion, 10° to 60° plantar flexion, −90° to 120° heel & toe rotation, a median target selection error of 11.77° for dorsiflexion, 6.31° for plantar flexion, 8.55° for toe rotation and 8.52° for heel rotation was found [183]. Pointing can be done using the tip of the foot or the position of the toe. However, the position of the ”hotspot” (e.g., the point on the shoe used for pointing) seems to differ substantially between users [12]. Foot taps in a standing posture towards targets arranged around the user in a semicircle yielded higher accuracy for near targets and for arrangements that favored division into columns (i.e. pie-like segments) over rows (i.e. near and far targets) [136].

Velloso et al. investigated unconstrained foot input under the desk for 1D and 2D tasks while seated. They found that moving the feet horizontally is easier than moving them vertically [207]. The preferred strategy to achieve his is rotation on the heel [188, 207]. Heel rotation further is preferred over dorsiflexion, plantarflexion and toe rotation while internal heel rotation is preferred over external rotation and angles below 100° were described as not difficult [183]. For discrete input, such as triggering an event or toggle between states, toe-tapping is considered a suitable method [188] and is preferred over swiping gestures due to balancing problems [94].

Challenges As holding the foot in a specific posture quickly becomes exhausting in unconstrained foot interaction, resting positions are suggested for seating and standing positions [188, 207]. When using conventional foot pedals, keeping the foot on top of the pedal easily leads to accidental activations [18]. In laparoscopic surgery, tools are controlled using foot pedals. When these pedals are identical in shape and size, they easily can be confused, which might lead to undesired consequences such as the wrong potential for ultracision equipment, which might result in cutting instead of coagulating [204, 182]. Wauben et al. surveyed ergonomics in minimally-invasive surgery settings. Foot pedals were found to be uncomfortable by 53% of the 284 participants, which mentioned no visual control, standing at one foot, too many pedals and difficulties in switching the table side [219].

Gaze Interaction

The information where a user is looking at can be used for human-computer interaction. It is a swift input method [218], suitable to gather the user's coarse area of attention [186, 195, 231] and a relatively intuitive method for pointing even though it lacks an activation command [43]. According to Sigut and Sidha [187], technical approaches can be roughly categorized based on the following properties:

- *intrusive/nonintrusive*: Intrusive systems require physical contact with the user, while nonintrusive ones use external cameras.
- *feature/appearance based*: Contours, eye corners or reflections in the eyes are used for gaze estimation in feature-based approaches. The widely used pupil center corneal reflection (PCCR) method is based on this approach, which estimates the gaze point based on the coronal reflection of a *glint* and iris center. Appearance-based ones aim to use image contents to directly map gaze to screen coordinates.

- *2D mapping/3D gaze point estimation*: 2D mapping-based estimation is based on a calibrated gaze mapping function, created using a set of 2D eye movement features. 3D gaze estimation uses a geometrical model of the human eye to calculate the gaze direction, which then can be intersected with the scene.
- *infrared/visible light illumination*: Infrared light allows easier tracking of eye features because it does not distract the user and produces higher contrast than visible light but suffers robustness when used in outdoor scenarios. Visible light, on the other hand, changes frequently, and there are uncontrollable specular reflections. Using IR illumination results in a dark pupil when the light source is placed close to the optical axis of the camera, compared to a dark pupil when placed away from the optical axis [64].

An in-depth analysis of models for eyes and gaze was performed by Hansen and Ji [64].

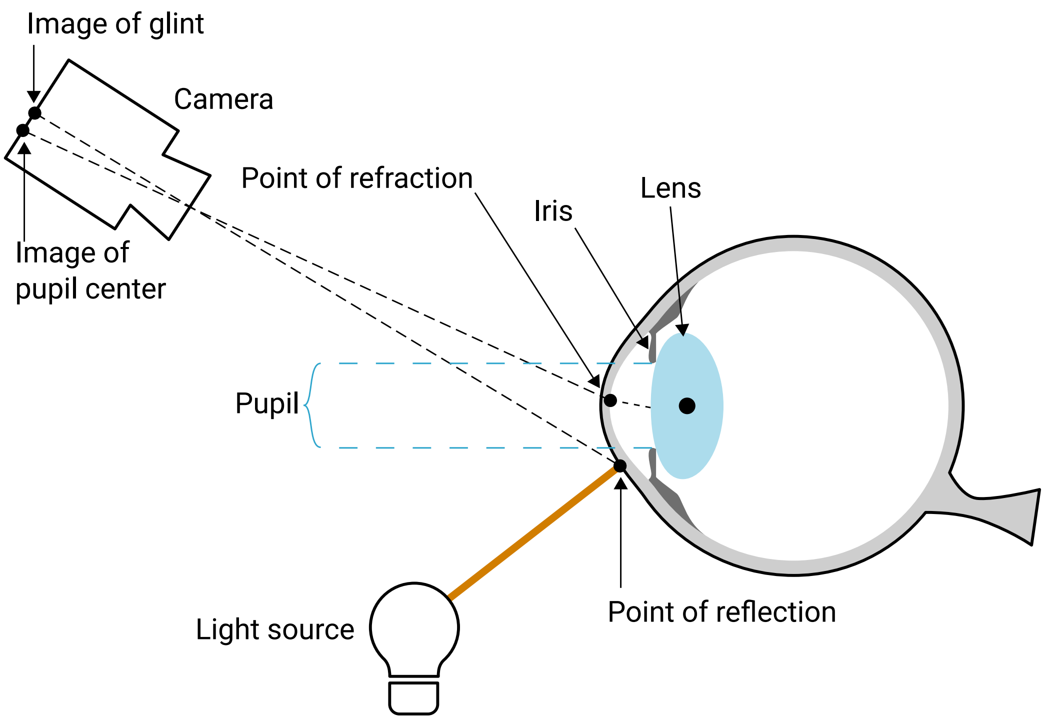


Figure 2.4: Model of the human eye and schematic setup for the pupil center cornea reflection (PCCR) method, including light source, camera and point of reflection ("glint"). Image adapted from Hansen and Ji [64].

Gaze is not very accurate as the size of the part of the fovea with a high spatial resolution restricts measurement accuracy to something between 0.5° [113] and 1.7° [233]. Even though the eyes are relatively stationary during fixations [159], jitter might be caused by inaccuracies during gaze position estimation of eye-tracking systems, which leads to positional accuracy between 0.25° to 1° for good eye-trackers [104]. A proposed approach is having the user view a magnified screen similar to a Fish-eye lens to overcome accuracy limitations [11].

In the medical domain, eye gaze has been used in robotic-assisted minimally-invasive surgery to support the generation of haptic constrains [138], to optimize an ablation path [197], to support 3D surface reconstruction [212], to improve autofocus [34] or to control a laparoscopic camera [50]. Further, functions that are often needed by medical assistants in the OR were made accessible using gaze and auditive feedback [22].

Challenges The eyes lack a method to perform discrete input, similar to a mouse click [43]. Instead, the eyes are ”always on” [91], which leads to the so-called ”Midas Touch” problem when eye gaze is used as input modality [92]. It emerges from the requirement of scanning a scene for visual perception and describes unintended interaction by looking at gaze-sensitive content [92].

A common approach to this problem is a dwell time, which requires the gaze position to stay at a target for a certain amount of time until an action is triggered [43]. For eye typing, dwell times usually range between 450 ms to 1000 ms [229], but shorter dwell times, such as 330 ms [229] or even 282 ms [111], can be used. For menu selection, 750 ms have been reported as the ideal duration for a simple button selection consisting of 400 ms to 500 ms plus extra time for visual search and decision [97]. However, false positives are still possible when looking at an object for longer than the set dwell time [97, 231]. On the other hand, searching for other information during dwell time without terminating a selection process is not possible [97, 231].

Another alternative is to use smooth pursuit eye movements, which only appear when following a target [166]. For this purpose, the similarity between the trajectory of the eye gaze point and the trajectories of all objects on the screen is calculated. The result allows to determine which object is being followed with the eyes [210]. This technique has been investigated for smartwatches [42], in VR [102] and for hand movements instead of eye gaze [29].

Voice Control

Today, voice assistants are available in consumer-level products allowing telephone calls, playing music, controlling smart home devices and many more by uttering a command [81]. As these devices need to listen all the time for a keyword and send user input to servers for processing, privacy is a major concern [81]. Based on Google Home, an application for device size recommendations in interventional radiology was developed [184].

As a touchless input method, voice control can be used for text input or discrete commands [36]. This is especially promising as it bears the potential to control computer systems without assisting staff, which in turn might lead to reduced costs and fewer errors resulting from misunderstandings and lack of experience on the proxy-user end [153]. However, voice input is not equally suited for all kinds of interaction tasks. Cursor control has been realized by mapping vowels to directions for motor-impaired users [65]. This approach has been evaluated using Fitts' law, which is a model of human movement in acquiring or selecting targets and, among other metrics, produces an *index of performance* [46]. The vowel-mapping approach by Harada et al. yielded an index of performance of 0.3 compared to 1.0 for mouse [65]. Utilizing the pitch of the voice for continuous input is possible but using the voice this way is considered unnatural and tires the throat [84]. In a clinical setting, however, voice control was deemed not suitable for continuous parameter adjustment but for discrete commands [141].

The command mode of a speech recognition setup in anesthesia performed better than free speech mode [6]. When using voice for discrete input, commands composed of two words are more intuitive and significant than single-word commands [10]. The recognition rate for typical anesthesia comments depends on the type of microphone and background noise; a headset microphone performed better than a handheld one for louder background noises and other people talking [6]. Perrakis, Hohenberger, and Horbach compared the two voice recognition systems (Siemens Integrated OR System SIOS and Karl Storz Operating room 1 with Storz communication BUS) and found manual control (remote control for SIOS, touch-screen for OR1) faster than voice input for both systems [153]. Combinations of voice input for discrete commands with other input channels, such as gestures that are more suited to continuous input were proposed [141]. Speaking freely during a collaborative discussion in the OR is an important use case that requires alternative input methods to voice. Further, voice commands might be useful to specify a direct, deep link into a Picture Archiving and Communication System (PACS) database [125].

2.2.2 Natural User Interfaces

The described input methods can be used to provide a more natural way of interacting with computers. According to Lee, a ”natural user interface” (NUI) is highly intuitive to a degree at which it becomes invisible during use [109]. Design guidelines for NUI demand to create an experience that feels like an extension of their body to experienced users while at the same time, it provides a natural experience for novices and experts alike [225, p. 13]. Designers of NUI are advised not to start by mimicking the real world or copying existing user interface paradigms. Instead, the experience must be authentic to the medium and considers the context in terms of metaphors, feedback, visual indications and input/output methods [225, p. 13]. It is essential to understand that NUI is not a category of input devices other than keyboard and mouse, but an approach that aims to capture the intent of a user based on its behavior [109]. This can be summarized by interpreting NUI, not as a natural *user interface*, but as a *natural user* interface instead [225, p. 13].

For spatial interaction with 3D content, gestures can be inspired by interaction with real objects [127]. In the medical domain, however, natural manipulation of relatively abstract values might be difficult. Intensity windowing, which, visually speaking, describes the brightness and contrast at which an image is displayed, might serve as an example. In clinical software, both values are often mapped to the mouse axis and modified simultaneously. To build on previous domain knowledge, touchless hand-based approaches applied a similar mapping to hand movements in mid-air [53, 40, 145, 228].

2.3 Multimodal Interfaces

Another approach to improve human-computer interaction is the combination of several modalities. According to Oviatt, multimodal systems process multiple user input modes in a coordinated manner and provide multimedia system output [146]. One of the earliest examples is Bolt’s demonstration of ”Put That There”, which allowed manipulation of different shapes on a large screen using voice commands in combination with pointing gestures [23]. Multimodal input systems might possess several advantages over unimodal systems: Mutual disambiguation allows recovery from unimodal recognition errors when semantically rich input modes are used, multimodal interfaces are expected to be easier to learn and use and allow human-computer interaction in more challenging applications [146]. Recent literature reviews see the potential for multimodal approaches to fulfill clinical requirements for touchless interaction [128] or to provide more reliable cues on whether to ignore or process gestures [36]. Zorzal et al. presented a multimodal interface to support laparoscopic interventions. A video feed and patient imaging data are displayed with a head mounted augmented reality device while image navigation can be performed using head gaze and heel rotation [234].

2.4 Research Methodology

The experiments presented in this thesis aim to investigate the performance of human-computer interaction methods in comparison to existing or alternative approaches. For this reason, quantitative user studies are employed. Qualitative data is gathered using unstructured or semi-structured post-test interviews. This section gives background information on study designs, dependent variables, research methods and tools that are employed in the subsequent chapters.

User studies are the primary research method, which are designed as lab experiments in a controlled setting. In general, lab experiments offer higher internal validity, the precise control of independent variable levels and the possibility to analyze the quantitative data using inferential statistics [144]. In the context of HCI research for the medical domain, this results in the following advantages:

- **Higher control over environmental factors** such as auditory or visual signals, interruptions, interpersonal communication or patient handling.
- **Control over the degree of task abstraction** in terms of required expert knowledge and task complexity.

Due to the high level of control, lab experiments allow a systematic investigation of different input modalities that are pursued in this thesis. This would be difficult to achieve in real medical scenarios as the complexity of procedures and variations in clinical practice yield a high potential for confounding variables.

There are some limitations of controlled lab experiments to be noted: Even though it is an advantage to control external factors regarding internal validity, these factors might influence the performance of an interaction technique or device in a real-world scenario. Experimental research might yield low *external validity* [144]. In the context of HCI in the medical field, simulating external factors to increase the validity of the outcome is a challenging task as they might vary between institutions or professions. Field trials would be the logical next step towards product development but do not take place within the scope of this thesis.

2.4.1 Subject Assignment

User studies follow a between-subject design or a within-subject design. In a between-subject design, each participant is exposed to one condition only. In a within-subject design, each participant is exposed to all experimental conditions [108, p. 49-51]. The advantages of a between-subject design are the lack of carryover effects such as learning effects or fatigue [55, p. 206] [108, p. 51]. Further, an experiment lasts shorter for a single participant if only one condition is to be tested [108, p. 50]. On the downside, individual differences have a stronger impact on a between-subject design, which makes it harder to detect significant differences. A large sample size is required as the same number of participants is needed for each condition [108, p. 50-51].

A within-subject design requires a much smaller sample size than the between-subject design as the performance of the same participants under different conditions is compared. This also reduces the impact of individual differences [108, p. 51]. Disadvantages of the within-subject-design are the possible influence of the aforementioned carryover effects. Learning or practice might occur during one condition and alters the performance in the following one [108, p. 51]. Further, physical or cognitive fatigue might alter the performance in subsequent conditions [55, p. 206]. There are different strategies to reduce the impact of these effects. A learning effect can be reduced by providing sufficient time for training as the learning curve tends to be steeper during initial interaction for many kinds of tasks [108, p. 55]. Another approach to control carryover effects is by counterbalancing, which ensures that each condition appears equally often each time. Complete counterbalancing requires all possible combinations of condition orders to be calculated and is only feasible for small numbers of conditions. As a practical compromise, latin-square designs ensure that each condition occurs equally often in each position but these are incomplete counterbalancing arrangements [55, p. 206][139, p. 245-246]. A better solution than a standard latin square is a balanced latin square design, where not only the number of appearances but also the preceding condition is balanced [55, p. 206].

2.4.2 Dependent Variables

The goal of a user study is to find out in which way dependent variables are connected to independent ones. Independent variables are the factors that are controlled by the researcher. Dependent variables are the outcomes the researchers are interested in, which might be the time required to complete a task or a measure for the difficulty of the task. ”Dependent” in this context means that this variable depends on the participant’s behavior or the independent variable [108, p. 30]. In human-computer interaction, independent variables may be an input device or input technique, while dependent variables are factors such as the required time or usability. Any factor which is not an independent variable and may influence the dependent variables is called a confounding variable and should be kept the same for all conditions [108, p. 39]. In the following, an overview of the most commonly used dependent variables in this work and the tools used to measure them is given. Less frequently used dependent variables or tools are described in the corresponding sections.

Time Performance A way of measuring performance is to record the time it takes a participant to complete a particular task. This includes all the processes during the measurement, such as cognitive processing or motor task performance. Time measures might be gathered and saved automatically or recorded manually during a user study or afterward using video logs.

Subjective Workload The term ”workload” describes the cost that is required to accomplish a task [67]. A widely used tool for subjective workload ratings is the NASA task load index (NASA-TLX) questionnaire [226]. It consists of the six subscales *Mental Demand*, *Physical Demand*, *Temporal Demand*, *Frustration*, *Effort*, and *Performance* [67] and was developed by Hart and Staveland [68].

The application of the original NASA-TLX questionnaire consists of two phases. A weighting scheme aims to account for the differences in the contribution of each dimension to the personal definition of workload. This is determined by selecting one out of each possible pair of subscales and calculating a weight for each subscale. In the second phase, the actual rating of one or more given conditions on the six subscales takes place one a scale from 0 to 100 with tick marks every five units. The scales are arranged in a way that the negative end of the scale is always on the left and the positive one on the right. The overall workload is then calculated from the weighted results of the subscale ratings [68]. The most common modifications of the NASA-TLX questionnaire is to eliminate the weighting process and to analyze the subscales individually, which has been referred to as Raw TLX (RTLX) [67]. Omitting the weighting process makes it simpler and faster to apply, while there seems to be no clear evidence for a negative impact on validity or sensitivity [67, 226]. Analyzing subscales can help to pinpoint problems [67] and might reveal differences that would level out in the overall workload.

There are, however, limitations to the NASA-TLX. McKendrick and Cherry found that this tool might be sensitive to task demands but lacks sensitivity to personal capacities [122]. Winter listed four persistent open questions: The lack of a recommendation regarding subscore weighting, the tickmarks in the paper-and-pencil which are prone to be misinterpreted for checkboxes, the easily confusing anchors ”good” to ”bad” for ”own performance”, which might be interpreted as inverted, and two existing versions with the different anchors ”high/low” and ”very high/very low” [226].

Another issue is the scale on which NASA-TLX and RTLX are assessed and scored, which might lead to misinterpretation of results. The original NASA-TLX questionnaire requires the tick marks representing five units to be counted, one subtracted and multiplied by five to get a value from 0 to 100. When scoring is omitted, results might be reported without scoring on a scale from 0 to 20 or from 1 to 21, representing the tick marks. Such results might be misinterpreted easily as truncated 0-100 point scales even though truncating axis in bar charts is considered a distortion technique [149]. This should be avoided by providing a description of the minimium and maximum value. In this dissertation, only RTLX is used and the results are reported on a scale from 0 to 20.

Usability and User Experience According to the International Organization for Standardization (ISO), the Term ”Usability” describes the ”extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [85]. A tool for measuring subjective usability is the System Usability Scale (SUS) [27]. Even though there are several alternatives such as the ASQ, PSSUQ, SUMI or QUIS [16, 178], the SUS questionnaire has several positive attributes: It seems to be the most sensitive post-study questionnaire [178, p. 232], is technology agnostic, relatively quick and easy to use and to interpret [16]. The questionnaire consists of ten questions to be answered on a 5-point Likert scale. After scoring, the SUS yields an overall system usability value between 0 and 100 [27]. The meaning of SUS scores has been investigated in relation to adjective ratings on a 7-point scale ranging from ”worst imaginable” to ”best imaginable”, which revealed SUS scores at around 50 to match the adjective ”OK” and at around 70 to ”Good” [15] In terms of acceptability, a score below 50 can be considered not acceptable, above 70 acceptable and in between as marginally acceptable [16].

User Experience as defined by the ISO is the ”user’s perceptions and responses that result from the use and/or anticipated use of a system, product or service” [85]. AttrakDiff [69], UEQ [107], and meCUE [132, 124] are the most recognized standardized questionnaires for evaluating User Experience [37]. In contrast to other questionnaires, meCUE addresses all central user experience (UX) components in a unified way. Module I ”product perception” assesses usefulness, usability, visual aesthetics, status and commitment, module II ”user emotions” covers positive and negative emotions and module III ”consequences of use” relates to product loyalty and intention to use. Additionally, module IV allows an overall evaluation. As the four modules were validated separately, they can be combined to fit different kinds of research questions [132].

3

Touchless Gesture Interaction for MRI-guided Interventions

In this chapter, a direct, sterile input method for controlling an MRI scanner is presented and compared to proxy-user interaction, which is a common method during MRI-guided interventions in clinical practice. The proposed approach is demonstrated for MRI-guided percutaneous liver tumour ablation following the general workflow described in [168]. The human-computer interaction methods, however, can be applied to different concrete tasks, which makes the overall concept adaptable for a broader range of MRI-guided procedures.

Compared to US or CT, MRI images provide better soft-tissue contrast [168, 44] and therefore allows better visualization of the targeted tumor, surrounding tissue and adjacent risk structures such as blood vessels [222]. Further, MRI allows multiplanar image acquisition, which means that the plane of the image to be acquired can be placed arbitrarily instead of being bound to the orientation of the imaging device. Despite these advantages, MRI has not yet become the standard imaging modality for percutaneous interventions. The reasons lie not only in the higher costs, but in the strong magnetic field which requires specialized hardware [17], and the current focus on diagnostic use cases. Sufficient control over the MRI scanner can only be achieved by using non-sterile input devices in the control room. A common workaround for interventional scenarios is the operation of the MRI by an assistant on the radiologist's demand. This method suffers from high noise levels during scanner operation [121] and causes delays when the radiologist needs to leave the scanner room to specify instructions [169].

3.1 Direct Control of MRI Scanners

In addition to the limitations in a sterile environment described in Section 2.1.2, touchless direct input methods for MRI scanner control require hardware that can cope with the strong magnetic field inside the scanner room. Therefore interactive image plane adjustment was realized using mechanical manipulators [32], a wireless active tracking device [163], in-bore optical moiré phase tracking [96], or handheld devices activated by foot pedals [120]. Approaches that allowed access to a wider range of functionalities use optical detection of hand gestures [61] or tablet PCs [169].

A proof of concept for touchless control of an MRI scanner via hand gestures was presented by Güttler [61]. The number of extended fingers and the position of the hand was detected by an MRI-compatible RGB camera. The approach allowed moving the field of view (FoV), tilting the image acquisition planes by 90°, and starting image acquisition [61]. The system used the manufacturers MRI user interface and did not include a graphical user interface adapted for gesture input. A UI to directly controlling the MRI scanner from within the scanner room was developed by Rube et al. The goal of their approach was to reduce delays caused by the need to leave the scanner room for communication and image viewing [169]. A web-based UI was available on a tablet PC mounted on a pole next to the MRI scanner. For evaluation purposes, cadaver experiments with a novice (physician in training) and an expert (radiologist with ten years of experience) were performed. They found mean puncture times comparable to literature [169, 193, 45, 168]. Even though direct interaction is improved with this concept, touching surfaces is not optimal as it increases the risk of break asepsis and should be avoided. Further, sterile drapes reduce image quality and might cause interaction errors [77].

3.2 Requirement Analysis

An online survey among intentional radiologists was conducted to understand the current clinical method for scanner control and the needs of radiologists when performing MRI-guided interventions.

The survey aimed to find answers on two questions: what is common practice for MRI-guided interventions and which methods and functions are deemed useful? Therefore, an online survey with four questions was conducted. The first three questions assessed clinical practice by asking in which way MRI control is achieved and how often certain functions are used pre- and interventionally. The last question asked for a rating of the potential usefulness of the same functions during interventions. The answer options were created in cooperation with clinical partners from the Hannover Medical School (Hannover, Germany). To make sure as most practical approaches as possible were covered within the survey, the workflow of an MRI-guided percutaneous liver tumour ablations was discussed with an interventional radiologist with a focus on alternative HCI methods. The resulting answer options regarding the current method for MRI control were:

- Delegate tasks to an assistant verbally
- Use the workstation in the control room
- Use a trackball/MRI-safe terminal
- Other

Further, the functions to be rated in terms of frequency of use **preinterventionally**, **intrainterventionally** and **potential intrainterventional usefulness** are described in the following. The term ”sequence”, in the context of MR imaging, describes combinations of radiofrequency pulses and gradient switching schemes which determine contrast and resolution of the acquired images [135].

- F1: Windowing
- F2: Change sequence
- F3: Start/stop sequence
- F4: Change sequence parameters
- F5: Translate image plane
- F6: Rotate image plane
- F7: Switch between planes that are parallel or orthogonal to the needle
- F8: Switch between sagittal, coronal and axial plane
- F9: Show planning data sets (including tumour in 3D, entry point, planned trajectory, target)

Eleven radiologists took part in the survey. The participants had 14.6 years (SD = 8.4 years) experience with needle interventions and 13.4 years (SD = 7.9 years) with MRI-guided interventions on average. The most common method for controlling the MRI scanner is by verbally delegating a task to an assistant, as can be seen in Figure 3.1. Less used is the workstation in the control room (5), an MRI-safe terminal (4) or other means (2). For this question, multiple answers were possible.

Translation of the image plane (F5) was reported to be used preinterventionally ”always” or ”often” eight times (see Figure 3.2). The most used functions during interventions are starting and stopping sequences (F3) as well as translation of the image plane (F5), which was reported to be used ”always” or ”often” by 10 out of 11 participants (see Figure 3.3). The least used functionality was changing sequence parameters (F4). It was reported to be used ”never” or ”rarely” five times for the preinterventional phase and six times interventionally. The same functionality was deemed to be the least useful during interventions being rated ”useless” or ”somewhat useless” by

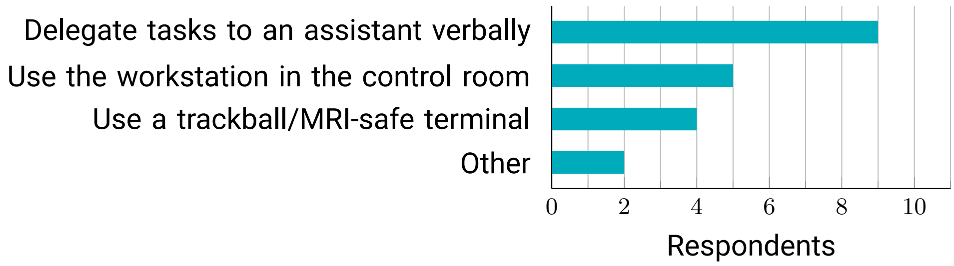


Figure 3.1: Survey results on "How do you currently control the MRI scanner?". Multiple selections were possible.

three participants and only rated as "useful" four times. The reported rare usage of changing sequence parameters indicates that there is no need for elaborate adjustments of image sequences when performing an intervention. On the question of how useful certain functions would be, translating the image plane (F5) was rated as "useful" by all participants, starting and stopping sequences was rated "useful" the second most (see Figure 3.4). Functionalities that allow switching between image planes (F7, F8) were always rated as "somewhat useful" or "useful".

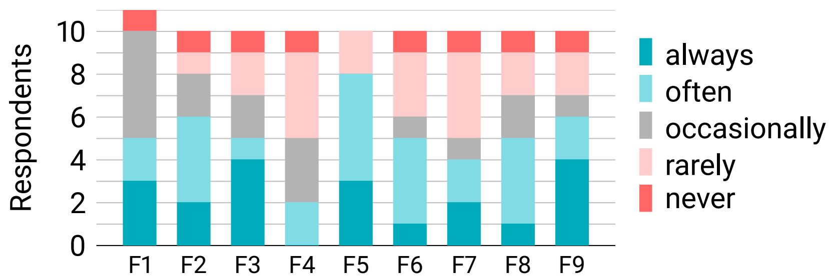


Figure 3.2: Survey results on "How often do you use the following functions **preinterventionally?**"

Further, the workflow of an MRI-guided needle intervention was discussed with clinical partners and found similar to approaches described in literature [168, 222, 44].

3.3 User Interface

A set of functions was defined based on the results of the preliminary survey and the observation of an MRI-guided intervention. It should be possible to start and stop sequences (F3) and to translate the image plane (F5), as these functions were most often rated as "useful" regarding their potential

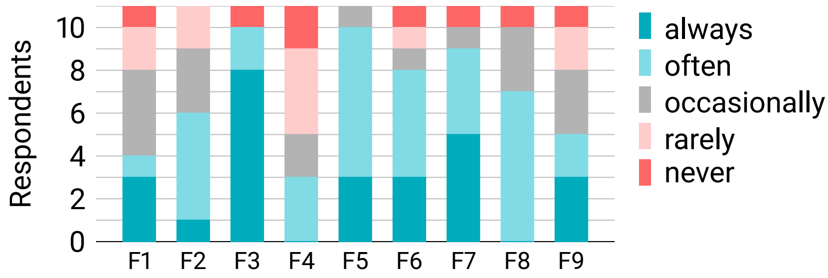


Figure 3.3: Survey results on "How often do you use the following functions **interventionally?**"

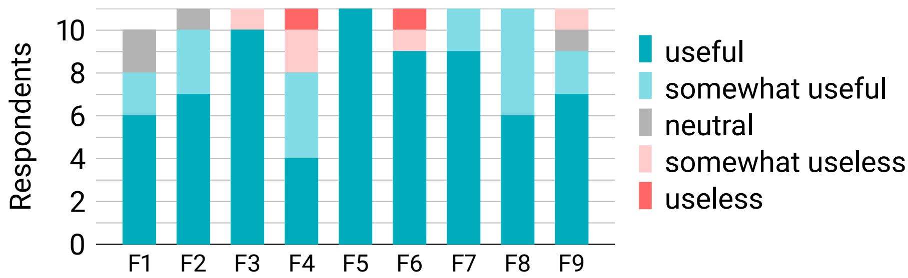


Figure 3.4: Survey results on "How **useful** do you think the following functions would be during interventions?"

usefullness during MRI-guided needle interventions in the online survey (see Figure 3.4). Further, sequence selection and windowing have to be accessible to support the workflow described by clinical partners.

These functions were included in an interventional user interface. It is based on two ideas: First, a suitable input modality and a corresponding input method for direct human-computer interaction under sterile conditions have to be found. Second, the graphical interface has to be adapted to the input method to the reduced set of functionalities and the interventional workflow. Design considerations regarding both parts are presented in the following.

3.3.1 Input Modality and Method

Inside the MRI scanner room, communication is difficult due to the high noise levels during scanner operation [169]. Special solutions for audio communications have to be employed [60] or imaging sequences have to be stopped or paused for verbal communication and commands. A method from clinical practice to signal such a break is to use hand gestures. As the needle can be kept in place with one hand, the other one can be spared for signals and gestures. Therefore, the input modality of choice is the hand, specifically one-handed touchless gestures. To make the desired functions accessible via hand gestures, two types of gestures were employed: The *deictic* (pointing) gesture of an extended index finger is used to control a mouse cursor. Discrete commands such as a selection of a sequence can be realized as buttons.

Further, prior knowledge of windowing with a conventional computer mouse can be leveraged. The finger position is mapped to absolute cursor positions. Touching the second joint of the middle finger with the thumb corresponds to pressing a mouse button (see Figure 3.5b). This method is derived from Mewes et al. [127]. A *manipulative* gesture is derived from an *iconic* (illustrating) one for the interactive movement of the image acquisition plane. By doing a flat hand, an initial position is set, symboling the image plane. Moving the flat hand to the front or back translates the image acquisition plane. The rate of translation depends on the distance of the hand to the position where the flat hand gesture was initially performed. By changing the hand gesture or leaving the sensor area, the translation can be stopped.

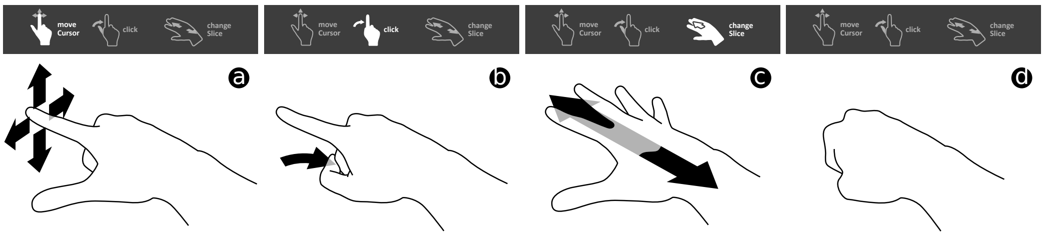


Figure 3.5: Hand gesture set for touchless MRI-control: cursor control (a), selection (b), moving the image acquisition plane (c), no interaction (d). A icon bar at the top of the user interface indicates the currently detected gesture. Above each gesture, the corresponding icon is highlighted. The icon for gesture (c) has extra states indicating forward, backward or no movement of the acquisition plane. Image from Hatscher et al. [74] © Springer Nature 2019[1].

3.3.2 Optimized Graphical Interface

During the intervention, real-time images are displayed on the MR-safe in-room monitor. In clinical practice, the manufacturer's software optimized for diagnostic use cases and mouse input is displayed. Therefore, the graphical

[1]Adapted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, International Journal of Computer Assisted Radiology and Surgery, Touchless scanner control to support MRI-guided interventions, B. Hatscher et al., © Springer Nature 2019.

user interface had to be adjusted to make the reduced set of functions gathered during requirement analysis accessible using the hand gestures described in the previous section. As prior knowledge should be applicable, live medical image data is presented as a tiled view of three, similar to the MRI manufacturer's software. Contrary to a diagnostic use case, higher image acquisition rates are required, which naturally leads to a smaller set of suitable sequences. Three preselected MRI-sequences are provided, each in two variants, favoring either speed or image quality. Each sequence, as well as start, stop and pause, can be triggered using large buttons, accounting for the lower accuracy of mid-air hand gestures compared to mouse input [13]. Further, visual feedback on detected hand gestures is given. The graphical interface can be seen in Figure 3.6.

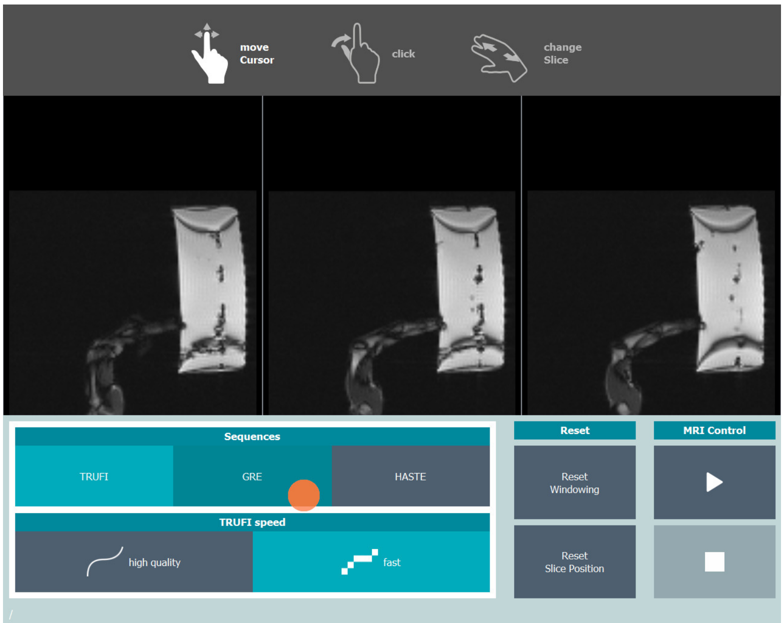


Figure 3.6: The prototypical graphical user interface (GUI) tailored to the interventional requirements of direct, touchless MRI control. The bar at the bottom provides a set of imaging sequences (left) and MRI control elements (right). At the top, visual feedback for hand gesture recognition is provided. In this screenshot, three parallel oriented images are acquired and displayed while the finger tipping method [44, 48] is performed to find a needle entry point. Image from Hatscher et al. [74] © Springer Nature 2019[2].

3.4 Evaluation

A user study was conducted to evaluate the proposed user interface. The *gesture input* approach was compared to *task delegation*, which is the most used input method in clinical practice according to the requirement analysis (see Figure 3.1). The study only considered physician-computer interaction inside the scanner room. Workflow steps that could be carried out on a conventional workstation, such as trajectory planning, were left out.

3.4.1 Participants

Ten radiologists (six male, four female) at an age between 27 and 50 years (M = 33.4, SD = 7.0) took part in the study. The average experience with MRI-guided interventions was 0.35 years (MIN = 0, MAX = 2, SD = 0.75), experience with needle interventions ranged from 0 to 18 years (M = 4.4, SD = 6.1), mostly with CT-guided interventions (M = 3.8, SD = 6.3). All participants were right-handed.

3.4.2 Apparatus

For hand gesture detection, a Leap Motion Controller was used. It tracks hand and finger positions in an area of approximately $0.5m \times 0.5m \times 0.5m$ using an infrared stereo camera. The device was modified to work reliable in the strong magnetic field of an MRI scanner and not to influence the image quality [150]. As an unshielded USB cable leading from the gesture sensor to a computer outside the scanner room degrades image quality, the data flow was converted to optical signals and transmitted via a fiberoptical cable. The sensor was initially powered by USB, which required batteries and a charge controller to be added on the sensor-side of the fiberoptical cable. The batteries provide power for approximately five hours. During imaging, voltage spikes induced on the line can cause the electronic components of the gesture sensor inside the scanner room to shut down. A surge filter (Würth Electronics, Niederhall, Germany) was installed before the sensor to keep the voltage below a safe threshold. An aluminum housing contained the surge filter, USB 3 to USB 2 converter, power supply, and USB to fiber-optic cable converter. A shielded USB cable connected the gesture sensor and the aluminum box. The MRI compatible technical setup is shown in Figure 3.7.

[2] Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, International Journal of Computer Assisted Radiology and Surgery, Touchless scanner control to support MRI-guided interventions, B. Hatscher et al., © Springer Nature 2019.

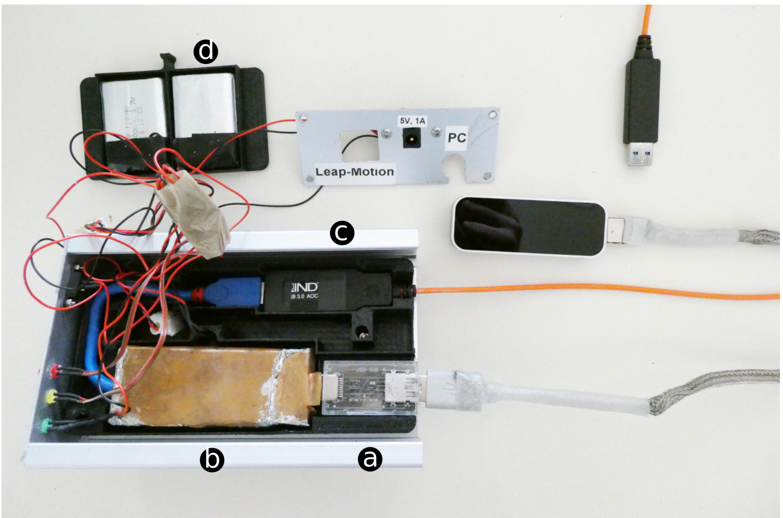


Figure 3.7: The modified gesture sensor made MR compatible by adding a surge filter (a), a shielded USB 3 to USB 2 converter (b), a converter from USB to fiber-optic cable (c) and batteries (d) as power supply. Image from Hatscher et al. [74] © Springer Nature 2019[3].

A prototypical software was developed to realize the optimized graphical user interface described in Section 3.3.2 and control over the MRI scanner. The software uses the Qt application framework (Qt Group, Helsinki, Finland). DICOM images are processed with the Grassroots DICOM library [114] and displayed using the visualization toolkit [181]. The commercially available *Scanner Remote Control* (SRC) plugin (Siemens Healthcare GmbH, Erlangen, Germany) provides a REST interface that grants external client applications direct control over the MRI scanner. For safety reasons, the manufacturer's workstation can stop running sequences and take over control at any time. Further, the API was used for logging all scanner-related events for duration measures.

A 1.5 T MRI scanner (MAGNETOM Aera, Siemens Healthcare GmbH, Erlangen, Germany) controlled by a workstation running the Siemens Dot Cockpit Software was used during the study. A plastic container filled with candle gel and three O-rings near the bottom served as phantom. It was placed inside the bore of the MRI scanner. The gesture sensor and the aluminum box

[3]Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, International Journal of Computer Assisted Radiology and Surgery, Touchless scanner control to support MRI-guided interventions, B. Hatscher et al., © Springer Nature 2019.

containing the components to make the sensor MRI compatible were placed at the side of the patient table. The sensor was connected via fiber-optical cable to a PC (SZ270R9, Shuttle Computer Group, City of Industry, CA, USA) outside the scanner room. The PC ran the prototypical software and was connected to the MRI host via ethernet. Depending on the input method, the manufacturer's workstation (task delegation) or the PC (gesture input) were connected to the MR compatible monitor next to the MRI scanner inside the scanner room.

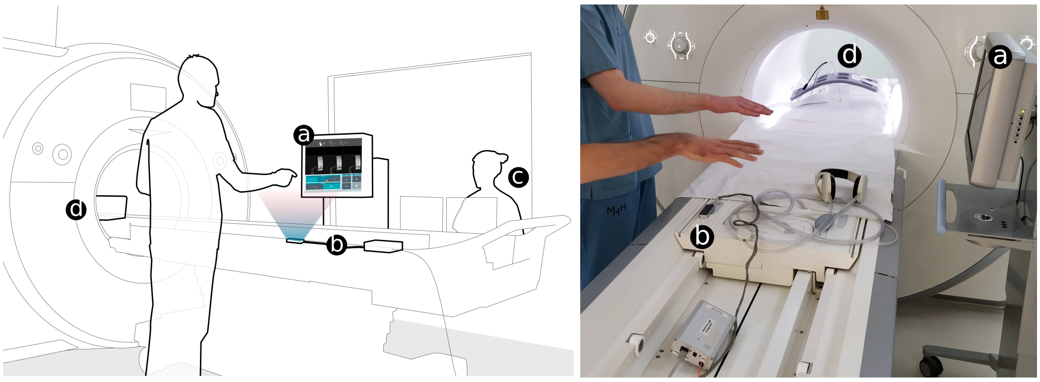


Figure 3.8: Study setup consisting of an MR compatible monitor (a), an MR compatible gesture controller (b), an assistant at the workstation in the control room (c) and a phantom inside the MRI scanner (d). Image from Hatscher et al. [74] © Springer Nature 2019[4].

3.4.3 Study Design

The study followed a within-subject design. Six tasks were performed with both of the input modalities *gesture input* and *task delegation*. The order of input methods was randomized to minimize carryover effects.

3.4.4 Tasks

The tasks to be performed resembled the physician-computer interaction workflow that appears during an MRI-guided percutaneous intervention:

- *Start* sequence

[4]Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, International Journal of Computer Assisted Radiology and Surgery, Touchless scanner control to support MRI-guided interventions, B. Hatscher et al., © Springer Nature 2019.

- *Move* slice position
- Change *window width* and *window level* to given values
- Insert *needle* into the gel phantom and hit the target

As starting a sequence was expected to take only a few seconds, the first task was executed three times per input modality to reduce the influence of varying reaction times. To perform the *start* task, the following steps had to be taken: stop the currently running sequence, select the given sequence and start it. The *move* task required the user to translate the displayed slice group 50 mm along the left-right-axis to a predefined position. To fulfill the *window* task, the window width and level needed to be set to a target value with a tolerance of 10 for width and 60 for level. Performing the *needle* task required no interaction with the MRI scanner but the insertion of a needle into the phantom guided by real-time MRI images. The last task completed the workflow, which was aimed to support less experienced participants in comparing both input modalities in the context of an intervention. The order of tasks was fixed.

The input modalities required different execution of the tasks. *Gesture input* allowed to perform all tasks directly. *Task delegation* required the user to attract the assistant's attention, who was sitting in the control room. The assistant then stopped the currently running sequence to allow verbal communication. The sequence to be started was communicated by the participant via the intercom. The assistant confirmed the command and performed the required steps using the Siemens Dot Cockpit software. The assistant knew the workflow beforehand but awaited each command to be spoken out loud completely to simulate the behavior of an experienced medical technician in a well-rehearsed team. "Left", "right", "stop" and "more" were the only commands to be used while performing the *move* task. The restriction to this set of commands was to avoid explicit instructions such as "go to position X" and instead simulate the situation during an intervention when the exact coordinate of the target slice is not known. The participant had to leave the scanner room for the *window* task and instruct the assistant at the workstation. For the same reason as in the *move* task, only "wider" and "narrower" were allowed for window width adjustment and "brighter" and "darker" for adjusting the windowing level.

3.4.5 Measures

As a measure of performance, task completion time (TCT) was gathered. The subjective workload was assessed with the RTLX [67]. With this tool, the subjective perception of mental demand, physical demand, temporal demand, performance, effort and frustration are assessed. The answers are given on a scale with 21 tickmarks ranging from ”low” to ”high” (”good” to ”poor” for performance, respectively). Subjective usability was measured using the SUS, which consists of ten questions that can be rated from ”strongly agree” to ”strongly disagree” on a 5-point Likert scale [27]. A score between 0 for the lowest and 100 for the highest subjective usability can then be calculated [27]. At the end of test, a structured interview with the following questions was conducted:

- What was particularly noticeable?
- How would the input method affect your daily work?
- What are the biggest advantages, disadvantages or differences compared to existing workflows?
- Where do you see potential for improvement?

3.4.6 Procedure

The user study was conducted in an MRI suite of the Hannover Medical School (Hannover, Germany) during off-hours by two study observers. One instructed the participants in the scanner room, while the second one remained in the control room. The observer in the control room operated study-related software parts (task selection, logging) and acted as assistant at the workstation during the *task delegation* modality. All tasks were known beforehand to the observer at the workstation. To provide equal conditions for all participants and to simulate the behavior of an experienced medical technician, every command was executed without delay at the same pace, significantly determined by the speed of the workstations default software. For each participant, a phantom was prepared with the needle already placed at the entry point.

Initially, a demographic questionnaire was filled out by the participant. Next, the overall procedure and the required workflow steps were described by an instructor. Similar to an interventional setting, the participant stood next to the bore of the MRI scanner (see Figure 3.8). For both input modalities, the same protocol was applied: The first of the two input methods was explained and demonstrated by the instructor at the MRI scanner. Every interaction required to fulfill the tasks was shown and upcoming questions were clarified. The participant was given time to become familiar with the *gesture input* method. Even though there was no time limit, the mean training time was 360 s (SD = 117 s). After that, the instructor explained in detail the six tasks to be performed in sequence. The beginning and end of a task were conveyed by a short visual signal towards the observer in the control room.

After finishing all tasks, the participant filled out the RTLX questionnaire and a SUS questionnaire. The whole procedure was then repeated for the second input method. After finishing both input modalities, the post-test interview was conducted.

3.5 Results

All tasks were completed successfully by all participants. The log files from the first participant were not recorded due to a software error. Therefore, the task durations of only nine users were gathered, but the qualitative measurements (RTLX, SUS, interviews) contain data from ten participants.

Task completion time for each task is shown in Figure 3.9. No difference between input modalities was found for the *start* task and the *needle* task. It has to be noted that no software interaction was required during *needle* insertion, hence no difference between input modalities was expected for this task. Performing slice translation during the *move* task was slightly slower with *gesture input* (M = 43.9 s, SD = 25.3 s) than delegating the task (M = 39.1 s, SD = 24.7 s). *Windowing* was faster when *gesture input* was used, but task completion times deviated stronger (M = 72.5 s, SD = 44.5 s vs. M = 80.8 s, SD = 11.2 s).

For subjective usability, the average SUS score for hand *gesture input* is 74.8 (SD = 14.9) and 58.2 (SD = 25.4) for *task delegation* (see Figure 3.10).

RTLX scores are shown in Figure 3.11. Comparable results for both modalities were found for the overall score, perceived mental and physical demand, effort and frustration. Participants perceived their performance lower with *gesture input* (9.3 s, SD = 4.7 s than with *task delegation* (M = 5.2 s, SD = 3.6 s). Temporal demand is perceived higher for *task delegation* (M = 6.6 s, SD = 4.9 s) than for *gesture input* (M = 5.2 s, SD = 4.0 s). With a mean system usability score of 74.8 (SD = 14.9), the usability of *gesture input* was rated higher than the one of *task delegation* (M = 58.3, SD = 25.4).

In the following, results from the post-test interviews are reported. The interviews were analyzed and similar statements were clustered. Direct control was considered advantageous over task delegation by nine of the ten participants. *P8* thought it is more predictable than an assistant, it allows decision making more freely *(P4)* and grants a higher level of control *(P9, P10)*. *P3* explained that interventional radiologists are used to directly

[5] Adapted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, International Journal of Computer Assisted Radiology and Surgery, Touchless scanner control to support MRI-guided interventions, B. Hatscher et al., © Springer Nature 2019.

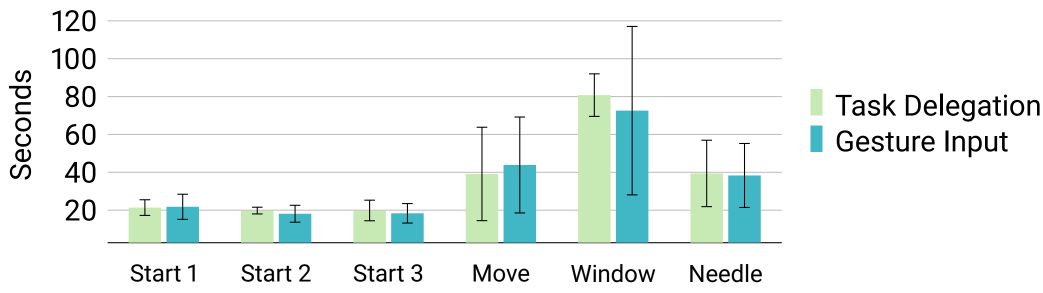


Figure 3.9: Task completion times for all six tasks including the *start* of sequences, *move* slice position, change of *window* width and level and *needle* insertion. Error bars represent standard deviation. Figure adapted from Hatscher et al. [74] © Springer Nature 2019[5].

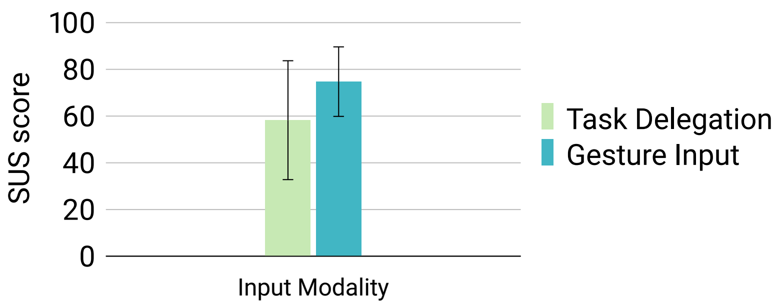


Figure 3.10: Mean score from the SUS questionnaire. Error bars represent standard deviation. Higher scores depicts better subjective usability. Figure adapted from Hatscher et al. [74] © Springer Nature 2019[5].

control the table movement of a CT angiography suite without assistance, thus the *gesture input* approach is more familiar. *P10* would prefer a gesture-controlled system to the use of extensive hand gestures for communication with the assistant in front of a conscious patient to maintain a professional atmosphere. It was assumed that using *gesture input* saves time for different reasons. There is no need to leave the scanner room or interrupt a task *(P4)* to additionally verbalize the desired input *(P5)* or await the assistant's response *(P7)*. Drawbacks of proxy user interaction that might be overcome by *gesture input* were mentioned by four participants. They include having an assistant without much experience *(P1)*, depending on someone else *(P4)*, relying on error-prone verbal communication *(P6, P8)* or lacking the necessary precision *(P8)*.

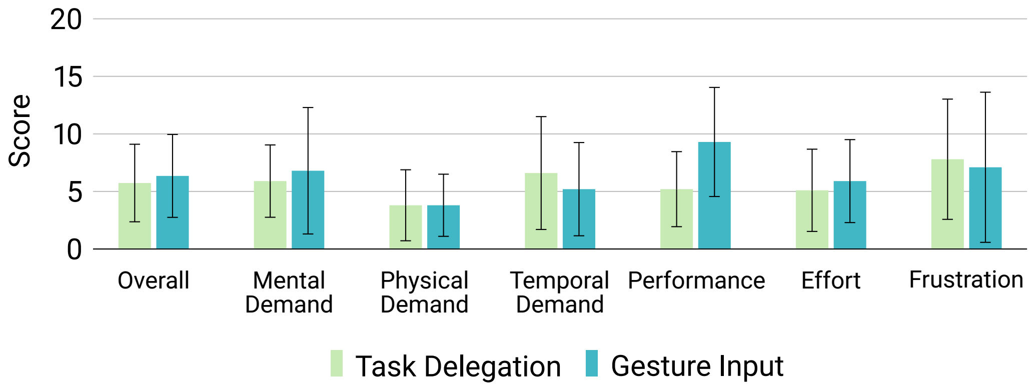


Figure 3.11: Unweighted mean scores for the NASA-TLX questionnaire on a scale from 0 to 20 (0 = low/good performance, 20 = high/poor performance). Error bars represent standard deviation. The overall workload is the mean value from all the subscales. Figure adapted from Hatscher et al. [74] © Springer Nature 2019[5].

Concerns could be categorized into three topics regarding the gesture set, the working area and the implications on the established workflow. For *gesture input*, three participants found it difficult to use different planes in 3D space. Using the same plane for windowing and image plane adjustment was deemed easier *(P3)* and gestures more similar to mouse movements on the table were suggested *(P4, P7)*. Three participants missed an indicator for the gesture sensor's FoV even though visual feedback for the currently detected gesture was given. Fatigue in the raised arm during gesture interaction was reported *(P8)*. Three participants raised concerns regarding integration in the current workflow and working environment. A different sensor placement in front of the physician rather than beside the patient was suggested *(P3)*. Well-timed preparation of supporting tasks might get difficult when the assistants outside the scanner room are not kept updated. Further, the procedure is hard to follow for staff, observers or students *(P5)*. Gesture interaction was considered a potential source of distraction during needle placement *(P7)*.

3.6 Discussion

In terms of task completion time and subjective workload, hand *gesture input* performed comparable to *task delegation* (see Figure 3.9 and 3.11). For the *move* task, slightly higher durations were measured for *gesture input* than for *task delegation*. One reason might be that the target position was on the left of the patient coordinate system, while the starting position was on the right, which was easily misinterpreted. When using direct input, that mistake was hard to spot while verbal communication and confirmation might have forced the participant to double-check the direction of movement. However, slice positions were expressed the same way as in the manufacturer's software; thus, prior knowledge should be applicable. Subjective performance was rated lower for *gesture input* than for *task delegation*, but the objective average task completion time does not show such a difference. This could mean that participants felt like there is potential to perform better with *gesture input*.

According to Bangor, Kortum, and Miller, SUS scores can be interpreted in terms of acceptability [16] and as adjective ratings [15]. Acceptability includes "not acceptable", "low marginal", "high marginal" and "acceptable" [16]. On this scale, the rating for *gesture input* corresponds to "acceptable" while *task delegation* falls in the range of "low marginal acceptability". Adjective rating categories are "worst imaginable", "poor", "ok", "good" and "excellent" [15]. The score for *gesture input* falls between "good" and "excellent", while *task delegation* falls between "ok" and "good". These results indicate that direct, touchless control of an MRI scanner increases the usability compared to the clinically common task delegation method.

In clinical practice, various factors come into play, which are difficult to replicate in a user study. *task delegation* depends on the assistant's experience and knowledge. As there is no standardized way of communication or vocabulary, it depends on the physician's preference. In the worst case, the assistant's attention is taken by other events, rendering interaction temporarily impossible. Further, comparing results with literature is difficult as similar study designs produce different outcomes [78, 227]. This might be caused by different implementations of task delegation. The position of the gesture sensor was fixed during the study. For clinical practice, the sensor placement should be more flexible as the available space in the bore and at the table, as well as preferences of the performing physician, may vary. Involuntary input happened when the participant did not pay close attention to the visual feedback on gesture recognition. This led to situations where the "virtual hold" was not yet released, but the hand was already moving out of the sensor area. A method to activate the system or "clutching mechanism" might overcome this drawback of touchless input, especially when contextual cues or multiple modalities are used [36].

3.7 Conclusion

In this chapter, a touchless user interface for direct, sterile input in an interventional setting is presented and evaluated. Clinically relevant functionalities to be accessible during interventions were gathered by discussion with clinical partners and conduction of an online survey among international experts. Based on the requirements, a user interface tailored to the needs of MRI-guided percutaneous needle interventions was created. The interface consisted of a hand gesture set derived from [127] and a customized GUI suitable for touchless gestures. The proposed approach was compared to the clinically common method of task delegation in a user study. The results revealed comparable subjective workload and task completion times but yielded higher usability for direct input. Further, physicians saw the potential of direct control over a system as it is deemed to be less interrupting and more independent from other team members.

This chapter shows that direct physician-computer interaction is advantageous over the common workaround of delegating tasks to a proxy user. However, the presented method is tailored to support needle guidance during MRI-guided percutaneous interventions. In terms of human-computer interaction, the medical scenario rules out voice commands due to high noise levels during image acquisition, which might be a promising approach for quiet settings. Situations, where the hands are occupied, require different approaches to maintain direct control over digital information when needed. This problem is addressed in chapters 4 and 5, which investigate foot-based and hands-free input methods. Further, the presented workflow steps are carried out sequentially, which is not always the case for minimally invasive interventions. For this reason, Chapter 6 investigates the influence of interaction tasks on manual tasks when performed simultaneously.

This chapter is based on the following publications:

B. Hatscher, A. Mewes, E. Pannicke, U. Kägebein, F. Wacker, C. Hansen, and B. Hensen. "Touchless scanner control to support MRI-guided interventions". In: *International Journal of Computer Assisted Radiology and Surgery* 15 (2020), pp. 545–553.

E. Pannicke, B. Hatscher, B. Hensen, A. Mewes, C. Hansen, F. Wacker, and R. Vick. "MR Compatible and Sterile Gesture Interaction for Interventions". In: *Proceedings of the 12th Interventional MRI symposium.* Boston, MA, USA, 2018, p. 43.

The content of the publication "Touchless scanner control to support MRI-guided interventions" [74] is a joint work between Dr. André Mewes, the second author, and the author of this thesis, whereby both authors contributed equally to the review of related works, a workflow observation and communication with the clinical partners, the conception, planning, performance, and evaluation of the user study as well as to the discussion. The author of this thesis conducted the online survey regarding clinical interaction practices, created the GUI and the connection with the MRI, and integrated the gestures, which were conceptualized, tested, and initially implemented by Dr. André Mewes. The authors of the publication confirmed by mutual agreement that the work may be used by all of them within the main focus of each persons work, which, in this case, is the comparison of direct interaction and proxy-user interaction in a medical setting.

The author of this thesis contributed the GUI and the connection with the MRI to the publication "MR Compatible and Sterile Gesture Interaction for Interventions" [150], which integrated the gestures conceptualized, tested, and initially implemented by Dr. André Mewes. Mr. Enrico Pannicke, the first author of the publication, contributed the concept for the MR-compatible Leap Motion Controller, hardware realization and evaluations regarding device functionality and MRI image quality.

4

Foot Input Techniques for Medical Image Manipulation

Footswitches are a common input method in the medical domain, for example, to trigger the X-ray source during radiological interventions [95, 82] or to control instruments in laparoscopic surgery [204]. Foot input, therefore, can be seen as an already established method of human-computer interaction in a sterile environment. Further, using the feet keeps the hands free for more important tasks and might reduce the need to pause a manual task to directly interact with devices. This chapter aims to extend the range of tasks to be performed by foot gestures and motions. To do so, different foot-based approaches for basic image manipulation tasks described in Section 2.1.2 are proposed. They are evaluated regarding their suitability for human-computer interaction of a single user in a standing pose.

The first section in this chapter utilizes a tactile floor to provide a foot input approach that does not require additional body-worn hardware but allows position-independent interaction. In the second section, the approach is extended towards less required floor space to match the confined setting in an OR. Pressure-sensitive sole inlays are repurposed for human-computer interaction in the third section. Different methods to perform continuous input by rotating the foot on the heel are presented, and compared in section four. The chapter is closed with a conclusion of the findings.

4.1 A Responsive Interface for Tactile Floors

This section presents a method to interact with medical image data using a tactile floor. Instead of physical pedals, rectangular areas on the floor are determined as input elements. Interaction elements may be out of reach or blocked by another user [82] foot pedals occasionally get lost [204] or have to be moved in a comfortable position [82]. Therefore, the interface adapts to the user by calculating an interaction area depending on the user's position and foot distance during an activation procedure. The virtual buttons are then associated with certain functionalities and are used to control a medical image viewer. A user study suggests that the proposed method is easy to learn in general and reveals user expectations as well as difficulties for certain foot gestures.

4.1.1 Interaction Method

During minimally invasive interventions, the clinician relies on live image data from imaging modalities such as X-Ray or MRI and therefore keeps its attention on the screen when navigating instruments [95]. In the following, design considerations regarding interaction and feedback in an interventional scenario and the resulting concept are described.

Input

Controlling medical image data is an auxiliary task that needs to fit interventional constrains and workflows. Therefore, the proposed interaction method is independent of the user's position on the floor. Instead, a specific activation gesture that can be performed anywhere on the floor causes virtual foot pedals to be arranged around the user's current position. Further, the activation method serves as protection against unintended input.

For activation, foot taps, which is lifting and lowering the foot or the ball of the foot, are employed. However, people lift and move their feet to maintain a stable stance in an upright position. Therefore, a triple-tap gesture triggers system activation as it is distinguishable from walking and single taps during stance correction. The time window to perform the triple-tap is two seconds. During activation, a virtual rectangle spans around the user's feet and is further used to define the positions of the virtual buttons for image manipulation. Once the system is ready to take input commands, triggering virtual buttons is done by tapping or stepping on the corresponding area on the floor. Manipulation of continuous values with a single button uses a two-rate approach: Upon activation, a low rate of change is applied, which increases to a higher rate after a fixed time duration. This allows both fast navigation to distant values and precise control. Deactivation can be achieved in two ways: Either by performing a triple-tap inside the virtual rectangle again or by leaving the virtual rectangle. This approach ensures that no further manipulation takes place in case the user abandons the system.

Feedback

In the OR, the feet are not always visible, which might lead to hitting the wrong pedal [204]. This applies to floor interaction as well and requires adequate feedback techniques. For the proposed system, visual feedback on a screen is employed for every interaction with the floor and additional auditive feedback for the triple-tap activation gesture. At all times, the area of the floor on which pressure above a certain threshold is detected, is displayed in the user interface. Activated cells are depicted in red. Feedback for the triple-tap activation gesture is provided in the form of a bar that represents detected taps and auditive signals for each tap. The bar empties in case of exceeding the time window of two seconds, or successful activation of the system. Upon activation, this is further indicated by another audio signal and the floor representation changing from greyed-out to colored and a green bar around one of the viewports (see Figure 4.1). The button position is depicted in the visual representation of the floor in the user's vicinity at the screen. The feet are represented by activated pressure cells. Upon stepping on an area depicted as a button, it turns green and the corresponding functionality is triggered. The button for switching between viewports is on the right of the user and labeled "3D->2D" or "2D->3D", depending on the currently active viewport. Buttons for incrementing and decrementing a value are located in front of the user. When interacting with 2D data, incrementing and decrementing is labeled "Slice+" and "Slice-". For 3D rotation, it is "X+" and "X-" or "Y+" and "Y-". Switching between the rotation axis is done using the button "RotaSwitch" on the left, which is only available in 3D mode.

4.1.2 Evaluation

A user study was conducted to investigate the suitability of the presented approach for foot interaction with medical image data. Ten participants (3 female, 7 male) in the age group between 21 and 29 years took part in the study. Nine participants majored in computer science, one majored in business administration.

Apparatus

A tactile sensor floor from the Fraunhofer Institute for Factory Operation and Automation (IFF, Magdeburg, Germany) was used, which was developed for the context of human-robot interaction for industrial applications [137, 49]. The tactile sensor system measures 160×95 cm and consists of a matrix of piezoresistive sensors, which are embedded in a robust plastic shell. To simulate the conditions in the OR as best as possible, the sensor floor was equipped with a linoleum coating, similar to the flooring of an OR. The entire floor comprises 608 individual sensor cells (taxels) arranged in 32 columns and 19 rows. The spatial resolution of the sensor floor is approximately $5 \times$

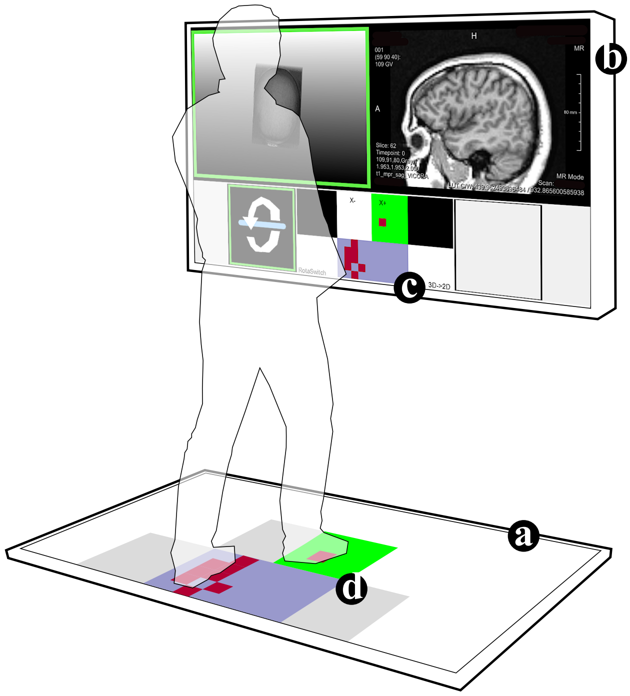


Figure 4.1: Study setup with a tactile floor (a) and a display showing the software prototype (b). Visual feedback (c) shows the virtual rectangle initially spanning around the user's feet position and virtual buttons (d). (virtual buttons on the floor were not visible during the study).

5 cm, which is sufficient to localize persons and determine foot orientation. A central sensor controller is used for data acquisition. It has a 12 bit A/D converter. Within the scope of this application, the measurement data are provided with a sampling rate of approximately 50 Hz via an USB interface.

A software prototype integrates the proposed input and feedback methods, the medical image viewer to be controlled and the data interpretation from the tactile floor. The feet are determined by performing a connected component analysis (8-connectivity) if a pressure threshold value is exceeded. The graphical user interface consists of a 2D and a 3D viewport as well as dedicated areas in the lower third of the screen for visual feedback, activation gesture and currently performed interaction. The software allows switching between viewports, scrolling through 2D images and rotating the 3D volume data.

Table 4.1: Tasks to be performed with foot input during the user study.

Task	Description
1	Activate the system
2	Navigate to slice Nr. 45 of the MRI-Data in the 2D view
3	Rotate the 3D-representation in a way that the face is visible
4	Rotate the 3D-representation in a way that the left ear is visible
5	Deactivate the system

Procedure

The entire procedure was recorded on video to perform the evaluation afterward. To assess the self-explanatory qualities of the proposed approach and the expectations of the user's regarding foot interaction, no introduction to the system was given, only the activation gesture was explained at the beginning.

Every participant performed a sequence of tasks, which covered all the functionalities of the system (see Table 4.1). To make sure the participants could perform all the tasks without expert knowledge, the descriptions asked bringing external features of a head in a MRI-dataset into view. After each subtask, users were asked to explain what problems they had with the task and what actions could be taken to support their understanding of the system. After completing the tasks, the user's filled out a software evaluation questionnaire.

Results

Participants did not immediately know how to interact with the system. However, task four, which was based on the previous ones, was solved much faster, after tasks two and three were completed. This shows that after a short familiarization phase, users were already getting used to the system.

The evaluation of the video protocols can be found in Table 4.2. Four participants had problems with the successful execution of the activation gesture. Three participants did not recognize the buttons of the user interface as interactive elements at first. Therefore, attempts to solve tasks two and three were by slide or scroll gestures were observed. However, a much more common problem faced by five participants was that they did not know how to switch between 2D and 3D mode in task three. The main reason was the incomprehensible label "2D->3D" of the corresponding button. In task five, six participants came up with the idea that deactivation might work the same way as activation. Three participants simply left the virtual, central rectangle

Table 4.2: Observed interaction approaches during the user study. Actions that resulted in the desired outcome are empathised.

Task	Observation	Distribution
1	*Gesture recognition at first attempt*	*6/10*
	Gesture executed too fast to be recognized	3/10
	Gesture not executed within the given time window	1/10
2	*Use of the provided buttons*	*9/10*
	Slide-gestures	1/10
3	*Use of the provided buttons*	*5/10*
	Slide-gestures	2/10
	Assistance needed	3/10
4	*Use of the provided buttons*	*10/10*
5	*Gesture analogous to 'activation'*	*6/10*
	Leaving the central rectangle	*1/10*
	Leaving the tactile floor	*2/10*
	Assistance needed	1/10

or the tactile floor, which both led to a deactivation of the system after three seconds. One participant was not able to deactivate the system immediately. The visual and acoustic feedback was perceived as intuitive and adequate by nine participants and helped them to understand the system. Figure 4.2 shows the results of the final questionnaire, which asked participants to rate the software/user interface on the six presented scales.

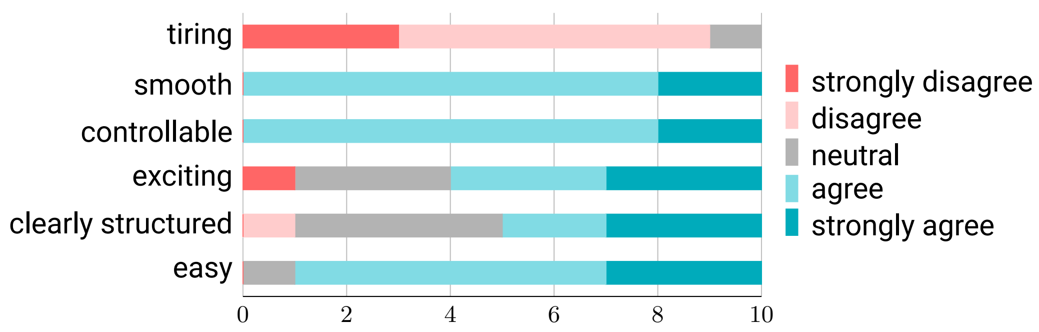


Figure 4.2: Results of the post-test questionnaire.

4.1.3 Discussion

The study results show that a tactile floor is suitable for interaction with medical image data sets. The need for optimization is evident in gestures such as the *triple-tap*, which is poorly recognized despite visual and acoustic feedback. From observations, it can be concluded that the term *tap* for foot gestures is too unspecific and not established, which leads to different interpretations. The prototypical implementation of the system captures only taps with the complete foot, ball taps are omitted. In addition, participants performed gestures at very different paces. A gesture was not recognized when exceeding the predefined time window, which occurred once. The constant recognition rate in the first and last tasks show that the corresponding gestures are difficult to execute even after learning. Due to the required robustness of an activation gesture, either clear instructions or personalized gestures might solve this issue.

Virtual buttons in the user interface were not immediately recognized as such. Attempts to trigger image navigation with slide gestures took place in the second and third task but were increasingly observed in the 3D task. This might be due to the familiarity of these kinds of gestures from touchscreen devices. The labeling of the buttons was partly criticized for being hard to understand, especially "RotaSwitch" for changing the rotation axis in the 3D display. This might also be attributed to the prototypical appearance of the graphical user interface. It should be emphasized that task four, as a combination of tasks two and three, was successfully performed by all participants without assistance. The system, therefore, seems to require only a brief training period. The post-test questionnaire shows that the system was mostly perceived as not very tiring, smooth, controllable and easy. The prototypical appearance of the user interface might account for the wide range of responses in terms of excitement and structure.

Even thought this work provides meaningful insights, there are various aspects a lab study cannot take into account. Therefore, the findings must be viewed critically. To allow video evaluation of the user interaction, the participants did not stand at a table covering the view of the floor, as it would be the case in the OR. However, there was no visual feedback on the floor, which means that looking at the feet only provided visual feedback on the position of the feet in case proprioceptive feedback was not sufficient. During the user study, the participants were able to concentrate entirely on operating the tactile floor instead of being forced to divide their attention between interaction with the medical image data and a manual task. Further, medical interventions might last longer than the presented user study, which means that fatigue might play a more significant role in a real setting. Nevertheless, the presented approach shows the potential of a tactile floors as input devices for hands-free, position-independent interaction with medical image data.

4.2 Floor Interaction in Confined Spaces

This section compares two hands-free interaction concepts for navigation in medical image data using foot gestures. The approach investigated in the previous section is adapted to off-the-shelf hardware. Additionally, a concept that requires less free floor space is presented. The concepts are compared in a user study in terms of TCT, UX and subjective workload.

4.2.1 Hardware

Currently, sensor floors are not a standard piece of hardware in state-of-the-art operating rooms. This section investigates the suitability of off-the-shelf hardware for foot gesture interaction to make real-world application of this research possible. SensFloor® (FutureShape GmbH, Höhenkirchen, Germany) is a commercially available capacitive sensor flooring that is used for a variety of applications, in particular for gait analysis and fall detection. Sensor cells detect capacity changes of the area above them so that the position of a person can be determined [194]. A cell is declared active when the capacity value exceeds a certain threshold. The sensor floor is an underlay covered with quadratic modules of approximately $48 \times 48\,\mathrm{cm}$, each consisting of eight triangular capacitive sensor cells. Due to the modular structure, the floor can cover an area of any size. This hardware was already investigated for medical image navigation in a user study but served as an auxiliary input for switching between monitors and activating/deactivating hand gesture recognition [94]. The sensor floor used in this work covers an area of $2\,\mathrm{m}^2$. The sensor floor underlay was covered with linoleum, which is a common flooring in operating rooms.

4.2.2 Interaction Methods

In the following, two interaction concepts are described. They differ in the gesture set and required space around the user. However, both concepts implement the same functionalities, which are:

- A system activation method to prevent unintentional input
- Changing the 2D layer in an image dataset (CT, MRI)
- Rotation of 3D volume data

Further, both concepts share the same activation method. A double-tap is used instead of a triple-tap as in Section 4.1 to improve recognition yet still make it distinguishable from walking or corrective foot movements. A double-tap with the whole foot has to be performed within one second to activate the system.

Concept 1: Virtual Buttons

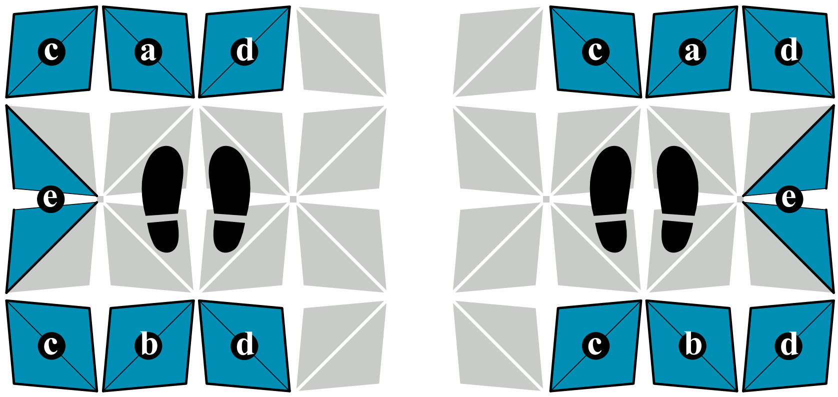


Figure 4.3: Section of the sensor floor with *virtual button* positions and the corresponding sensor cells. Depending on the foot used for activation, the left or right button layout was used. (a) increment slice/rotate up, (b) decrement slice/rotate down, (c) rotate left, (d) rotate right, (e) switch between 2D and 3D mode. Image adapted from Wagner et al. [216].

This concept is based on the approach presented in Section 4.1. Defined areas on the floor are triggered by stepping on it. In comparison to Section 4.1, the mapping between buttons and executable actions was optimized, so that the button positions correspond to the direction of navigation. Further, the concept was refined based on clinical user feedback to such an extent that all functions are accessible with one foot. To ensure faster navigation, a two-rate approach similar to Section 4.1 is used. If the user stays on a button for longer than 2.5 s, the time interval between automatically triggered navigation steps is reduced from 0.5 s to 0.2 s. An image that shows the button positions relative to the position of the user is used as visual feedback (see Figure 4.4). Icons describe the function of each button. For navigation with this concept, $1\,\mathrm{m}^2$ of free space is required.

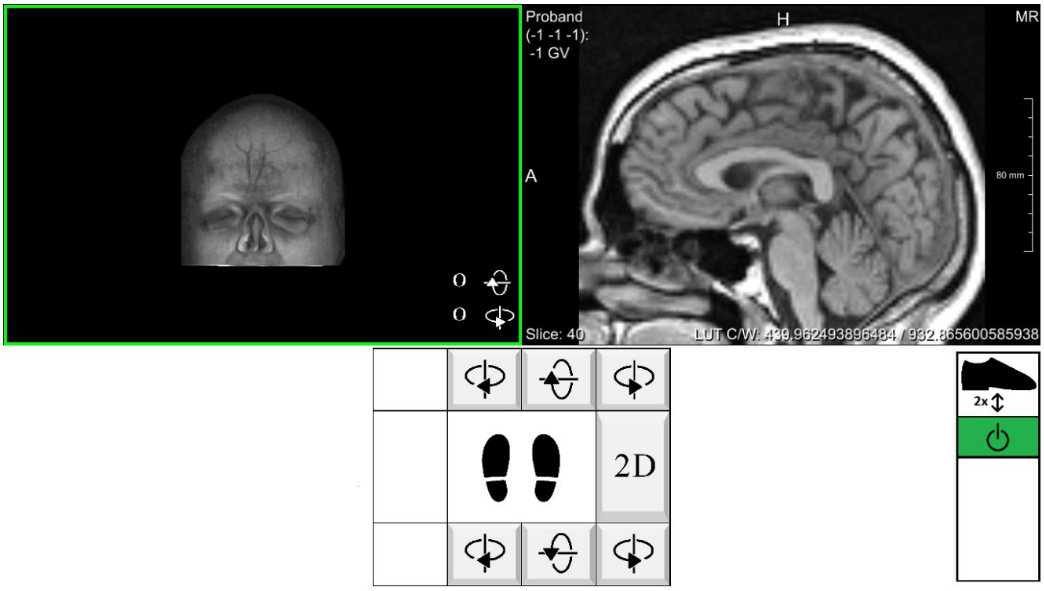


Figure 4.4: Graphical user interface for the *virtual buttons* concept showing input elements for the currently active viewport and a button to select the 2D viewport. Icons instead of textual labels depict button functionality when possible. Image from Wagner et al. [216].

Concept 2: Minimal Interaction Space

This concept aims to enable the interaction with the image data using minimal floor space. It is based on foot gesture sets by Sangsuriyachot and Sugimoto, which use pressure distribution, foot rotation and transition [174]. Due to the resolution of the sensor floor, lateral foot rotation could not be detected robustly. Instead of only rotating the foot on the heel, a slight transition was added to allow detection with the given sensor cell pattern (see Figure 4.5 c and f). Further, a dorsal and plantar flexion of the ankle were adopted from Sangsuriyachot and Sugimoto [174]. The resulting gesture set is shown in Figure 4.5. Compared to the *virtual buttons* concept, all functions can be accessed directly so that no mode change is necessary. Slice navigation is performed with the right foot only, and the rotation upwards and downwards is performed with the left foot only. The use of both feet is necessary for lateral rotation. The distribution of gestures on the individual feet was chosen to provide a reference between the gestures and the corresponding image viewers. The 3D dataset was placed on the left half of the screen, and the 2D dataset in the right one.

The same mechanism for accelerating interaction as in the *virtual buttons* concept is used. Another difference to the *virtual buttons* concept is that a

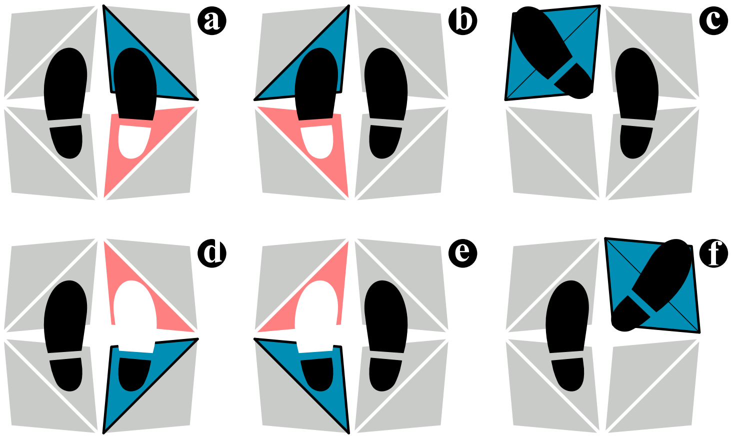


Figure 4.5: Gesture set for navigation with *minimal interaction space*. The white part of the footprints indicate where the foot needs to be lifted to trigger a functionality i.e. red cells must not be activated. (a) increment slice, (b) rotate up, (c) rotate left, (d) decrement slice, (e) rotate down, (f) rotate right. Image adapted from Wagner et al. [216].

gesture needs to be held for 0.5 s before the gesture is recognized. Because the user's foot is rarely completely even when being lowered and, therefore, might activate sensor cells in an arbitrary order that would trigger a different command than the final foot position, this mechanism was introduced to avoid unintended execution of functions. A legend provides visual feedback, which represents the relationship between function and gesture as pairs of icons. In order to support the user, the functions were grouped and visualized below the corresponding image viewer. Furthermore, these pairs were positioned according to the navigation direction. The user interface is shown in Figure 4.6. For navigation with this concept only a free area of approximately $0.25\,\mathrm{m}^2$ is required, which is a quarter of the space required for the *virtual buttons* concept.

4.2.3 Evaluation

In order to evaluate the developed concepts, a user study was carried out. The study followed a within-subject design, which means that all participants were exposed to all conditions. The task order was randomized to minimize learning effects. The subjective workload was measured with a RTLX questionnaire [67], and the user experience was measured with the dimension "usability" of the meCUE questionnaire [132].

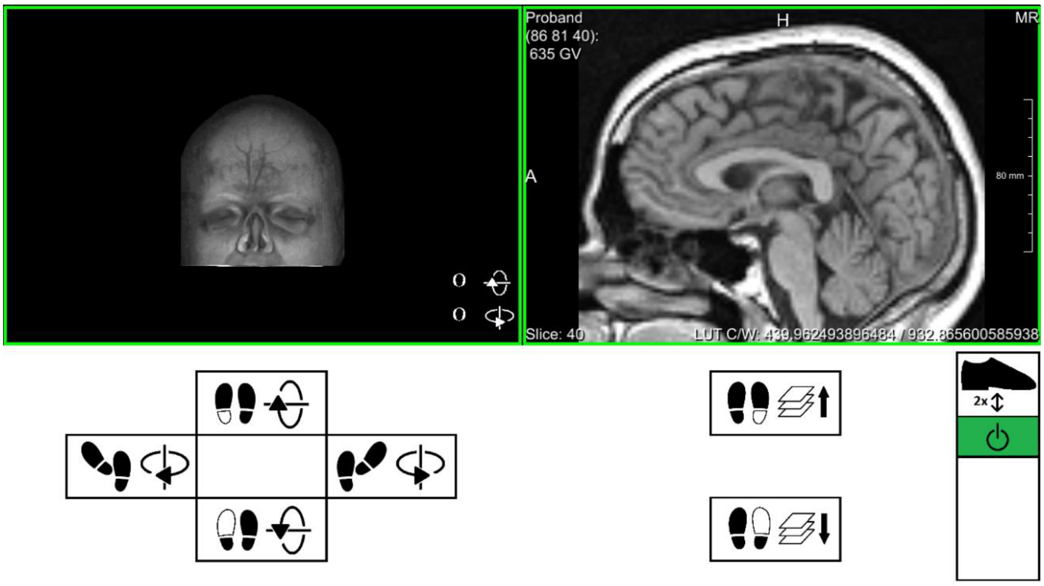


Figure 4.6: Graphical user interface for the *minimal interaction space* concept. Due to the availability of all functionalities without switching between a 2D and 3D mode, gestures are displayed below the corresponding viewports. Image from Wagner et al. [216].

Participants

Thirteen medical students (eight female, five male) took part in the study. The average age was 23.9 years, the average shoe size 40.6 Paris points (European Continental System). One participant reported basic knowledge of foot interaction from a previous study. All other participants reported no experience at all.

Apparatus

In order to make the conditions as similar as possible, the participants wore plastic shoes throughout the study, similar to the ones used in operating rooms. A marker on the floor indicated the standing position during the study to ensure an equal distance to the sensor areas. A 27-inch screen displaying visual feedback was placed in front of the participant about 1.5 m away. An additional screen displayed the current task.

Table 4.3: Tasks to be performed by the participants during the study.

Task	Description	Category
1	Activate the system	Activation
2	Deactivate the system	Activation
3	Activate the system	Activation
4	Go to slice 48 (starting at slice 40)	Slicing
5	Go to slice 45	Slicing
6	Rotate to the right by 10 steps	Rotation
7	Rotate to the top by 15 steps	Rotation
8	Go to slice 25	Slicing
9	Rotate to the bottom by 20 steps	Rotation
10	Rotate to the left by 3 steps	Rotation
11	Go to slice 32	Slicing
12	Deactivate the system	Activation

Task

There were three categories of tasks:

- Activating/deactivating the system
- Slice change in an image data stack
- Rotation of a 3D volume display

The participants had to perform four tasks from every category for each concept, which resulted in 24 tasks in total (4 tasks $\times$ 3 categories $\times$ 2 concepts). The exact sequence of tasks can be found in Figure 4.3.

Measures

Task completion time (TCT) was logged automatically during the study. The subjective workload was assessed using a RTLX questionnaire. For measuring usability, the corresponding dimension from the meCUE questionnaire was employed. Further, errors in gesture detection were recorded.

Procedure

A thematic introduction was given at the beginning of the study, and demographic data were collected. The participant performed a training phase of five minutes to try out the system and to clarify any questions. Four test tasks ensured that the participant understood the basic functionality. Subsequently, a total of 12 tasks per concept were performed. Following each condition,

Table 4.4: Summary of statistical tests for subjective workload (SW), usability (U) and task completion time (TCT)

	df	F	T	p	sig	η^2	d	effect
SW	12		-1.669	0.121			0.46	small
U	12		1.399	0.187			0.39	small
TCT								
Concept	1, 12	6.31		0.030	*	0.35		large
Assignment	1.59, 19.02	64.12		< 0.001	*	0.84		large
Interaction	1.50, 18.04	1.19		0.310		0.09		medium

a RTLX questionnaire and the described part of the meCUE questionnaire was filled out. At the end of the study, an interview was conducted to gather comments on the interaction concepts.

Results

The developed concepts were implemented for an evaluation. The results of the user study are described below. Table 4.4 shows the results of the analysis of variance (ANOVA). The results for the subjective workload across all six dimensions of the RTLX questionnaire were not statistically significant. When navigating with *virtual buttons*, the subjective workload overall was slightly lower (M = 6.08, SD = 3.12) than when interacting on a *minimal interaction space* (M = 7.30, SD = 2.38). Only the physical demands were estimated to be lower (M = 6.31, SD = 3.37) when using the *minimal interaction space* concept than when using *virtual buttons* (M = 7.92, SD = 5.50), which can be seen in Figure 4.7. Also, no significant differences between the two concepts could be observed in terms of usability. *Virtual buttons* (M = 6.29, SD = 0.91) performed slightly better overall than *minimal interaction space* (M = 5.90, SD = 0.68).

However, significant differences could be identified between the concepts concerning the processing time of the tasks. On average, the tasks could be completed 2.32 s faster with *virtual buttons* (M = 6.25 s, SD = 0.28 s) than with the *minimal interaction space* concept (M = 8.57 s, SD = 0.87 s) (see Figure 4.8).

The activation gesture was not successfully detected in 12 (23%) of the 52 performed activation tasks (13 participants × four task repetitions) with *virtual buttons* and 10 cases (19%) with the *minimal interaction space* concept. It has to be noted that both concepts use the same double-tap activation gesture. Navigating through image slices yielded 35% input errors by users for

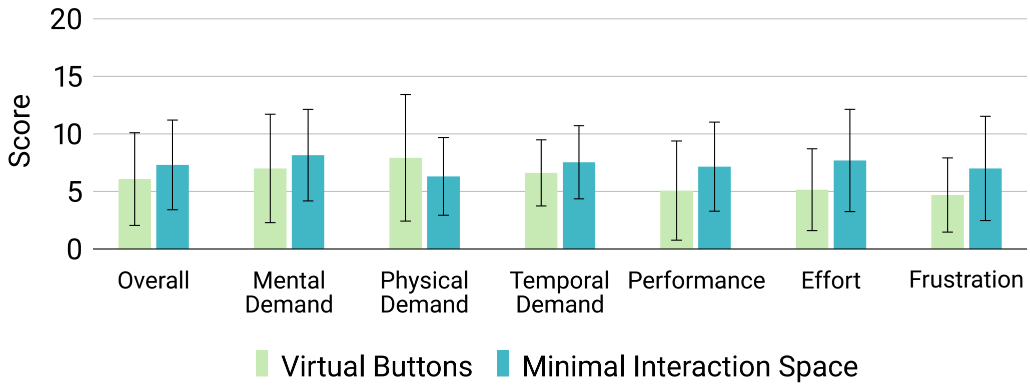


Figure 4.7: Unweighted mean scores for the NASA-TLX questionnaire on a scale from 0 to 20 (0 = low/good performance, 20= high/poor performance). Error bars represent standard deviation. The overall workload is the mean value from all the subscales.

virtual buttons and 58% for *minimal interaction space*. During the rotation task, errors occurred in 25% of the tasks when the *virtual buttons* concept was used, and in 46% of the tasks for *minimal interaction space*. An overview of all user error percentages can be found in Figure 4.9.

The visual feedback was reported to be very good and intuitive. Problems with gesture detection resulted from slight position changes during the use of the system. All participants were able to interact successfully with the system after a short familiarization phase.

The evaluation of the interaction concepts showed that *virtual buttons* performed better overall. Significant differences were found for task completion times. Regarding usability and subjective workload, interaction with *virtual buttons* also performed slightly better. Only physical demand was rated higher for *virtual buttons* compared to *minimal interaction space*. Participants found the positions of the buttons too far away, which required big steps and, in turn, caused balance problems.

4.2.4 Discussion

In this section, two interaction concepts for hands-free navigation of 2D image series and 3D datasets were compared. The concepts differ in the applied gesture set, the required floor space and the visual feedback. The visual feedback proved to be appropriate and enabled the participants to navigate through the medical image data without continually having to look at their feet. After a short training phase, all participants were able to interact without

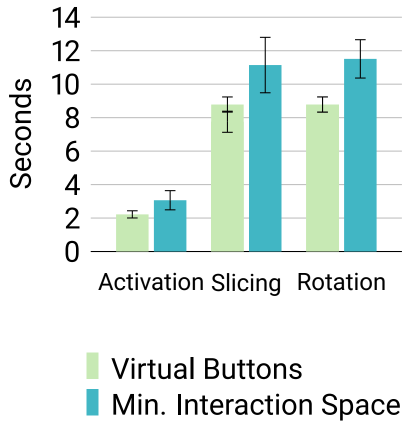


Figure 4.8: Mean results for the task completion time (TCT). Error bars show standard error.

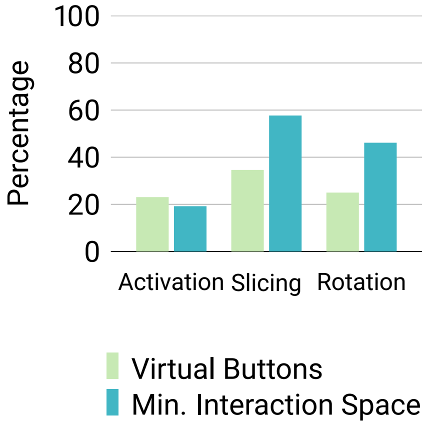


Figure 4.9: Mean user errors for each task type.

assistance. For *minimal interaction space*, difficulties with the robustness of gesture recognition were revealed, which might explain the higher degree of frustration. Further, unintended functions were executed, which increased the task completion time due to the required correction steps. The difficulties during navigation were mainly due to the acceleration mechanism, which should enable faster navigation. However, this led to an opposite effect, which might emerge from the lack of control over the exact point in time when the speed change happens. A possible solution for this problem might lie in controlling the navigation speed with foot plantar pressure. Since no mode change is necessary for interaction on a minimal interaction surface, a slight time saving was initially expected. Due to the difficulties encountered in navigation and the resulting longer task completion time, no statement can be made about the time saved by omitting the mode change during interaction for *minimal interaction space*. Overall, the interaction concept *virtual buttons* was slightly better in terms of usability and subjective demands. Further, *virtual buttons* yielded significantly lower TCT.

There is always more than one person in the OR, so it is necessary to recognize different users. Additional sensors or recognition algorithms developed for this purpose could be used [110]. Identification using wearable sensors would void the advantage of intelligent floors, which is interaction without additional hardware. In the long run, it needs to be investigated if gait detection in confined spaces can be used to identify users reliably. Furthermore, a calibration phase for individualized gesture recognition should be considered. This could result in more robust gesture recognition [90].

This work was able to confirm the results of the previous section regarding the feasibility of hands-free navigation in medical image data using foot gestures. In addition, a concept was developed that allows navigation within a *minimal interaction space*. However, the robust recognition of the foot gestures in order to counteract the navigation inaccuracies and the resulting longer task completion time needs further improvement to match established input methods. Due to the resolution of the employed hardware, the buttons used as *virtual buttons* were relatively far away from the user, so that relatively large steps were necessary. Floors with higher resolution or foot-mounted approaches might lead to more robust results and allow personalization without additional identification steps. However, this work may inform further research regarding the development of gesture-based systems for navigation in medical image data under sterile conditions.

4.3 Foot Input using Plantar-Pressure

Both previous sections outline the benefits and drawbacks of floor-based foot input approaches. Especially section 4.2 reveals difficulties for gesture recognition at relatively low floor resolutions. Close proximity between users might further hinder user identification and, in turn, renders personalization impossible without technological additions.

This section focuses on controlling a medical image viewer using foot plantar pressure, which is the pressure field present between the sole and the surface. By shifting the weight or flexing the toes, the pressure distribution can be changed and thus used as a hands-free input channel. Plantar pressure has already been used for a variety of applications such as user and surface identification [117] or to perform subtle, non-observable foot gestures [51]. Fitzke et al. presented a plantar pressure-based input approach for mouse cursor control with adaptable speed, intended for medical applications [47].

4.3.1 Interaction Method

In this section, an approach that uses different zones of the foot's sole to trigger certain functionalities is described. Even though gesture sets for plantar pressure input using the heel, ball, outsides and insides of the foot have been proposed [51, 174], informal studies suggested using the heel and the ball rather than the sides of the foot. As an alternative to directional approaches, this work investigates overlapping pressure zones. As can be seen in Figure 4.11, pressure-sensing areas are defined in a way that allows using

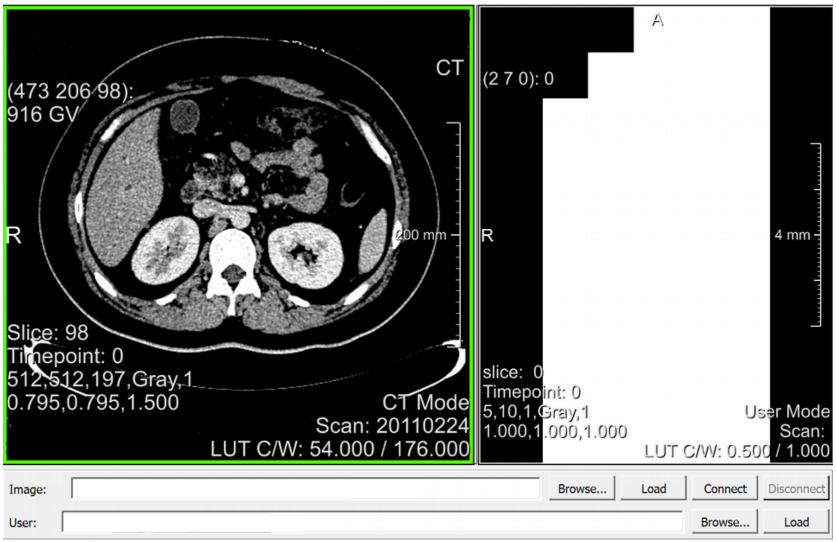


Figure 4.10: User interface showing the medical image data set on the left and the pressure sensor matrix on the right. Cells exceeding a threshold are depicted in black. The green border indicates an active state of the system, hence image slice navigation is enabled. Image from Solovjova et al. [192].

the ball of the foot or the toe. The same pattern is used at the heel, where a small area is defined to detect pressure at the back of the foot in addition to heel pressure.

These input areas are used to navigate medical image data by scrolling up and down an image stack. Apart from this basic functionality of a 2D medical image viewer, an activation mechanism to avoid unintentional input is required. From a human-computer interaction viewpoint, both functionalities serve as examples for two types on input tasks, namely discrete and continuous input, and therefore cover a variety of applications where hands-free interaction is important.

4.3.2 Evaluation

A user study was conducted to assess the subjective workload while using the proposed input method for plantar pressure-based human-computer interaction for manipulating medical image data.

Measurements

The subjective workload was assessed using the NASA-TLX questionnaire without the extra step of weighting the subscales (RTLX [67]).

Participants

Fourteen students (11 male, 3 female) took part in the study. The age ranged from 20 to 28 (M = 22.43, SD = 2.53). The mean shoe size was 42.21 Paris points (SD = 2.58). Only one of them reported previous experience with foot input methods.

Apparatus

For the study, a shoe insole equipped with 50 pressure sensors was used. The PlantaPress™ system (Thorsis Technologies GmbH, Magdeburg, Germany) consists of two inlays with a small box each that holds the power supply and the electronic components for Bluetooth communication. The box is to be mounted on the ankle of the corresponding foot. A Bluetooth dongle receives the pressure data for further processing.

A laptop ran a prototypical software that interpreted the data. As weight, physiology and personal preferences of users differ, so do readings from pressure-sensing hardware. Therefore, the threshold value for each person, and each gesture had to be determined. This was done by calculating the threshold value using the following formula while the gesture in question was held:

$$P_{threshold} = \frac{1}{t} * \sum_{i=0}^{t} p_{gesture}(i) * constant$$

This way, specific values for each gesture, and each user were gathered and applied during the study. The software further provided a medical image viewer and visual feedback for the plantar pressure. The graphical user interface of the software can be found in Figure 4.10.

Procedure

The study took place in a computer laboratory. The participants had to wear a pair of clogs, equipped with the pressure-sensing inlay. Clogs were provided in 40/41, 42/43 and 44/45 Paris points. The chosen pair was then equipped with the inlay of the corresponding size. At first, an instructor explained the gestures. For each participant, a user profile consisting of the calculated thresholds for each gesture was created, using the method described in section 4.3.2. The participant performed the tasks listed in Table 4.5. In case of a

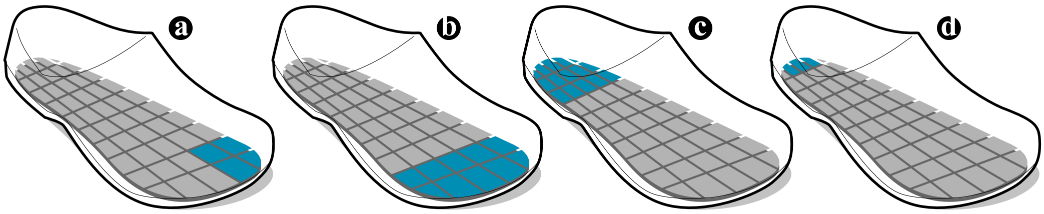


Figure 4.11: Groups of cells of the pressure-sensing inlay to be activated simultaneously to trigger system activation (a), increase of slice number (b), decrease of slice number (c) and system deactivation (d).

failed attempt, the task was repeated up to three times, which was sufficient for all of the participants. After completing a task, a RTLX questionnaire was filled out.

Table 4.5: Tasks to be performed during the user study with corresponding actions to be executed.

Task	Description	Action
1	Activate the system	Pressure on the big toe, while lifting the heel
2	Increase slice number	Shift the weight to the ball of the foot
3	Decrease slice number	Shift the weight to the back of the foot
4	Deactivate the system	Pressure on the heel, while lifting the ball of the foot
5	Activate the system	Pressure on the big toe, while lifting the heel
6	Go to slice number 151	Shift the weight to the ball of the foot
7	Got to slice number 68	Shift the weight to the back of the foot
8	Deactivate the system	Pressure on the heel, while lifting the ball of the foot

Results

When it comes to subjective workload, the RTLX scores show similar ratings for all gestures (see Figure 4.12). Comparing the individual subscales, however, reveals differences that level each other out in the overall score. None of the gestures was mentally very demanding, whereas the gesture for scrolling down got the highest score for effort, temporal demand, performance and frustration. In terms of performance, a higher score means a lower perceived

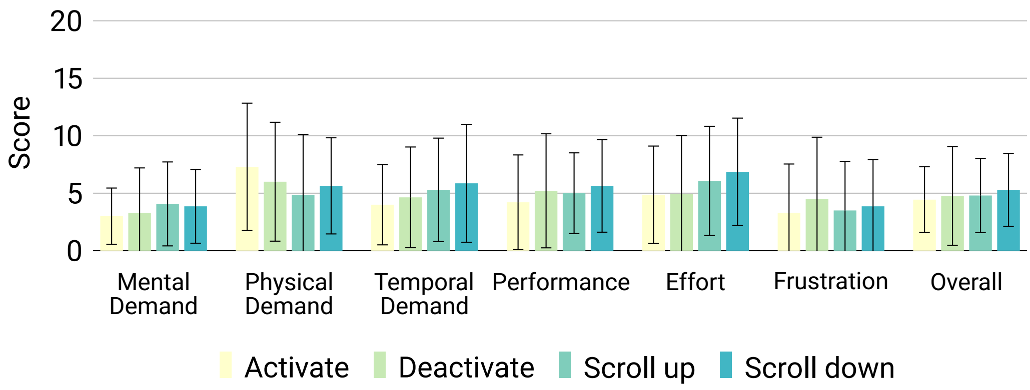


Figure 4.12: Unweighted mean scores for the NASA-TLX questionnaire on a scale from 0 to 20 (0 = low/good performance, 20= high/poor performance). Error bars represent standard deviation. The overall workload is the mean value from all the subscales.

performance. Compared to that, the activation gesture scored best on all subscales except on physical demand.

During the 112 measured tasks (14 participants × 8 tasks), involuntary deactivation of the system happened four times. While attempting to disable the system, the slice number was changed seven times. Six failed attempts to reach a specific slice number were observed.

4.3.3 Discussion

This section introduced a new interaction concept using overlapping pressure-sensing zones for plantar pressure input. The perceived task load using the presented approach is relatively low overall. Especially the low mental demand indicates the potential to be used in demanding scenarios where more critical tasks have to be performed in parallel, such as the OR. However, physical and temporal demands have to be lowered to provide a suitable user interface, which might, in turn, lower the perceived effort and increase the perceived performance. During the user study, the presented calibration had to be corrected manually in several cases, indicating the need for a more elaborate approach. In addition, user preferences such as scroll speed, size and position of the pressure areas should be added to the user profiles to leverage the advantages of a personalized input method.

However, given that the hardware was design for passive, diagnostic purposes, the approach performed reasonably well as users got accustomed to the system quickly. Adaption to the proposed use case by adding an accelerometer and an angular velocity sensor would allow for a broader range of possible foot gestures and a more reliable gesture detection by applying sensor fusion algorithms. This would allow interaction with medical images in multiple ways, depending on the user's preferences and requirements, but without the need to put the current manual task at rest.

4.4 Interaction Techniques for Heel Rotation

Foot interaction based on pressure data with different approaches, covering floor-mounted and body-worn sensors was investigated in the previous sections. These approaches rely on supplying pressure to a surface with the foot. However, the feet are capable of performing a much more extensive range of motions that can be used for interaction, which do not necessarily require floor contact. The approach presented in this section draws suitable foot movements from literature and connects these findings to ergonomic guidelines for foot pedals. On this basis, suitable input techniques are determined and converted into concrete interaction concepts. These input concepts are then applied to manipulating a continuous value, which covers a wide range of practically relevant interaction tasks such as navigation through medical images or adjustment of brightness and contrast of DICOM-images.

To evaluate the approaches, a prototypical system consisting of an input device and a medical image viewer is developed and used to compare the concepts in a user study in terms of performance, subjective workload, usefulness, usability, positive and negative emotions and intention of use.

4.4.1 Design Considerations

When performing a radiological intervention, the physician is standing in front of the patient table [82], which limits foot gestures to single foot input to maintain balance [208]. Delicate manual tasks such as navigating a needle or a catheter to a target structure require a steady hand. Both of these factors limit the range of foot movements available for human-computer interaction in such a situation. Even established input methods such as foot pedals bear some disadvantages. They can be uncomfortable, out of view and easily get lost under the operating table [219]. Keeping the foot on the pedal, on the other hand, causes fatigue and might lead to involuntarily triggering [18]. Further, heavy lead aprons and covers in radiology make it hard to perform complicated foot gestures.

Difficulties in distinguishing foot pedals were reported during laparoscopic surgery, which results in occasionally hitting the wrong switch. This can be dangerous, for example, when triggering cutting instead of coagulation on diathermic equipment [204]. In the process, ergonomic guidelines for the use of foot pedals were developed [204]. This includes ergonomic as well as technical aspects. This work follows these guidelines with minor modifications. Two of the ten guidelines were discarded as they did not apply to the development of interaction techniques. Clog dimensions the system has to work with are given in guideline five. This guideline was discarded because even when exceeding these dimensions, the proposed approach should work. Guideline eight requires the system to work with or without clogs. The goal is to identify suitable interaction methods rather than developing an actual input device for clinical use. Therefore there is no need to comply with guideline eight at this stage of the research process.

The guidelines derived from [204] therefore are as followed:

- The design of the input device must avoid a static standing posture.
- Dorsal flexion of more than 25° to control the device is not allowed.
- The force for activation must not exceed 10 N.
- A frequent dorsal flexion of the foot should be avoided.
- The input device must be controlled without looking at the foot pedal.
- The chance of accidentally activating the wrong function must be minimal.
- The chance of losing contact with the input device must be minimal.

When it comes to the kind of human-computer interaction that is required during radiological interventions, it is crucial to understand the workflow and challenges. During needle intervention or when navigating a catheter through the patient's blood vessels, neither the device nor the target structures are visible directly. The radiologist, therefore, relies on an imaging modality such as an angiography system. Unfortunately, blood vessels are effectively transparent to X-rays, which necessitates the use of a contrast agent. This agent is administered into the blood vessels and is distributed by the blood flow. During a short time window, the vessel tree is visible on X-ray by means of the contrast agent. The images acquired during this phase hold not only static information, but show the speed of the blood flow and can resolve overlapping that might be confusing when viewed in a static image. To understand medical image data, a practitioner needs to directly control the image sequence [142]. However, in terms of human-computer interaction, only a single degree of freedom is required to realize this kind of manipulation, which can be achieved using foot input.

4.4.2 Interaction Concepts

As described in Section 2.2.1, heel rotation combined with lifting and lowering the tip of the foot seems to be a promising input method. However, these movements can be mapped to a value in different ways. In the following, three input concepts that implement rate-based interaction, relative interaction and both of them are presented. For consistency, all concepts realize mapping between the value to change and the foot movements in a way that turning the foot inwards (i.e., the right foot towards the left one) corresponds to decrementing and turning the foot outwards corresponds to incrementing. Further, lifting the ball of the foot always disables input, similar to lifting a computer mouse for repositioning.

Discrete Buttons

In this concept, virtual buttons are arranged in a fan-like fashion (see Figure 4.13). The layout is similar to [232], who used it for menu item selection. In contrast, this work uses the buttons for rate-based input. Rate-based input means that a parameter is changed at a specific rate, as long as a condition is met. Typical examples are joysticks or a car's accelerator pedal. Often, the rate at which the change takes place can be varied. Rate-based input has been used for human-computer interaction utilizing pedals [103], pressure distribution [174, 51] or kicks [7].

Each button occupies 20° to allow safe selection at a heel rotation selection error of 8.52° [183]. Further, the position and number of buttons are based on the active range of motion found by Zhong, Tian, and Wang [232], which results in five slots, ranging from $-40°$ to $60°$. 0° corresponds to a relaxed foot position, pointing straight forward. To activate a button, the foot has to be rotated over the corresponding area and lowered to the ground. The slot at 0° is left without function to allow resting the foot when no interaction is intended. Both the innermost buttons change the value by one every 800 ms while the outermost buttons provide a faster way by changing the value every 200 ms.

Foot Scrolling

Building upon previous knowledge, *foot scrolling* implements one degree of freedom (DoF)-interaction similar to scrolling on a touchscreen device. Foot movement is mapped relative to the manipulated value. As long as the the foot is on the floor, rotating on the heel increments or decrements the value by one every 10°, depending on whether the foot is turned inwards or outwards. In case the desired value cannot be reached with a single motion, lifting the foot allows repositioning without changing the value. As no interaction takes place without movement, there is no need for a dedicated resting area as in the *discrete buttons* concept.

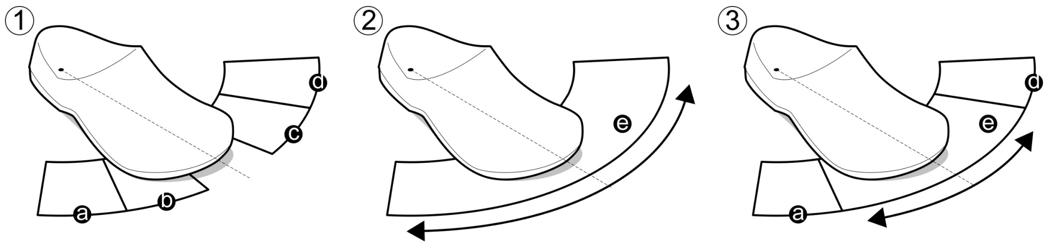


Figure 4.13: Concepts for manipulating one degree of freedom by rotating the foot on the heel. The *discrete buttons* concept allows to (1) increment (a,b) or decrement (c,d) the current image by keeping the foot on the floor in the corresponding areas. The image is changed at a rate of 0.2 seconds by the outmost buttons (a,d) and 0.8 seconds by the innermost ones (b,c). *Foot scrolling* (2) allows to change the current image continuously every 10° by rotating the foot on the heel with a lowered foot (e). Repositioning can be done without affecting the current image by lifting the tip of the foot. *Step and scroll* (3) combines rate-based input (a,d) and scrolling by rotating the foot on the heel (e). Image from Hatscher, Luz, and Hansen [72] © Walter de Gruyter and Company 2018[1].

Step and Scroll

Both the former concepts bear potential advantages and disadvantages. *Discrete buttons* provide two fixed rates, which most probably does not include the preferred rate for every user. *Foot scrolling*, on the other hand, requires continuous motion of the foot, which might cause fatigue when used for langer changes of the value. *Step and scroll*, therefore, integrates both concepts by providing two buttons for fast scrolling at the outermost positions and a scrolling area in between. This way, fast manipulation without causing fatigue can be achieved by keeping the foot on a virtual button while fine-grained control is possible via the scrolling area.

[1]Republished with permission of Walter de Gruyter and Company, from i-com, 17, 1, 2018, Foot Interaction Concepts to Support Radiological Interventions, Hatscher et al.; permission conveyed through Copyright Clearance Center, Inc.

4.4.3 Evaluation

The presented concepts are compared in a user study. The study follows a within-subject design. Task completion time, subjective workload and subjective user experience were measured. Each of the ten participants performed ten measured tasks for each of the three concepts, leading to 300 data points ($10 \times 10 \times 3$).

Participants

Ten male participants recruited from the local university between 25 and 30 years (M = 26.2, SD = 1.8) took part in the study. In a questionnaire on previous experience with foot interaction, one participant rated himself as very experienced on a 5-point Likert scale from 0 (no experience) to 5 (very experienced). Two participants reported medium previous knowledge (3). The remaining participants stated no experience. Shoe size ranged from 42 to 49 Paris points (M = 44.1, SD = 2.3).

Apparatus

Based on the requirements of the concepts, a prototypical, shoe-mounted input device was created to measure foot rotation and whether the tip of the foot is lifted or lowered (see Figure 4.14). Two measurements are gathered by the hardware: the orientation of the foot is assessed using a 3-axis gyroscope (MPU-9250, InvenSense, San Jose, CA, USA) where only readings of the Z-axis are taken into account. Whether the foot is lowered or lifted is assessed using a downward-facing time-of-flight distance sensor at the tip of the shoe (VL6180X, STMicroelectronics, Geneva, Switzerland). Both sensors are queried over an I^2C-Bus by a microcontroller with an integrated Bluetooth low energy stack (RFD22102, RFduino Inc., Hermosa Beach, CA, USA). Two standard AAA batteries serve as power supply. The components are attached to a clog using a velcro fastener, which allows using clogs in different sizes. Detecting lifting and lowering of the foot uses a bi-level threshold approach with an upper threshold of 10 mm above ground and a lower threshold of 5 mm above ground to avoid flickering. Due to the influence of the floor color on the sensor readings, the distance sensor had to be calibrated by setting the value read at a lowered position to zero.

The readings are sent to a laptop running a custom MeVisLab [164] application. The software has three main tasks: It interprets the sensor readings and

[2] Republished with permission of Walter de Gruyter and Company, from i-com, 17, 1, 2018, Foot Interaction Concepts to Support Radiological Interventions, Hatscher et al.; permission conveyed through Copyright Clearance Center, Inc.

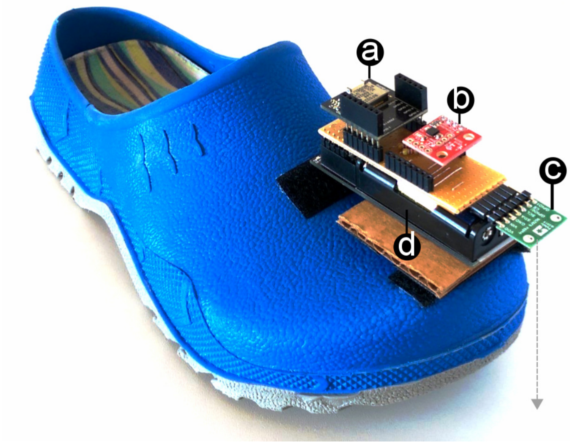


Figure 4.14: An OR clog with the hardware prototype fixed on top. It consists of a microcontroller with integrated Bluetooth stack (a), gyroscope (b), distance sensor (c) and power supply (d). Image from Hatscher, Luz, and Hansen [72] © Walter de Gruyter and Company 2018[2].

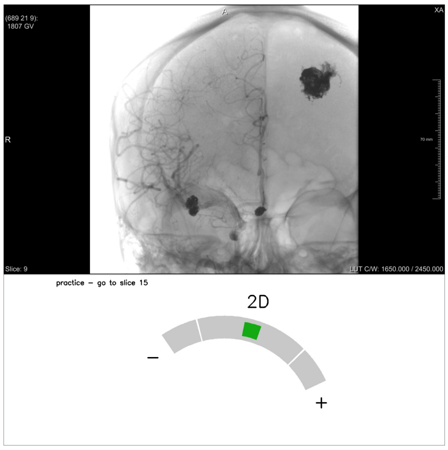


Figure 4.15: GUI including a medical image viewer on top and visual feedback for the foot input method at the bottom. The foot position is displayed as a green cursor. A lifted foot is indicated by a cursor shrinked to 60% of its size. Image from Hatscher, Luz, and Hansen [72] © Walter de Gruyter and Company 2018[2].

maps them to a 1D cursor position using the proposed concepts, it provides visual feedback for the user and it displays the medical image data to be navigated. Additionally, a line of text informs the participant about the goal of the current task during the study. This is to minimize the cognitive effort of remembering the task, which might influence the study results. A screenshot can be found in Figure 4.15. The upper part resembles a medical image viewer for DICOM images, with the number of the current slice displayed in the lower-left corner. The lower part provides visual feedback on input elements. A green cursor depicts the position of the tip of the foot. When the foot is lifted, the cursor shrinks to 60% of its original size. Underlying, gray areas show the location of interactive areas corresponding to Figure 4.13. There was no visual feedback on the floor.

An area of 170 × 105 cm was covered with linoleum, similar to the one used in operating rooms to provide comparable friction between shoes and the floor. At the height of 140 cm at its center, a 40 inch monitor was mounted in front of the floor. Resolution and size were comparable to the large screen display of a Siemens angiography suite, which has the same resolution of 3840 × 2160 pixels and 56 inch.

Measurements

As a measure of performance, the task completion time was assessed. The subjective workload was measured using RTLX [67]. User experience was gathered via the modules I, III, V and the dimension "intention to use" from module IV from the meCUE questionnaire [132]. The second dimension from module IV "product loyalty" and module II with the dimensions "visual aesthetics", "status" and "commitment" were left out as the system was not a finished product, hence ratings for these dimensions were of no interest.

Tasks

During the study, the participants had to navigate to a given slice number in a sequence of 18 radiological images showing the distribution of contrast agent in the blood vessels of the brain. The target slice numbers were identical for all participants. The position was chosen in a way so that scrolling in both directions was required. The current slice number was displayed in the lower-left corner of the image, the target slice in below that in the upper left corner of the visual feedback area. An auditory signal indicated the start of the task. The end of the task was conveyed verbally by the participant. The tasks were designed that way to measure the time until the user finished the task to its satisfaction, as it would be the case in a real-world scenario. To filter out any hesitation in verbal communication, time was measured from the start signal to the last sensor data that indicated foot movement before the participant conveyed task completion.

Procedure

The study took place in a computer laboratory. At first, a questionnaire regarding demographics, shoe size and experience with foot interaction was filled out. The participant selected a pair of clogs out of tree pairs in sizes 41/42, 43/44 and 45/46 Paris points. The sensor described in Section 4.4.3 was then attached to the right shoe. The participant took a position approximately 100 cm from the display but was not restricted to that distance. The task order was identical for all participants. The order of the input methods was randomized. For each concept, the interaction was explained by the instructor, followed by five practice tasks and ten measured tasks. Participants were asked to navigate to a given slice number, using the current interaction concept. Additionally, the target slice number was displayed in the graphical user interface all the time. Before each task, the foot had to be placed in a neutral, centered position. After performing the measured tasks, the RTLX and the subset of questions from the meCUE questionnaire were filled out. This procedure was followed for each input concept.

Results

The statistical results of the ANOVA and post-hoc tests can be seen in Table 4.6.

Table 4.6: Summary of the test statistics for subjective workload (SW), user experience (UX) and task completion time (TCT) for concepts *discrete buttons* (DB), *foot scrolling* (FS) and *step and scroll* (SAS).

Dependent variables	df	F	t	p	sig	η^2_{part}	d	Effect
SW	2, 18	5.93		0.01	*	0.4		large
FS vs. SAS	9		3.67	0.02	*		1.16	large
DB vs. SAS	9		1.59	0.44	n.s.		0.5	medium
FS vs. DB	9		1.88	0.28	n.s.		0.6	medium
UX								
Overall	1,49, 13,41	6.37		0.02	*	0.42		large
FS vs. SAS	9		4.99	< 0.01	*		1.58	large
DB v. SAS	9		2.48	0.11	n.s.		1.11	large
FS vs. DB	9		0.76	1.00	n.s.		0.24	small
TCT	2, 18	12.25		< 0.01	*	0.58		large
FS vs. SAS	9		5.34	< 0.01	*		1.69	large
DB vs. SAS	9		1.70	0.37	n.s.		0.54	medium
FS vs. DB	9		3.06	0.04	*		0.97	large

Significant results are revealed by the ANOVA for TCT. The target images were reached fastest using the *foot scrolling* method (M = 4.03s, SD = 0.88), followed by *discrete buttons* (M = 5.11s, SD = 0.88). The *step and scroll* method performed worst (M = 5.75s, SD = 1.13). Post-hoc test revealed that *foot scrolling* was significantly faster than *discrete buttons* and *step and scroll*.

The results for the six RTLX subscale ratings and the overall workload score are presented in Figure 4.16. The lowest overall workload score is achieved using the *foot scrolling* concept (M = 4.42, SD = 2.47) followed by *discrete buttons* (M = 5.40, SD = 2.96) and *step and scroll* (M = 6.67, SD = 3.28). A statistically significant difference was found between *foot scrolling* and *step and scroll* using a one-way ANOVA and subsequent post-hoc tests.

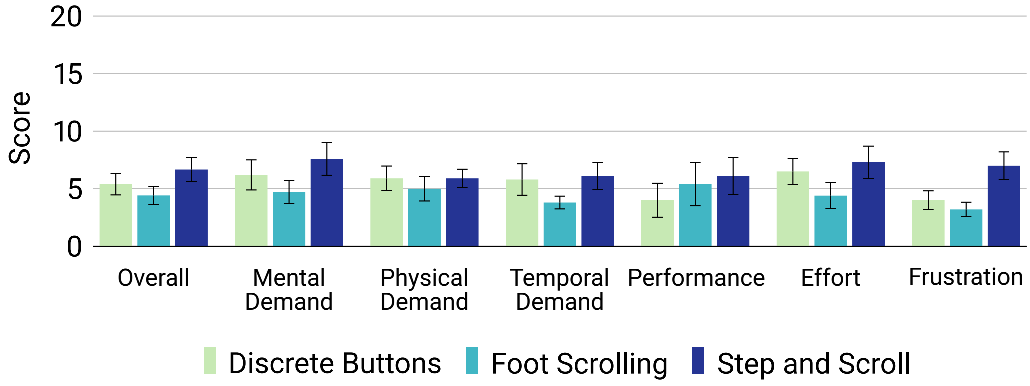


Figure 4.16: Unweighted mean scores for the NASA-TLX questionnaire on a scale from 0 to 20 (0 = low/good performance, 20= high/poor performance). Error bars represent standard deviation. The overall workload is the mean value from all the subscales.

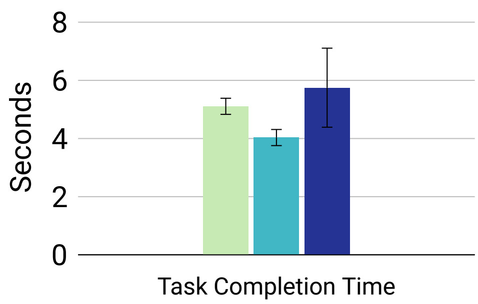


Figure 4.17: Mean results for the task completion times (TCT) in seconds with standard error bars.

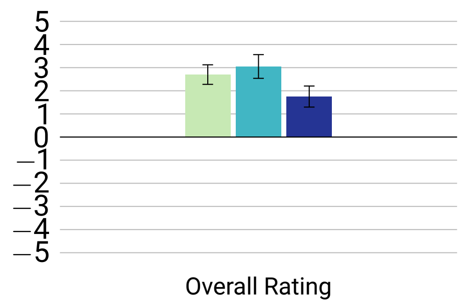


Figure 4.18: Mean results for the meCUE overall rating with standard error bars. (-5 = bad, 5 = good)

Figure 4.19 shows the results for the meCUE dimensions assessed. Analyzing the meCue overall rating score revealed significant ANOVA results (see Figure 4.18). All concepts were rated positively. The concept *foot scrolling* (M = 3.05, SD = 1.62) achieved a slightly better score than *discrete buttons* (M = 2.70, SD = 1.34). *step and scroll* (M = 1.75, SD = 1.44) scored worst in the meCUE overall rating. A significant difference between the concepts *foot scrolling* and *step and scroll* was found via post-hoc tests.

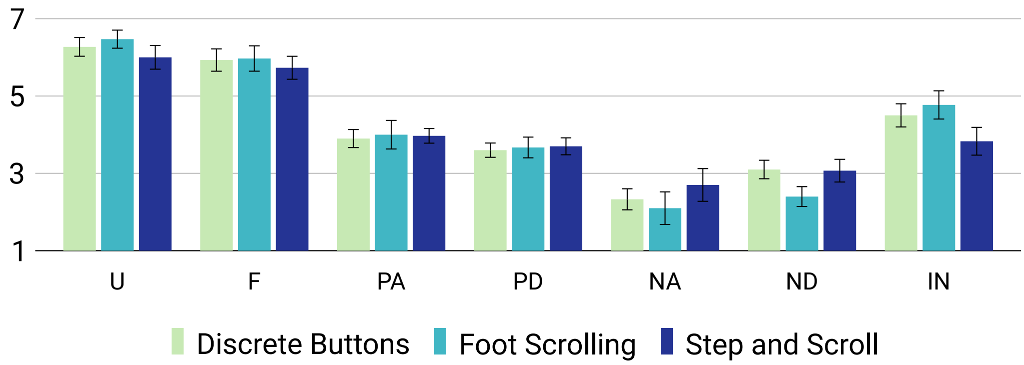


Figure 4.19: Mean results for the meCUE dimensions Usability (U), Usefulness (F), Positive Emotions (PA, PD), Negative Emotions (NA, ND) and Intention to use (IN) with standard error bars. (1 = strongly disagree, 7 = strongly agree)

4.4.4 Discussion

In this section, three methods for using heel rotation to control medical image data are presented. A user study revealed that the *foot scrolling* concept is superior in terms of task completion time, subjective workload and user experience. This is interesting as lifting and repositioning the foot when using relative input requires repeated movements, which is potentially exhausting [188]. In contrast to these findings, Alexander et al. found displacement-based interaction slower than rate-based approaches for standing positions [7]. The results might be caused by the familiarity of sliding gestures, which are similar to hand gestures employed for touchscreen devices. This explanation is supported by the RTLX dimensions Mental Demand and Effort, which are relatively low for *foot scrolling* while Physical Demand differs less compared to *step and scroll* and *discrete buttons*. The concept *step and scroll* was rated worst in terms of subjective task load. This is evident for Mental Demand, Temporal Demand, Effort and Frustration. Observations suggest that this is due to the border between sliding and pedal-like behavior. As there is no haptic feedback or restriction, participants hat to pay close attention to where precisely the cursor is located. Otherwise, reaching the end of the scrolling area easily leads to involuntary rate-based interaction.

However, the proposed input methods can meet almost every requirement listed in Section 4.4.1. A static standing posture is avoided as lifting the foot triggers no functionality. The empty zone of the *discrete buttons* allows moving the foot without interacting with the system. The upper threshold is well below 25° as the tip of the foot has to be lifted 10 mm off the ground.

The activation force is below 10 N, as there are no physical parts to be pressed. Dorsal flexion is kept to a minimum, whereas *foot scrolling* requires it the least. Due to the visual feedback, the input device can be controlled without looking at the feet. The chance of accidentally activating the wrong function is small as the cardinal direction can be easily distinguished by the direction the foot is rotated. This approach is similar to the disc-like pedal presented by van Veelen et al. As the device is mounted on the clog, losing the pedal is impossible.

The prototypical input device is theoretically able to recognize the use of conventional foot pedals by detecting the dorsal flexion angle using the gyroscope and comparing it to the measures of the distance sensor. In this case, interaction should be deactivated to avoid interference with the existing workflow. Walking around should also deactivate the system. This is not implemented in the proposed system, but a preliminary test suggests that a negative angle for dorsal flexion by lifting the heel is a reliable indicator for an interruption. This can be supported by a second distance sensor at the back of the clog. Interventions can last several hours, which might cause fatigue in legs or feet. Further, hospitals use different flooring and clogs, which might prevent sliding. Reinschluessel et al., therefore, implemented scrolling in a way that requires lifting the foot prior to rotation [162].

Even though this chapter adds to the body of knowledge by comparing different foot input methods in a lab study, questions regarding interplay with environmental factors are not answered exhaustively. In the OR, the cognitive workload is, among others, influenced by verbal communication with assistants, consultation of colleagues, alerts and interaction with auxiliary systems. Therefore, the influence of manual tasks on multimodal input techniques is investigated in Chapter 6.

4.5 Conclusion

In this chapter, foot-based human-computer interaction methods for medical image manipulation have been investigated. Section 4.1 and 4.2 described approaches for sensor floors. The initial problem of utilizing a sensor floor as an interface for a single user is tackled in section 4.1 by placing virtual buttons dynamically around the user, based on its position and foot distance. Additionally, a graphical user interface allowed to focus on the screen by providing visual feedback on feet and floor positions and on gesture recognition. Section 4.2 built upon the previous section but aimed to reduce the required space by using small steps and different combinations of weight on heels and balls of both feet. Both sections revealed important issues when using sensor floor data on a technical as well as on a theoretical level. Foot gestures such as triple-taps are not very common, and users might perform them very fast or forcefully and slow. Systems, therefore, need a reasonably fast recognition approach, but time windows for gesture detection, on the other hand, have to take slow execution into account (see Section 4.1).

Special care has to be taken regarding the resolution of sensor systems as people tend to shift their position when using their feet as described by Saunders and Vogel [175], and taking big steps causes higher physical strain (see Section 4.2). Section 4.3 and 4.4 take a different approach and focus on input methods utilizing in-place foot movements and weight shifting. Using plantar pressure for human-computer interaction yields relatively low mental demand, which might be beneficial to support demanding tasks. The user study revealed the need for a personalized user interface due to different weight and pressure distributions as well as preferences. This input method seems to be suitable for simple input tasks but relies on contact to the floor, which naturally excludes mid-air foot movements. Section 4.4 compared different mappings for heel rotation onto a one-DoF task and found foot movements resembling sliding on a touchscreen superior to the proposed alternatives. Further, the approach is oriented on criteria for ergonomic foot pedal design in clinical applications [204].

Foot-based input techniques and methods, in general, proved to be suitable for simple tasks. Each of the presented approaches bears advantages and downsides, which have to be considered in designing future user interfaces for the OR. Sensor-equipped floors allow input without additional preparation and equipment for the user. Distinguishing users, on the other hand, is difficult, which might be an important factor in a hierarchical working environment. Depending on the resolution of the floor, gestures might be challenging to detect when the user is placed between sensor cells. Interaction methods based on plantar pressure inlays do not require additional infrastructure and allow simple commands, while overlapping pressure patterns such as a toe and the ball of the foot are difficult to distinguish. On the other hand, personalization according to personal preference and hierarchical level should be possible. Measuring the angle of the foot on the other and allows gestures composed wholly or partly out of mid-air movements such as taps, heel rotation, tilting the foot or kicks, which can be tuned for intuitive due to similarity to sliding gestures or physical foot pedals. In general, it can be said that there is no such thing as a perfect foot input device or approach as it depends on the requirements of the task and environmental restriction which method is most suitable.

This chapter is based on the following publications:

B. Hatscher, M. Luz, and C. Hansen. “Foot Interaction Concepts to Support Radiological Interventions”. In: *Mensch und Computer 2017 - Tagungsband*. Regensburg: Gesellschaft für Informatik e.V, 2017, pp. 93–104.

B. Hatscher, M. Luz, and C. Hansen. “Foot Interaction Concepts to Support Radiological Interventions”. In: *i-com* 17.1 (2018), pp. 3–13.

B. Hatscher, S. Wagner, L. Grimaldi, M. Fritzsche, N. Elkmann, and C. Hansen. “Navigation in medizinischen Bilddaten mittels eines taktilen Fussbodens”. In: *Proceedings of the Annual Meeting of the German Society of Computer- and Robot-Assisted Surgery (CURAC)*. Bern, Schweiz, 2016, pp. 33–38.

A. Solovjova, D. Labsch, B. Hatscher, M. Fritzsche, and C. Hansen. “Plantar pressure-based gestures for medical image manipulation”. In: *Mensch und Computer 2018 - Tagungsband*. Bonn: Gesellschaft für Informatik e.V, 2018, pp. 421–425.

S. Wagner, B. Hatscher, M. Luz, B. Preim, and C. Hansen. “Konzepte mit unterschiedlichen Platzanforderungen zur Navigation in medizinischen Bilddaten mittels eines Sensorfußbodens”. In: *Proceedings of the Annual Meeting of the German Society of Computer- and Robot-Assisted Surgery*. Hannover, Germany, 2017, pp. 220–225.

The responsive interface for tactile floors described in Section 4.1 was implemented and evaluated by Sebastian Wagner and Luigi Grimaldi, and the plantar pressure input method described in Section 4.3 was implemented and evaluated by Alina Solovjova and Dominic Labsch within student projects. The foot interaction concept for confined spaces described in Section 4.2 was implemented and evaluated by Sebastian Wagner within his master’s thesis [215]. These works were supervised by the author of this thesis.

The publication ”Foot Interaction Concepts to Support Radiological interventions” [71] received the Best Paper-Award at the conference Mensch und Computer 2017, Regensburg, Germany. Upon invitation, an extended version was published at the i-com Journal of Interactive Media [72].

Section 4.4 is based on *Foot Interaction Concepts to Support Radiological Interventions*, Hatscher, Luz, and Hansen [71], from i-com, 17, 1, 2018, © Walter de Gruyter and Company 2018

5

Multimodal, Hands-Free Interaction

The capabilities and limitations of touchless hand gestures and foot input methods were shown in the previous chapters. In a complex scenario, however, a single input channel might suffer from several disadvantages: they might be overloaded, not the perfect choice for the task at hand, or the placement of sensor equipment is not optimal. The use of multiple input modalities might solve this issue.

This approach is investigated in two steps: At first, a deeper understanding of modalities is gained by comparing hand, foot and voice input in a simplified image interaction task. Second, different combinations of eye gaze and foot input are proposed for hands-free coarse pointing and manipulation.

5.1 Alternative Touchless Input Modalities

During an intervention, each potential human-computer input modality can be occupied: The hands may be fully engaged in performing the actual intervention [95, 130, 141], conventional, physical foot pedals might render the feet temporarily unavailable for additional interaction tasks and voice commands might interfere with verbal communication between medical staff. A solution to this problem might lie in the availability of alternative input methods to adapt to different situations. It has already been shown that, in the OR, hand gestures are preferred over voice input for situations when image manipulation is performed to support a discussion [125]. Therefore, the input modalities *hand, foot and voice* are compared in an image manipulation task to provide alternatives in case a specific modality is not available. To do so, for every step of an abstract image manipulation workflow, suitable input techniques according to literature were implemented. A user study with 19 medical students in the fourth year of studies or higher who had experience in the OR was conducted. The input techniques were compared in terms of task completion time, perceived usability, perceived usefulness and an overall rating. Qualitative feedback was gathered by conducting post-test interviews.

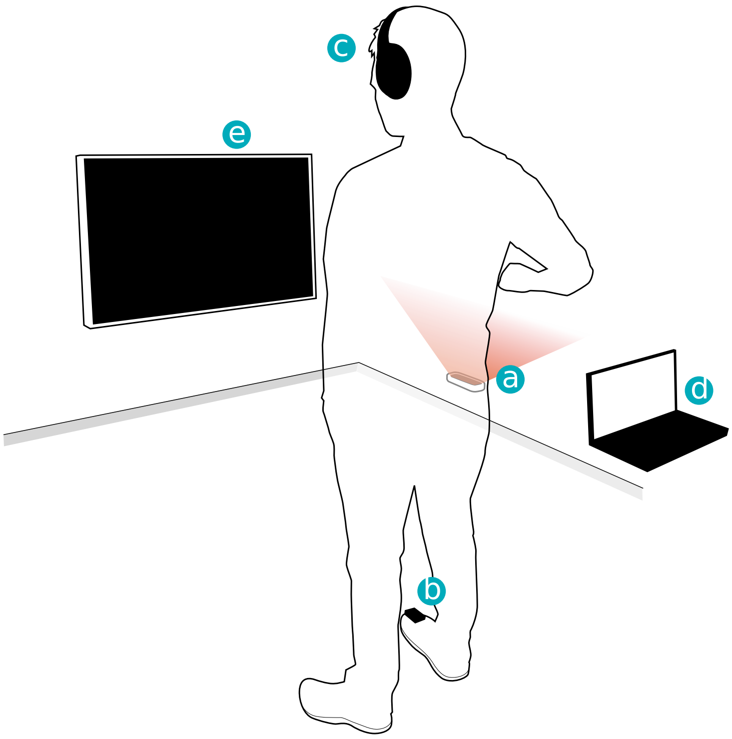


Figure 5.1: Technical setup of the user study: Hand gesture sensor (a), inertial measurement unit attached to a shoe (b), headphones with integrated microphone (c). A laptop (d) processed the data and ran the medical image viewer to be controlled on a large screen (e). Image from Hatscher and Hansen [70].

5.1.1 Workflow for Medical Image Manipulation

The overall abstract workflow to manipulate image data during interventions consists of the following steps:

- System activation: Change the state of the system, so that images can be manipulated.
- Function selection: Choose which kind of manipulation should be performed.
- Image Manipulation: Interact with the data.
- Deactivation: Put the system back into an inactive mode.

Even though all steps are required to interact successfully, activation and manipulation are especially important as different requirements collide. To illustrate the reasons behind this, their role during interventions is described in the following.

A suitable system activation method is difficult to find because a middle ground between the two contradicting requirements of easy access and prevention of involuntary input has to be found. The physician handles medical instruments but gets indirect visual feedback by using an X-ray imaging device or a laparoscopic camera. When the input is detected continuously, involuntary manipulation might happen quite easily. This might lead to decisions based on wrong information or requires extra time to revert these changes. Hence, the user interface needs to be robust against involuntary activation. On the other hand, the system needs to be activated every time, which requires easy and fast input methods to not lengthen interventions unnecessarily.

Further, different levels of precision are required, depending on the situation. Radiologists might navigate back and forth through a stack of images in fast succession to understand the structure of arteries by observing the distribution of contrast agent [95] or look for a specific image that provides the best overview to be used as "road map" [78]. While the first use case requires fast but less accurate control, the second one demands an accurate selection of a single image.

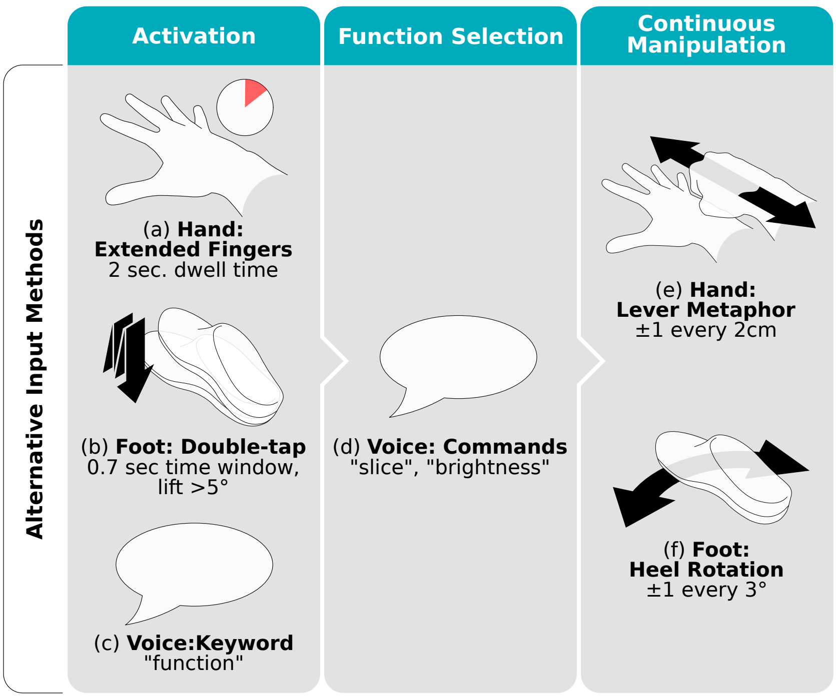


Figure 5.2: Workflow steps and the input methods provided for each step. Image from Hatscher and Hansen [70].

5.1.2 Suitable Input Methods

A variety of input methods for similar tasks is proposed in literature. In the following, the most relevant approaches are discussed regarding their suitability for the steps ”activation”, ”function selection” and ”continuous manipulation”. Mentis et al. used a very similar sequence of tasks with an initialization keyword, an event command and voice commands or hand gestures for manipulation [125]. Aspects of this workflow are further part of existing research such as activation [119], function selection [25] or image manipulation use one DoF [162, 230].

Activation

First tests by Mauser and Burgert suggested that a flat hand gesture with at least four extended fingers and palm down with a dwell time of three seconds avoids unintended activations [119]. Based on these findings, a hand dwell gesture was designed to fit the needs of the task. To make it more comparable to voice and foot input which do not require dwelling, the dwell time was shortened to two seconds. The number of fingers to be extended was increased to five to provide a well-defined static gesture.

Voice control is considered suitable for discrete commands [141] and was used for system activation in a clinical context [142, 141, 125]. Ebert et al. used the keyword ”change” to initiate voice commands [40]. Such a keyword provides a semantically meaningful connection to the subsequent command. Similar to that approach, activation via voice is done using the ”function” keyword while the following voice command selects the function to be used.

Heo et al. investigated ”one-bit input” methods for hands-busy situations using the forearms, bouncing, blowing and foot-tapping [76]. Given the medical context, the forearms are most likely sterile, blowing is prevented by facial masks and bouncing is difficult when performing delicate manual tasks during interventions. This leaves foot-tapping as a potentially suitable alternative. Short, slow foot movements while adjusting the stance [175] might be easily confused with a single-tap. Therefore, a double-tap as already introduced in Chapter 4, was employed.

In summary, the approaches implemented for sterile system activation using the feet, hands or voice were:

- **Hand: Extended fingers**, hold one hand palm-down with extended fingers for two seconds (Figure 5.2a).
- **Foot: Double-tap**, execute a double-tap with the ball of the foot (Figure 5.2b).
- **Voice: Keyword**, utter the activation keyword ”function” (Figure 5.2c).

Function Selection

Following the activation, a function has to be selected. The method for this task needs to deal with a list of arbitrary length, depending upon the number of functionalities implemented. The system presented by Ebert et al., for example, provides a choice of fourteen functionalities [40]. Voice input is considered the most appropriate input method for this kind of task as it allows to direct uttering the targeted information [125, 142], which is not possible by hand or foot gestures. For this reason, function selection is made by voice commands only.

Continuous Input

For medical image manipulation using hand gesture input, Hettig et al. compared input via Myo armband, Leap Motion Controller, Joystick and verbal task delegation. The Myo armband performed the fastest for 2D image manipulation while 3D rotation was accomplished fastest with verbal task delegation [78]. Single-handed and double-handed gestures for image manipulation were proposed, including image stack navigation, windowing, pan and zoom [40, 53, 196].Mentis et al. reported that clipping plane manipulation using hand gestures was especially beneficial as it allows simultaneous discussions compared to voice commands [125]. Reinschluessel et al. found no significant difference between hand gestures, foot gestures and verbal task delegation for navigating medical data in an virtual OR [162]. Aligning with these results, Zaman et al. reported comparable results for image scrolling by foot and verbal task delegation in terms of task completion time and usability [230].

Performing continuous input by voice was investigated by Igarashi and Hughes. The voice was used as an on/off switch; the pitch of the voice allowed rate-based continuous input and series of peaks such as syllables triggered discrete commands [84]. This input method is deemed unnatural [84] and prevents natural speech, which makes it less suitable for the targeted use case. Another approach for voice-controlled continuous input maps vowel sounds to directions, allowing motor-impaired users to gain control over a cursor. The system reached an index of performance of 0.3 compared to 1.0 for a mouse in a Fitts law study [65]. In the medical domain, voice control is considered not suitable for continuous commands [141] as it inhibits verbal communication [125]. Therefore, voice input was discarded for continuous input.

Based on these findings, two methods to continuously interact with the system were developed. For hand gesture interaction, a lever metaphor is used.

When the user closes the hand inside the sensor area, the value corresponding to the selected function is coupled to the hand's position. Moving the fist back and forth changes the value accordingly. The coupling is released when the hand is opened or leaves the sensor area. Fine-grained control during image stack navigation is possible as the image increments or decrements every two centimeters the hand is moved. Foot input builds upon the *foot scrolling* method investigated in Section 4.4. While the foot in on the ground, rotating around the heel changes the value every 3° the foot is rotated. When the tip of the foot is lifted, repositioning does not influence the target value. The implemented methods for manipulation of a continuous value are:

- **Hand: Lever metaphor**, move a closed fist back and forth. Movements with an open hand are ignored (Figure 5.2e).
- **Foot: Heel rotation**, rotate the foot on the heel while the ball of the foot is lowered. Movements with the ball lifted are ignored (Figure 5.2f).

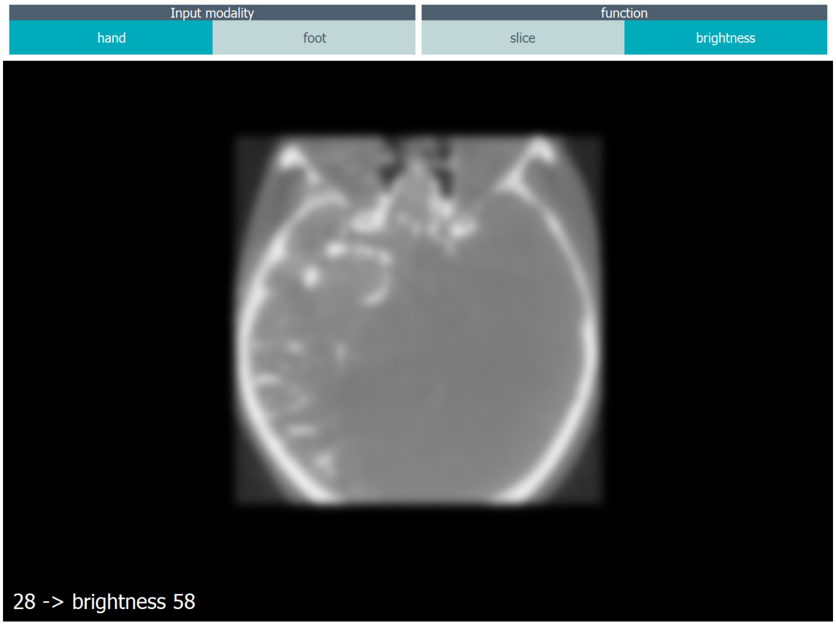


Figure 5.3: Graphical user interface of the medical image viewer. The top bar shows the currently used input modality and the selected functionality. On the lower left, the current value, the target function and the target value were displayed during the study. Image from Hatscher and Hansen [70].

5.1.3 Evaluation

The outlined approaches offer alternative ways to control medical image data in case a specific input channel is not available. However, it is vital to understand how potential future users perceive these techniques. For this reason, a comparative user study with medical students was conducted. Task completion times as a measure of performance, usefulness and usability were assessed.

Participants

Nineteen students majoring in human medicine (13 female, 6 male) at ages ranging from 22 to 31 years (M = 24.9, SD = 2.2) participated in the study. All participants were in their fourth year of studies or higher. They were recruited via a mailing list. The average prior experience with hand gesture interaction was low (M = 1.2, SD = 0.5 and M = 1.4, SD = 0.7 on a 5-point Likert scale where 1 = no prior experience, 5 = very experienced). Two participants were left-handed and left-footed. Shoe size ranged from 37 to 44 Paris points (M = 40.3, SD = 2.1).

Apparatus

A monitor with a resolution of 3840 × 2160 pixels and a diagonal of 40" was positioned at the height of 85.5 cm, approximately 150 × 150 cm of free space were provided for the participant. An overview of the technical setup is given in Figure 5.1. A Leap Motion Controller was used for hand gesture detection and placed on a table at the right-hand side. An MTw Awinda Wireless 3DOF Motion Tracker (Xsens Technologies B.V., Enschede, Netherlands) attached to a clog was used to gather foot movements. QuietComfort 35 II Headphones (Bose GmbH, Friedrichsdorf, Germany) with an integrated microphone were used for voice input and to deliver auditive signals. For speech recognition, a lightweight speech recognition engine based on the CMU Sphinx Natural Language Processing library called PocketSphinx was integrated [83, 158]. The underlying dictionary was customized and only contained the key phrases used in the study, which were "function", "slice" and "brightness".

A prototypical software processed the sensor data, implemented the interaction concepts, and provided a graphical user interface consisting of a medical image viewer and visual feedback (see Figure 5.3). Foot taps were recognized by reading the pitch of the sensor and applying a double-threshold approach. For calibration, the shoe was placed on the floor, and the sensor pitch value was used to represent zero degrees dorsiflexion. A tap corresponded to crossing the upper threshold of dorsiflexion of 5° followed by crossing the lower threshold of 3°. When this pattern was detected twice within 700 ms, a double-tap was registered. Tap recognition was disabled or interrupted when the pitch of the sensor went below zero, indicating that the user lifted the heel and is walking instead of controlling the software. To minimize additional cognitive workload during the user study and to avoid interruptions, the target of the current trial was displayed in the software.

Study Design

The study followed a within-subject design. The input techniques described in Section 5.1.2 were not evaluated individually but in a succession that might happen alongside a medical intervention. This aims at motivating the participants to think of the proposed techniques as part of a clinical routine. Therefore, the methods were combined into four conditions that cover activation, function selection and continuous manipulation . According to the literature, not all methods were considered suitable for each task, as described in Section 5.1.1. Therefore, function selection was made using voice commands in each condition while it was not used for continuous manipulation. An overview of all conditions can be found in Figure 5.4. Each participant performed two training trials and four measured trials for each of the four conditions. The order of the conditions was counterbalanced using a Latin square to minimize learning effects. The tasks were not randomized to maintain comparability. The workflow always consisted of (1) activation of the system (2) selection of a functionality and (3) manipulation of a continuous value until the target was reached. To avoid early starts, the system did not respond until a trial started and stopped responding after a trial ended.

Measurements

Task completion time was assessed for two subtasks: between the start of a task and successful activation of the system and between system activation and reaching the target value, which marked the end of the task. Subjective measures were gathered using two modules from the meCUE questionnaire [132]. Module I assesses the perception of instrumental qualities and consists of six questions to be answered on a 7-point Likert scale. Module V assesses overall evaluation on a scale from -5 to 5 at a scale interval of 0.5. As meCUE was designed to measure the user experience for a product, the questionnaire description and questions were adapted to ask for ”interaction technique” instead of ”product” to avoid confusion during the user study. The trials and post-test interviews were recorded on video. The video logs were analyzed to gather error rates. The participants’ opinions on the proposed interaction techniques, especially in the context of interventional scenarios, were gathered during the post-test interviews.

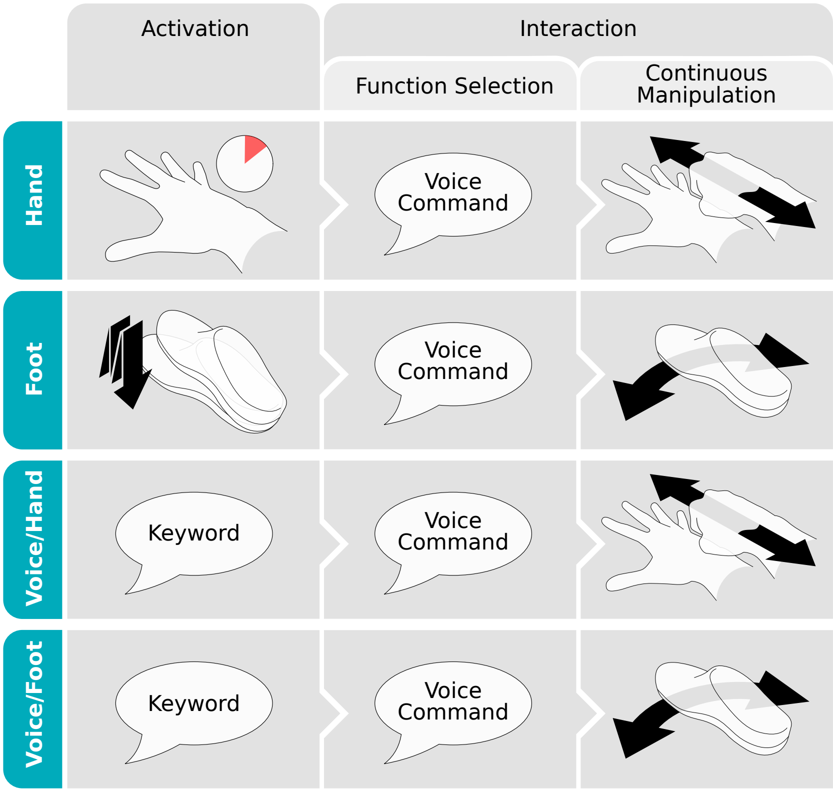


Figure 5.4: User study conditions and separation of the workflow steps in activation and interaction phase. Image from Hatscher and Hansen [70].

Procedure

The study took place in a computer laboratory. In the beginning, the participant filled out a questionnaire assessing demographic information, dominant hand, and foot as well as experience with hand and foot interaction. Out of three pairs of clogs, the best matching one was selected, and the motion sensor was attached. The shoes were available in 41/42, 43/44 and 45/46 Paris points. The first of the four randomized conditions was explained and demonstrated. The participant performed two non-measured practice trials and four measured trials. During practice, remaining questions regarding the current interaction method were clarified. An auditive signal (2000 Hz, 0.7 seconds), transmitted via the headphones, indicated the beginning of a task.

The system did not respond before the signal. In each trial, the participants had to activate the system, select the given function and modify the default value to match a target value. The target function to be selected, and the target value was displayed in the lower-left corner of the user interface, right next to the current value (see Figure 5.3). Reaching the target value was detected automatically. To avoid extensive correction loops but at the same time prevent intentional overshooting for fast task completion, the controlled value had to match the target value closely (± 1) for two seconds. A sound signaled task completion, and the system stopped responding. All parameters were set to fixed default values at the end of each task. After each condition, the usability, usefulness and overall rating was assessed twice using module I and V of the meCUE questionnaire, for the activation phase and the interaction phase, consisting of function selection and continuous manipulation. This procedure was repeated for each of the conditions.

Results

During the study, measures were taken separately for activation and for the interaction phase. Interaction includes the steps "function selection" and "image manipulation" while function selection was performed using voice commands in every condition.

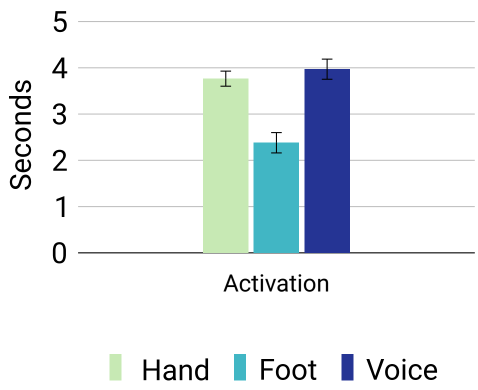


Figure 5.5: Mean task completion time for the activation step in seconds. Error bars show standard error.

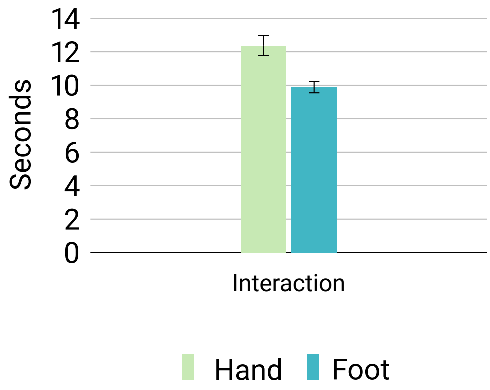


Figure 5.6: Mean task completion time for the interaction phase in seconds. Error bars show standard error.

Table 5.1: Summary of the test statistics for task completion time.

	df	**F**	**t**	**p**	η^2_{part}	**d**	**Effect**
ANOVA							
Methods	3, 54	14.75		<.05	.45		large
Task	1, 18	306.73		<.05	.95		large
Interaction	2.34, 42.15	10.43		<.05	.37		large
t-tests							
Activation							
Hand vs. Voice	18		.72	.48		.16	small
Hand vs. Foot	18		4.88	<0.125		1.12	large
Foot vs. Voice	18		5.34	< 0.125		1.23	large
Interaction							
Hand vs. Foot	18		5.16	< 0.125		1.18	large

Task Completion Time In general, participants were able to perform the tasks fastest for the ”foot” condition (M = 6.22 s), followed by ”voice/foot” (M = 7.02 s) and ”voice/hand” (M = 8.00 s). The condition ”hand” performed slowest (M = 8.07 s). Task completion time was analyzed using a 4 × 2 ANOVA (conditions: hand, foot, voice-hand, voice-foot × tasks: activate, interact). It revealed a significant main effect for condition and task as well as a significant interaction effect for condition × task (see Table 5.1). In case the assumption of sphericity was violated, a Greenhouse-Geissner correction of degrees of freedom was applied.

To understand the input methods in detail, they were further statistically analyzed in the context of the workflow step activation and interaction separately. For activation, hand input using the ”extended fingers” gesture, foot input to perform a double-tap and voice commands using the keyword ”function” were differentiated (see Figure 5.2). As voice activation was used in two conditions, these values were pooled. Three t-tests for paired samples were performed and revealed a significant (p <0.05) advantage when using the foot activation method compared to hand or voice activation (see Figure 5.5). The double-tap was the fastest activation method (M = 2.38 s), while no significant difference was found between the extended fingers gesture (M = 3.77 s) and the ”function” keyword (M = 3.97 s). For interaction, which includes the subtasks function selection and continuous manipulation, there were only two different input methods. Therefore, values across the conditions using the hand gestures and across the ones using foot gestures were pooled. Both methods were compared with a t-test for paired samples. Significantly lower task completion times for the heel rotation method (M = 9.89 s) were revealed by the analysis compared to the lever metaphor method (MD = 12.37 s, see Figure 5.6).

Usability, Usefulness and Overall Evaluation Subjective ratings for usability, usefulness and overall evaluation were gathered and analyzed for both the subtasks activation and interaction separately. Foot input scored slightly higher than hand or voice as an activation method for each of the measures. As an interaction method, foot-based heel rotation scored slightly higher than the hand-based touchless lever method.

Error Rates Recognition error rates were gathered by analyzing video logs. Due to file corruption, only 16 out of the 19 datasets could be analyzed. This was not related to the user or system performance; hence the impact on the validity of the error rate results is deemed low. Due to the controlled lab environment, there were almost no external factors influencing the participant or the input devices. Therefore, false-positive error rates for each of the input methods were less than 1%. The error rates reported in the following thus refer to false-negative errors.

Overall voice commands were not recognized correctly in 26% of all cases. Individual error rates for each of the commands were 35.3% for ”function”, 26.9% for ”slice” and 10.5% for ”brightness”. The error rate for the extended fingers gesture for activation by hand gesture was 8.9% and the double-tap was not recognized in 6.3% of all cases.

Post-test Interviews In the post-test interviews, impressions of the participants after using the proposed input methods were assessed. Five of the nineteen participants described the foot-tapping method as convenient for system activation. P02 and P13 said that they like to use the feet because various devices in the OR, such as angiography suites or electrocautery devices, are already operated with the feet using foot pedals. P09 could imagine using the feet more often in the OR. P02 expressed enthusiasm towards more elaborate foot and leg input methods such as tapping, foot tilting or knee movements. Five participants deemed voice commands less suitable due to the noise level in the OR and the unnatural way of speaking. P13 raised concerns that voice input might not be suitable due to the conversations that take place during interventions. According to P19, people become impatient when voice recognition does not work, and some surgeons would not use it if a second try is required. Despite these concerns, four participants named voice as the most convenient activation method.

The overall picture for function selection and continuous manipulation was more diverse. Hand gestures were considered unsuitable for the OR as the hands are busy and at risk of breaking asepsis when gestures are performed. According to P02, a physician who wants to view additional image data definitely would not have their hands free at that moment. Hand gestures were reported as the most straightforward method by four participants, foot gestures by three participants.

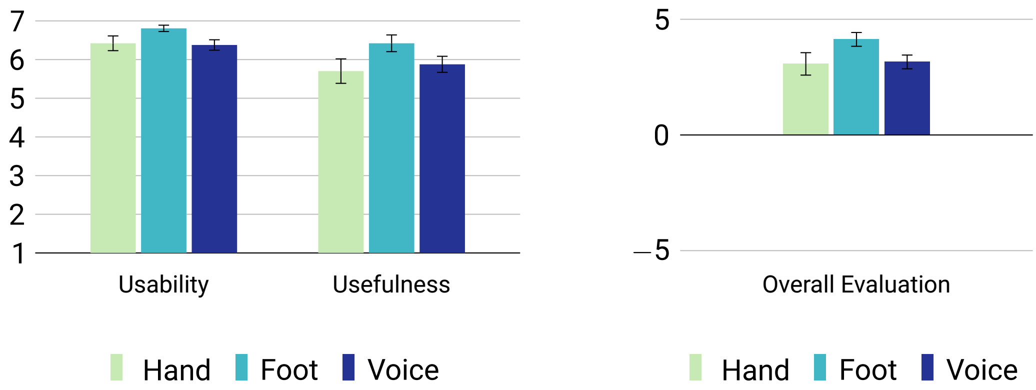


Figure 5.7: Mean results for the meCUE dimensions usability, usefulness (1 = strongly disagree, 7 = strongly agree) and overall evaluation (-5 = bad, 5 = good) during the activation phase.

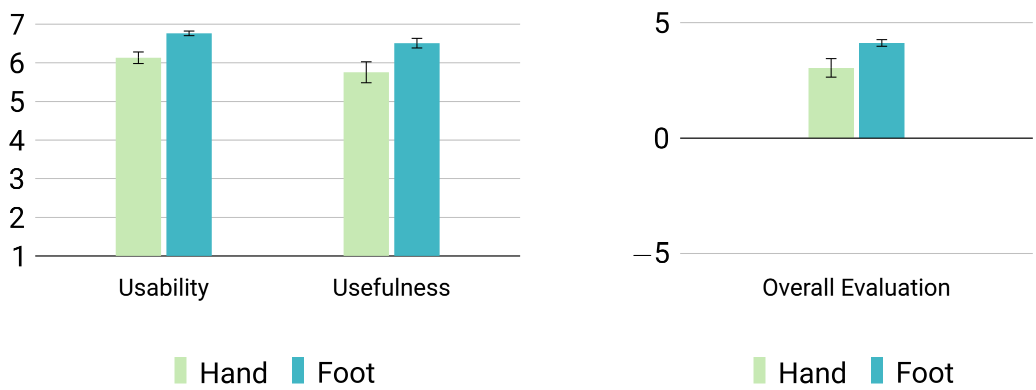


Figure 5.8: Mean results for the meCUE dimensions usability, usefulness (1 = strongly disagree, 7 = strongly agree) and overall evaluation (-5 = bad, 5 = good) during the interaction phase.

Table 5.2: Summary of the test statistics for Module I (Usability, Usefulness) of the meCUE questionnaire.

	df	**F**	**t**	**p**	η^2_{part}	**d**	**Effect**
ANOVA							
Methods	1.92, 34.63	5.40		<.05	.23		large
Task	1, 18	.19		.67	.01		small
Interaction	3, 54	5.92		<.05	.25		large
t-tests							
Activation							
Hand vs. Voice	18		.30	.77		.07	no effect
Hand vs. Foot	18		2.16	<0.5		.62	medium
Foot vs. Voice	18		2.26	<0.5		76	medium
Interaction							
Hand vs. Foot	18		3.35	< 0.5		1.06	large

Table 5.3: Summary of the test statistics for Module V (overall evaluation) of the meCUE questionnaire.

	df	**F**	**t**	**p**	η^2_{part}	**d**	**Effect**
ANOVA							
Methods	1.87, 33.72	3.55		<.05	.17		large
Task	1, 18	.64		.43	.04		medium
Interaction	3, 54	3.83		<.05	.18		large
t-tests							
Activation							
Hand vs. Voice	18		.18	.86		.05	no effect
Hand vs. Foot	18		2.12	<0.5		.59	medium
Foot vs. Voice	18		2.42	<0.5		75	medium
Interaction							
Hand vs. Foot	18		2.38	< 0.5		.69	medium

5.1.4 Discussion

Direct input concepts that consider several restrictions in sterile environments were presented in this section. The input methods aim to provide alternatives for common physician computer interaction tasks. This allows direct control even when one input channel is occupied, for example, when the hands are busy with medical tasks or the feet are needed to operate pedals. According to literature and depending on the task type, suitable input methods were determined and adapted to fit a prototypical, abstract physician-computer workflow. The proposed interaction methods were evaluated in a user study. Task completion times, overall subjective rating, usability and usefulness were assessed as well as qualitative data during post-test interviews. In scenarios where direct input must be provided at all times, there might be one primary input modality and alternatives in case the main input channel is occupied. Foot input performed significantly better than the alternative hand and voice input methods. Further, post-test interviews revealed a tendency towards foot input as it felt most convenient and seems to interfere the least with the existing environment in the OR. The reason for this might lie in the fact that foot control in the form of pedals is common for medical devices. Users with a medical background might prefer this input channel due to prior knowledge.

A. Zaman et al. compared voice and foot input for navigating radiological images in a virtual operating room [1]. In contrast to the presented results, they found no significant differences between both input methods in terms of task completion time and subjective workload. The reason might lie in the different voice recognition software or the input technique as A. Zaman et al. used forward and backward dragging of the foot for image navigation.

Limitations

The hand gesture for activation uses dwell time, which influences the task completion time. A time-independent activation method might lower the task completion time for this approach.

Speech recognition is under active research, resulting in various technical solutions with different restrictions and error rates. Commercially available voice assistant systems often perform the actual processing on a remote server [81]. As these systems have to listen all the time for a keyword and generate data that can get stolen, privacy concerns arise [81]. Open-source solutions, on the other hand, allow full control over the acquired data but are bound to the local infrastructure and require additional effort for results that can compete with commercial solutions. The speech recognition setup used in this study, therefore, might not represent an optimal configuration. It can be estimated that reliable and accurate speech recognition is possible even on moderate hardware in the future.

The study was designed in a way that focused on a given workflow context. Therefore, only input modalities that were deemed suited to the tasks by the literature were investigated as input methods. Outside the medical scope, the range of input channels and methods may be extended as less restricting factors come into play. The presented approaches and their evaluation, however, indicate that hand, voice and foot perform similarly for image manipulation scenarios and maintain the ability to directly interact even when one of these input channels is occupied. The question of additional workload emerging from directly controlling a computer while performing a manual task is investigated in Chapter 6.

5.2 Multimodal Gaze and Foot Input

In clinical routine, physicians take a look at the screen that displays the information required and return to the medical, manual task at hand. The approach proposed in this section is motivated by this natural interaction. Gaze input for direct interaction suffers from the Midas Touch Problem, which describes unintended interactions that emerge from the double role of the eyes when gaze is used for information retrieval and input [92]. Therefore, foot gestures are introduced to overcome this limitation. Foot input has been combined with eye gaze in the context of desktop computer workplaces to free the hands for other tasks [56, 103, 35]. In the presented approach, the input modalities eye gaze and feet are combined in different configurations to allow manipulation of medical images in a hands-free fashion and evaluated in two user studies.

5.2.1 Interaction Tasks during Radiological Interventions

During minimally-invasive procedures in an angiography suite, several images might be presented on a large screen. These images show fluoroscopy images from different detectors (i.e., from different angles) [52, 78], from different points in time or processed visualizations such as 3D datasets [78]. Additional to system activation, function selection and manipulation of medical image data (see Section 5.1.1), a method for selection of a viewport is required. This task can be broken down in the following subtasks:

- Choose a viewport
- Confirm the viewport

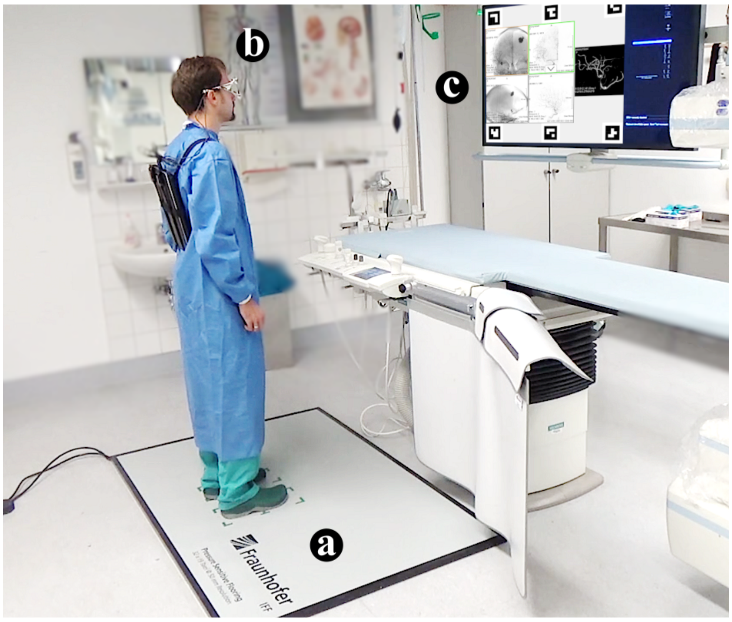


Figure 5.9: Technical setup for hands-free interaction consisting of a tactile floor (a), a mobile eye-tracking device (b) and the large screen of an angiography suite (c).

After that, the image data can be manipulated. For a series of images, this means scrolling back and forth [95] or search for a specific image that provides the best overview [78]. When 3D data is available, it can be rotated, requiring two DoF at least. Image manipulation thus can be split up into the following subtasks:

- Scroll through an image stack (one DoF)
- Rotate a 3D data set (two DoF)

This set of tasks lay the basis for the development of the following interaction techniques. This also includes the studies that were conducted for evaluation purposes.

5.2.2 Hands-free Image Manipulation Techniques

In this section, four techniques for hands-free interaction with image data in the given scenario are presented. Two of them build on the idea of combining eye gaze and foot input. For comparison, one approach employs foot input only and one relies on eye gaze only. Eye gaze reveals the user's current area of attention [186, 231], while foot input can be used to perform *single step* and *step and hold* commands [129] to allow hands-free input. The proposed multimodal approaches differ in how viewport selection is performed. *On-the-Fly Manipulation* follows a liberal approach by keeping eye gaze input enabled all the time. *Dedicated Lock Gesture* roughly follows the *gaze suggest* and *touch confirms* principle proposed by Stellmach and Dachselt [195] event though double-tap instead of a touch event is required for confirmation. When it comes to image manipulation, all approaches share the same foot input method. An overview of the methods can be found in Figure 5.10.

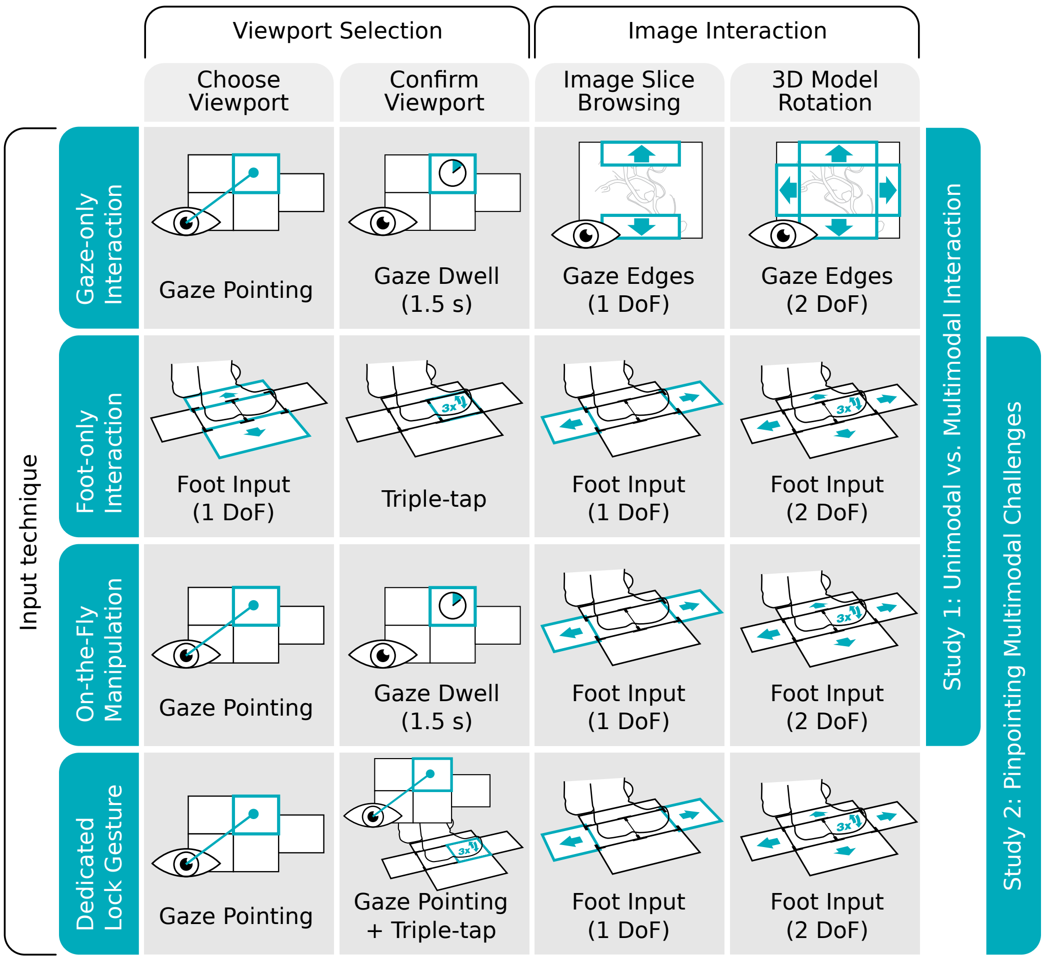


Figure 5.10: Overview of the input techniques. For viewport selection, both subtasks need to be performed, for image manipulation one of the variants is used depending on the type of image data.

On-the-Fly Manipulation

This approach uses continuous gaze input for the selection of the active viewport. The assumption is that involuntary deselection does not occur in the proposed use-case as a viewport is a sufficiently large target so that exact selection is not necessary. Further, the information to be assessed is displayed inside this area, which means leaving the area should be a strong hint that interaction with the corresponding content is not required anymore. Preliminary studies showed that the inaccuracy of eye-tracking influences this method when gazing at content near the borders. To account for that, a dwell time of 1500 ms is used to compensate for glances on adjoining viewports. This is a very high value for a dwell time approach where common values range between 400 ms to 1000 ms [112, 113, 218, 97, 43]. In this case, however, gaze is not used for the selection of small targets in short succession but for a task more comparable to mode switching. This is required less often and makes reliability more critical that a fast selection process. Further, information retrieval with the eyes during dwell time is possible as the dwell area covers a whole viewport. Therefore, dwell time duration is a less critical factor compared to use cases that require fast, fine-grained input, such as eye typing.

Dedicated Lock Gesture

Contradictory to the assumptions made for *On-the-Fly Manipulation*, this approach takes a more conservative estimate for the reliability of gaze input and requires a *Dedicated Lock Gesture* for confirmation of a selection. This is motivated by the findings of Mauderer, Daiber, and Krueger: For the selection of distant targets, their approach required a flick hand gesture combined with gaze selection. They reported that some subjects were looking at their hand on the touchscreen instead of the target[118]. A similar effect could be possible when using the feet for interaction while focusing on a screen.

To select a viewport, gaze is used. While looking at the viewport, a triple-tap with the ball of the foot has to be performed to confirm the selection. In contrast to *On-the-Fly Manipulation*, interaction is restricted to the selected viewport until another one is selected using the same triple-tap gesture. This means images can be manipulated no matter if the user looks at the screen or not. The intention is to prevent involuntary viewport switches by demanding an active action of the user while being less fatiguing by allowing the user to look somewhere else after the confirmation process. The triple-tap gesture was chosen to allow a reliable distinction from foot lifts during pose correction.

Foot only Input

For comparison with unimodal yet hands-free input, *Foot-only Interaction* does not rely on eye-tracking. Instead, a viewport can be selected by cycling through all available viewports in a serialized fashion by stepping to virtual pedals on the left and right. Confirmation is performed the same way as for *On-the-Fly Manipulation* and *Dedicated Lock Gesture* by using a triple-tap foot gesture.

Foot Input for Image Manipulation

Both the multimodal input techniques and foot-only input use the same approach for foot-based image manipulation. This is done by pressure-sensitive areas on the floor, which are referred to as ”virtual pedals”. The area underneath each foot is subdivided into two virtual pedals. This allows for the discrimination of ball and heel taps. Additional virtual pedals are arranged around the user's position and between the feet. These are used for mapping image manipulation onto the four directions, which corresponds to the directions that are used when navigating through image data with mouse or joystick. Navigating through image stacks is done using the virtual button in front of the foot and behind the heel. Rotation of 3D objects is performed using all four virtual pedals. It is realized using a Two-Axis Valuator approach as it was found to perform best for mouse input [14]. This means rotation happens relative to the user's perspective or the virtual camera. Stepping on the left or right virtual pedal rotates the object around the cameras up-vector while stepping on the pedal at the front or at the back rotates around the vector perpendicular to the up-vector and the view-vector of the virtual camera. Diagonal movements, which combine rotation around both axes and are part of the Two-Axis Valuator approach [14], are not considered as is difficult to step on two buttons at the same time with one foot.

Gaze only Input

Viewport selection is performed in the same way as with *On-the-Fly Manipulation*. Image manipulation with *Gaze-only Interaction* is performed by looking at predefined areas located near the edges of a viewport. The overall layout of these areas corresponds to the virtual pedals for foot input. Image stack navigation is done by looking near the upper or lower edge of the current viewport while 3D rotation is performed using all four edge areas.

Section 4.3 revealed the challenges when using plantar pressure for controlling continuous input, such as complicated calibration processes due to different body weight and preferences. For this approach, interaction is realized using virtual pedals as discrete buttons. A pedal changes the corresponding value (i.e., the slice index or angle) at a fixed rate as long as it is pressed. To reduce the time required to make significant changes to the value, the rate of change increases after three cycles. Initially, the value changes once every second. The increased rate is once every 200 ms.

5.2.3 Evaluation

The proposed interaction techniques were compared to answer the following questions:

- What are the challenges for unimodal input techniques?
- Which problems arise due to the interplay of multiple modalities?
- Which technique performs best?

With these questions in mind, two consecutive studies were conducted. At first, two unimodal approaches and a multimodal approach were investigated. *Gaze-only Interaction*, *Foot-only Interaction* and *On-the-Fly Manipulation* were compared concerning task completion times, perceived task difficulty, subjective workload and suitability for a given task. The Single Ease Question (SEQ) [177] was used to measure perceived task difficulty. A single question regarding the suitability of input setups for a task was asked. Subjective workload was assessed using the RTLX questionnaire [68, 67]. The *Dedicated Lock Gesture* technique was created based on the findings of the first study.

In the second study, the multimodal approaches *On-the-Fly Manipulation* and *Dedicated Lock Gesture* as well as *Foot-only Interaction* were compared in terms of task completion time and subjective workload.

Apparatus

To realize the proposed input techniques, the setup consisted of a head-mounted eye-tracking device, a tactile floor, a laptop running a prototypical user interface and a screen. The technical setup used in both studies was almost identical. Details and differences are described in the following. A head-mounted eye tracker from Pupil Labs [100] with an accuracy of 0.6° under ideal conditions was used. The device uses two cameras to determine the gaze point of the user: one for tracking the user's pupil and a second, forward-facing one to put the gaze point in relation to the world view. Calculations are performed by an additional laptop (Thinkpad E320, 4x Intel Core i5 @ 2.5 GHz, 8 GB RAM) worn as a backpack to which both cameras connect (see Figure 5.9). The Pupil Labs eye-tracking platform is an open-source project which consists of several apps and is easily extendable. Gaze point calculations were performed by the pupil labs capture app and relations to on-screen coordinates were determined using the Surface Tracker plugin. On-screen gaze coordinates were sent to the laptop that ran the prototypical software via ad-hoc Wifi by a custom UDP plugin. A tactile floor from the *Fraunhofer Institute for Factory Operation and Automation IFF* already described in Chapter 4 was used. It consists of a piezoresistive composite material sandwiched between two electrodes, effectively creating an array of tactile sensors [137]. It is covered with linoleum, similar to the typical flooring in operating rooms. The pressure data was accessible via USB. Gaze position and floor pressure data were received by a Thinkpad T540p laptop (Intel i5 Core @ 2.6 GHz, 8 GB RAM) running the prototypical software application and displaying it on the connected screen. In study one, a 56-inch monitor was used; in study two, the software UI was displayed on the large screen of an angiography suite.

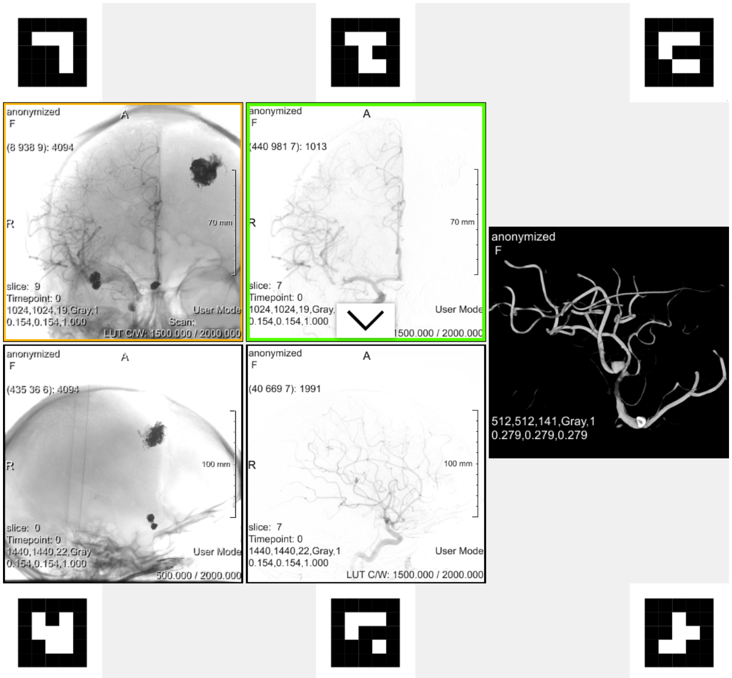


Figure 5.11: Prototypical software consisting of four viewports showing medical image stacks (left and center) and one 3D-viewport (right). Borders indicate the selected but not confirmed viewport (orange) and the currently selected viewport (green). Image manipulation is indicated by arrow icons at the viewport edges. Six fiducial markers around the viewports provide a reference for the mobile eye-tracking system. Image from Hatscher et al. [73].

The prototypical software implements the proposed interaction techniques and resembles the graphical user interface of an angiography suite. It was created with MeVisLab [164]. The interface corresponds to one of the available layouts on the angiography suite and consists of five viewports. The rightmost viewport displays a 3D volume data set, which is used to display a 3D representation of a blood vessel structure that results from a three-dimensional digital subtraction angiography (3D-DSA). The remaining viewports show series of angiography images that can be scrolled back and forth. When interacting with the system, visual cues communicate the current state regardless of the interaction technique. Viewports that are currently selected but not confirmed are surrounded by an orange border, and the selected viewport is surrounded by a green one. When interacting with image data, arrows show the direction. For 2D images, an arrow icon appears at the top when scrolling up and at the bottom when scrolling down the image stack. At the 3D viewport, arrow icons appear near the border corresponding to the rotation direction. The graphical user interface is shown in Figure 5.11. The software prototype was not connected to the angiography system except for the screen. Therefore, no live X-ray images could be acquired. During the study, anonymized X-ray datasets were used. Additional information on the system state was not included to reduce distractions during the study. Further, fiducial markers are arranged around the user interface. This is necessary for the Surface Tracking plugin to detect the screen and calculate the on-screen coordinates of the gaze point.

Study one: Unimodal vs. Multimodal Interaction

The first study aims to investigate challenges for hands-free image manipulation. Gaze and the feet are therefore compared as unimodal input channels. *On-the-Fly Manipulation* is introduced to find out which aspects of multimodal input are advantageous and which ones bear unsolved challenges.

Participants Ten participants (three female, seven male) between 22 and 31 (M = 24.8, SD = 2.7) with normal or corrected-to-normal vision took part in the study. All of them were students of medical systems engineering or computational visualistics. The participants were acquired by mailing lists and remunerated with 10 €. Little prior experience with foot interaction was stated by three participants (i.e., rating it with 1 on a 5-point Likert scale), little prior experience with eye-tracking as input setup by one. The remaining participants had no experience in either category.

Task	Type	Task Description
1	2D	Select the upper left viewport and go to slice 7
2		Select the lower center viewport and go to slice 7
3	3D	Select the right viewport and rotate to target position

Table 5.4: Overview of the task sequence for study one. The target position for Task three was provided as printout.

Measures The participants conveyed verbally when they started and finished a task. The investigator then logged the task completion time. After each task, a questionnaire assessing the SEQ and the suitability of this input setup for the given task was filled out. After all tasks for one input technique were completed, a RTLX questionnaire was filled out. The answers were conveyed verbally and noted by the investigator. A post-test interview was conducted to gather comments and issues that came up during the trials.

Procedure The study took place in a computer laboratory. Each participant completed a set of three tasks for each setup (within-subjects design). The sequence of the setups was counterbalanced over all participants. The participants wore clogs with a hard rubber outer sole while operating the tactile floor to avoid different recognition accuracy caused by different shoes. A position was marked on the floor to maintain the same distance to the screen for all participants. Regardless of the setup sequence, participants wore the eye tracker during the whole test. A demographic questionnaire was filled out beforehand. A 16-point eye tracker calibration was performed and repeated between tasks when participants experienced inaccurate results. For each setup, the investigator first explained the system, followed by a free training task until the participant felt confident dealing with the system. After that, the participant performed the three tasks listed in Table 5.4. Tasks one and two required a viewport selection and interaction with sequences of radiological images with one DoF, further referenced as 2D tasks. Task three required rotation of the vessel tree with two DoF (3D task). Target slices were given as numbers; the target orientation of the 3D vessel tree was provided as a printout.

Results For task completion time, task difficulty, and suitability of an input setup for a specific task, a two-way ANOVA for repeated measures was performed with the input technique as the first factor and task types (2D, 3D) as the second factor. The values for tasks one and two were averaged in

	df	F	p	η^2_{part}	Effect
Task completion time					
Setup	2, 18	7.58	<.01	.46	large
2D vs. 3D	1, 9	13.53	<.01	.60	large
Interaction	2, 18	6.54	<.01	.42	large
Task difficulty					
Setup	2, 18	8.87	<.01	.50	large
2D vs. 3D	1, 9	9.34	.01	.51	large
Interaction	2, 18	.98	.39	.10	small
Suitability					
Setup	2, 18	8.77	<.01	.49	large
2D vs. 3D	1, 9	11.25	<.01	.56	large
Interaction	2, 18	3.07	.09	.25	small
Subjective workload					
Setup	1, 9	5.58	<.01	.38	large

Table 5.5: Summary of the test statistics of study one for task completion times, perceived task difficulty (Singe Ease Question), suitability of input setups for a specific task type and subjective workload.

the factor task types because both tasks represent interaction with 2D-images. The test results are presented in Table 5.5.

The average training time the participants needed until they felt confident with the system was 2 minutes, 13 seconds. The task completion times are shown in Figure 5.12. To complete the 2D tasks, the participants needed a comparable amount of time for all three input setups. For the 3D task, the pattern of results differed considerably. Participants needed more time for 3D tasks (M = 137.3 s, SD = 99.6 s) compared to the 2D tasks (M = 20.8 s, SD = 10.4 s). They needed more time to perform the task with *Gaze-only Interaction* compared to *Foot-only Interaction* and *On-the-Fly Manipulation*. The results reveal significant main effects for the input setup and the task type and a significant interaction effect.

The analysis of task difficulty revealed a significant main effect for input setup and task type. The mean results for perceived task difficulty are presented in Figure 5.13. Participants perceived interaction with radiological images as more difficult using *Gaze-only Interaction* (M = 4.2, SD = 1.0) compared to *Foot-only Interaction* (M = 2.8, SD = 0.7) and *On-the-Fly Manipulation* (M = 2.8, SD = 1.5). Moreover, participants perceived interaction in the 3D

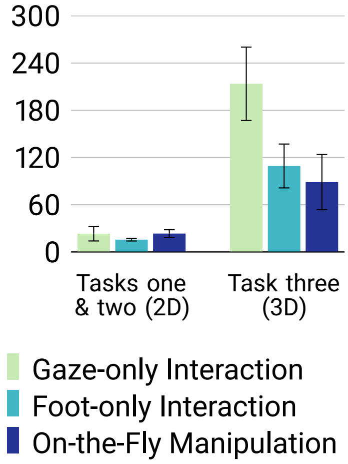


Figure 5.12: Mean task completion time in seconds for study one with standard error bars.

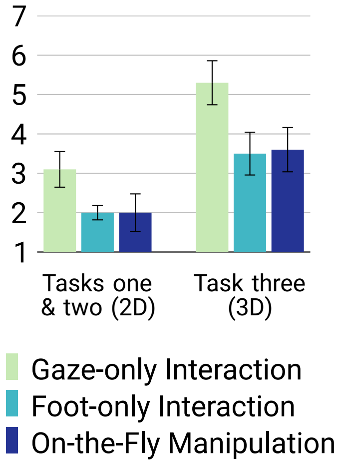


Figure 5.13: Mean results for task difficulty (SEQ) for study one with standard error bars. (1 = very easy, 7 = very hard)

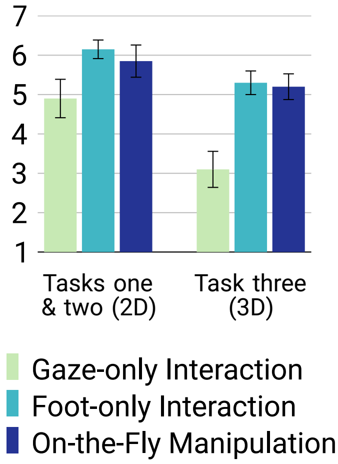


Figure 5.14: Mean results for suitability of input method for study one with standard error bars. (1 = not suitable, 7 = very suitable)

task with two DoF as more difficult (3D task: M = 4.1, SD = 1.6) compared to navigation of image sequences (2D task: M = 2.4, SD = 0.7). The interaction effect was not significant.

Similar results revealed the analysis of the perceived suitability of input setups for the given task types, which are presented in Figure 5.14. Participants stated that *Gaze-only Interaction* was significantly less suitable for 2D tasks (M = 4.0, SD = 1.2) compared to *Foot-only Interaction* (M = 5.7, SD = 0.6) and *On-the-Fly Manipulation* (M = 5.5, SD = 1.1), which was reflected in a significant main effect for input setup. Respectively, tested input setups were perceived as significantly more suitable for 2D tasks (2D: M = 5.6, SD = 0.6) than 3D tasks (3D: M = 4.5, SD = 0.9). The interaction effect was not significant for this variable.

Subjective workload assessed with a RTLX questionnaire was analyzed with a one-way ANOVA for repeated measures with input technique as the only one factor. The effects are presented in Figure 5.15. A significant main effect was found for the overall workload. The participants reported the highest workload for *Gaze-only Interaction* (M = 8.4, SD = 2.4) compared to *Foot-only Interaction* (M = 5.9, SD = 2.2) and *On-the-Fly Manipulation* (M = 5.1, SD = 2.1). The detailed analysis shows that this pattern was present in the dimensions of temporal demand and effort. For mental demand,

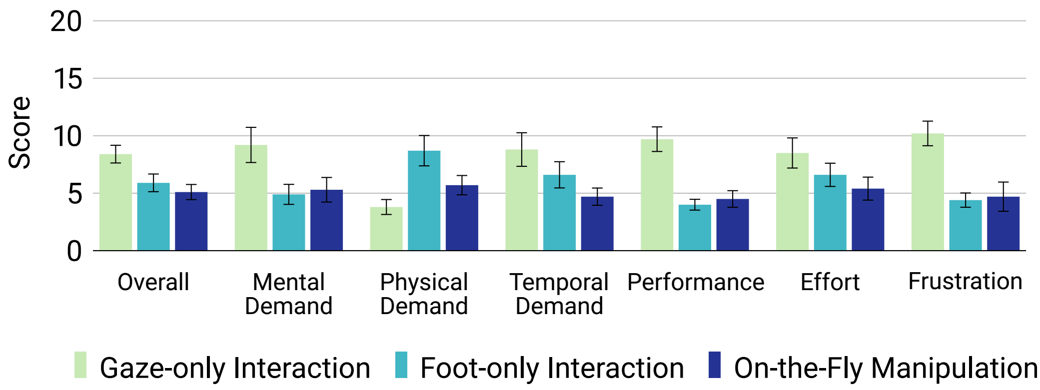


Figure 5.15: Unweighted mean scores for the NASA-TLX questionnaire from the first study on a scale from 0 to 20 (0 = low/good performance, 20 = high/poor performance). Error bars represent standard deviation. The overall workload is the mean value from all the subscales.

performance and frustration, the values of *Foot-only Interaction* and *On-the-Fly Manipulation* were comparable. In contrast, the participants perceived *Foot-only Interaction* to be the most and *Gaze-only Interaction* to be less physically demanding.

Summary Study one Study one shows that *Gaze-only Interaction* the slowest, most challenging and less suitable input method for the investigated tasks. However, *On-the-Fly Manipulation* performed comparably to *Foot-only Interaction*. The multimodal approach subjectively required less physical demand but no significant higher mental demand (see Figure 5.15).

It is unclear how always-on gaze interaction influences the results for *On-the-Fly Manipulation*. For this reason, a second study was conducted to find out how task completion time is distributed between subtasks (e.g., selection, confirmation, manipulation) for different multimodal interaction techniques. Further, observations suggest that the performance in the rotation task is influenced by the participant's spatial sense.

Study two: Pinpointing Multimodal Challenges

This study aimed to determine the influence of single modalities to task completion times. For this reason, an interaction technique using a *Dedicated Lock Gesture* was introduced. By not automatically confirming a viewport upon looking at it, the potentially disadvantageous influence of gaze interaction during image manipulation might be reduced. *Foot-only Interaction* was also included as subtasks were not analyzed in the first study (see Figure 5.10).

Participants Thirteen students (six female, seven male) between 22 and 31 years old (M = 25.5, SD = 3.1) with normal or corrected-to-normal vision participated in the study. Seven majored in human medicine, four in medical systems engineering, one in computational visualistics and one in computer science. The participants were acquired by mailing lists and remunerated with 20 €. Four participants reported medium prior experience with foot interaction and eye-tracking as an input method (i.e., rating it 3 on a 5-point Likert scale). The remaining participants did state no prior experience.

Tasks Similar to study one, a within-subjects design was followed. A set of six tasks had to be completed for each interaction technique (see Table 5.6). The tasks were grouped in two blocks of three tasks. In the first block, only interaction with 2D image stacks was required. The second block included 3D interaction tasks. In study one, performance at the 3D-task seemed to rely on the participant's spatial sense. To reduce the influence of this factor, 3D interaction tasks were restricted to rotation around one axis. In each task, the viewport had to be changed, and the content in the target viewport had to be manipulated.

Table 5.6: Overview of the task sequence used in study two.

Task	Type	Task Description
Training	2D&3D	-
Verification	2D	Select lower left viewport and go to slice 4
	2D	Select the upper center viewport and go to slice 15
	3D	Select the right viewport and rotate three steps right
1.1	2D	Select the the lower center viewport and go to slice 16
1.2	2D	Select the upper left viewport and go to slice 7
1.3	2D	Select the lower center viewport and go to slice 3
2.1	3D	Select the right viewport and rotate eight steps left
2.2	2D	Select the upper center viewport and go to slice 7
2.3	3D	Select the right viewport and rotate six steps down

Table 5.7: Summary of the test statistics for task completion times (TCT) and subjective workload (SW).

	df	**F**	**p**	η^2_{part}	**Effect**
TCT					
Interaction Technique	2, 24	2.27	.13	.16	large
Subtask	1.16, 13.95	207.90	<.01	.95	large
Interaction	1.89, 22.69	47.92	<.01	.80	large
SW					
Interaction Technique	2, 24	1.10	.35	.08	medium

Measures As a measure of performance, TCT was assessed. The time was logged manually by the investigator to account for the uncertainties of the user, including correction phases. To avoid early starts, input was deactivated between measurements. Time measurement was stopped as soon as the participant signaled task completion. Communication delays were corrected based on video logs, taking the last foot movement of the participant before conveying task completion as stop cue. Completion time for the subtasks *choose viewport*, *confirm viewport* and *image manipulation* was gathered by analyzing the video logs. Video analysis focused on the amount of time required to complete one subtask, not on cumulated times functionalities were used. This means that upon finishing a subtask (i.e., ”confirm selection”), time was measured until the next subtask (i.e. ”image manipulation”) started even if involuntary interaction and correction loops took place in between. Similar to study one, the workload was assessed using the RTLX questionnaire.

Procedure The study took place in an angiography suite. First, a questionnaire assessing demographics, experience with gaze or feet as input modality and shoe size was filled out by the participant. After that, the participant donned a pair of clogs, available in 41/42, 43/44 and 45/46 Paris points and was equipped with the eye-tracking device. As in study one, an area to stand in was marked on the floor. The order of input techniques was counterbalanced to reduce learning effects. As forward movements with the feet were found to be more effective than backward movements [176], the tasks were not randomized.

Each input technique was explained by the investigator. In case gaze interaction was used, a 16-point calibration was performed. Following, a free training phase took place, which lasted until the participant felt confident with the system. Three tasks had to be finished successfully in under 1:30 min to verify a certain level of confidence. The training phase would have been repeated in case one of the tasks was not completed in time, but every participant passed on the first try. Instructions were read out loud by the investigator before starting a trial to separate time for comprehension of a task from the actual task completion time. A RTLX questionnaire was filled out after each task block.

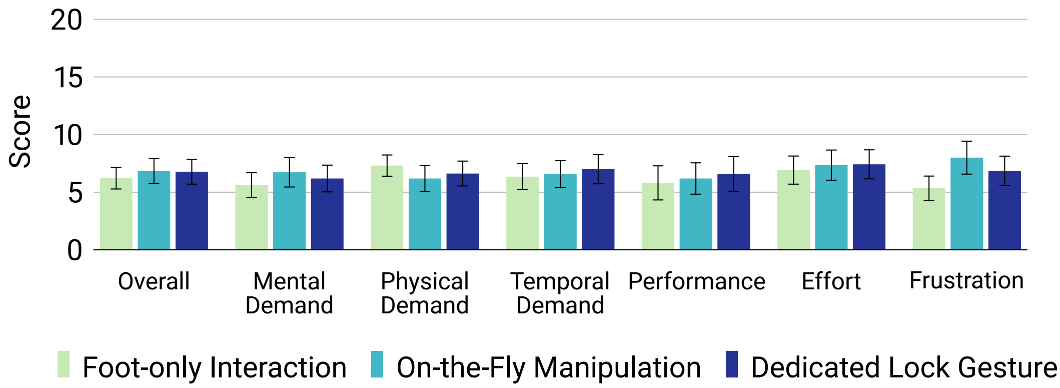


Figure 5.16: Unweighted mean scores for the NASA-TLX questionnaire from the second study on a scale from 0 to 20 (0 = low/good performance, 20= high/poor performance). Error bars represent standard deviation. The overall workload is the mean value from all the subscales.

Results

Overall, the interaction techniques compared in study two yielded similar results. However, the time required to complete each of the subtasks differed considerably. An overview of the subtask completion times gathered from analyzing the video logs can be found in Figure 5.17. The task completion times were analyzed by a 3×3 ANOVA with the factors *input technique* and *subtask* (choose viewport, confirm viewport, manipulate data). In case the assumption of sphericity was violated, a Greenhouse-Geisser correction of degrees of freedom was applied. The analysis revealed a significant main effect for the *subtask* and a significant interaction effect between *input technique* and *subtask* (see Table 5.7). For choosing a viewport, considerably shorter times were achieved with gaze input than by *Foot-only Interaction* (see Figure 5.17). Confirmation of a viewport selection was accomplished the fastest with the dwell time approach from *On-the-Fly Manipulation* and took the longest

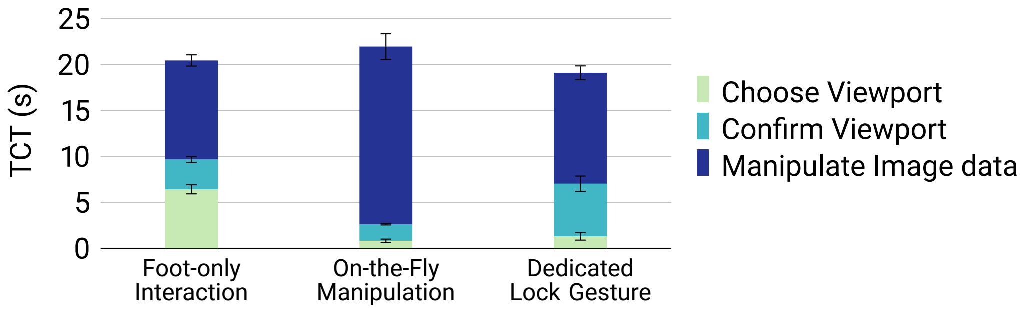


Figure 5.17: Mean task completion times from the second study divided into subtasks for *Foot-only Interaction*, *On-the-Fly Manipulation* and *Dedicated Lock Gesture* with standard error bars for each subtask.

when the *Dedicated Lock Gesture* had to be performed. Completing the image manipulation task took the longest with *On-the-Fly Manipulation*.

RTLX scores from all tasks were pooled due to insignificant differences and analyzed with a one-way ANOVA for repeated measures with the interaction technique as the only factor. There was no significant result, as can be seen in Figure 5.16.

5.2.4 Discussion

In this section, concepts for hands-free interaction with image data for a radiological use case were evaluated. The feet and eye gaze were used as unimodal input and combined in two ways into multimodal approaches. The resulting input techniques were compared in two user studies. The first study showed that *Gaze-only Interaction* performed worst in the given scenario while unimodal foot input and the first multimodal technique performed comparably. The second study showed overall comparable results for *Foot-only Interaction* and both multimodal approaches. However, an in-depth analysis of subtask completion times revealed that gaze pointing works fastest for selection task, but the combined use of gaze and the feet seems to lengthen the subtask completion time.

Velloso et al. separated and analyzed trial completion times for gaze and hand gesture interaction in a similar way. In their work, gaze was compared to 2D and 3D hand gestures for object selection, while object manipulation was done by hand gestures every time. The time required to complete the tasks was separated into three task steps (acquisition, confirmation, translation) [209]. Their results align with the findings presented in this chapter when it comes to fast selection via gaze but differ for subsequent steps.

In [209], object confirmation and translation yielded similar task completion times while confirmation and image data manipulation results presented in this chapter differ between input techniques. There might be two reasons for this: First, all interaction techniques presented by Velloso et al. implemented a finger pinch as *Dedicated Lock Gesture*, which avoids the Midas Touch Problem by disabling gaze input during the transition task, compared to *On-the-Fly Manipulation*. Second, the feet were not in the participant's FoV. Even though there was no visual feedback on the floor and the position on the virtual foot pedals except for the central position was not marked, it was observed that participants looked at their feet when shifting to maintain a stable stance, after stepping on virtual foot pedals or putting the feet back on the marked position. This indicates that proprioceptive feedback is not sufficient and visual checks have to be performed. Similar difficulties were reported in studies that required interaction outside the user's FoV or in peripheral vision [202, 118]. When gestures or interaction was done close to the screen (and therefore, in the user's FoV), no such problems were reported [30, 209, 154]. This problem might be tackled in a variety of ways. To avoid the need to look at the feet, the visual feedback presented to the user interface, as well as dynamic placement of the virtual buttons presented in Section 4.1, could be included. Saunders and Vogel compensated for minor foot movements by automatically adjusting the center of the virtual foot input areas [175], which could also be used to avoid the need to look at the feet. Foot-mounted input devices should not suffer from these problems. Further, the eye gaze position can be used to disable all input channels in case the user looks away from the screen.

Limitations Even though the proposed approaches offer valuable insights into unimodal and multimodal hands-free interaction in a standing position, there is room for improvement. With the delicate structures of medical image data in mind, no gaze cursor was displayed to avoid distraction. Switching the viewport involuntarily might be avoided when the detected gaze position is displayed. Further, offsets due to poor calibration can be detected easier. A solution might lie in calibration-less gaze input techniques such as *smooth path pursuit* [210], dynamical recalibration [155] or a method to temporarily display the gaze cursor.

The focus of this section lies in the general suitability of input methods. Even though the tasks and approaches are motivated by clinical needs, the study design did not take all aspects of an interventional scenario into account. A manual task representing the handling of medical instruments was not present. The influence of such a task when performed during multimodal interaction is investigated in depth in Chapter 6. Image control is easy to map on virtual foot pedals in a spatially meaningful way. Increasing functionality might require interaction with more abstract parameters and could cause a higher workload for the user. In the OR, lots of external factors come into play. Alarms and status notifications of medical systems are communicated using auditive and visual signals, interpersonal communication with the medical team and consultations between colleagues take place. Such interruptions are not present in a lab setting. Nevertheless, the findings from these sections give insights that might contribute to the development of direct, hands-free interaction methods for the medical domain.

5.3 Conclusion

In Section 5.1, foot, hand and voice input were used to perform a prototypical image manipulation workflow. As foot movements, double-tap for discrete commands and heel rotation for continuous input were used. Hand gestures followed a lever metaphor, including a grab gesture for discrete input and movement of the closed fist for continuous parameter adjustment. Voice commands allowed system activation and function selection via key phrases. In a user study, foot-based input excelled in terms of task completion time, usability, usefulness and overall evaluation. Further, qualitative feedback suggests that foot input is considered familiar because foot pedals are already established in clinical practice, it is deemed more robust against noise than voice input and more compatible with sterility guidelines than hand gestures. Even more elaborate use of foot and knee movements seems acceptable, based on personal preferences.

Section 5.2 extended foot input by adding gaze for pointing tasks. The proposed combined methods yielded comparable results for foot-only and gaze-supported approaches. In-depth analysis revealed that more time was required for simultaneous gaze and foot subtasks. Visual checks of the position of the feet might indicate that proprioceptive feedback is insufficient and has to be supported with visual feedback in the user's FoV. Using gaze, therefore, suffers from the double-role of the eye for confirmation and control, similar to the Midas Touch Problem.

In summary, the feet seem to be a suitable choice for hand-free input for simple tasks as they outperformed hand and voice input and are deemed more suitable for clinical scenarios. Extending foot input with gaze to allow more complex tasks suffers from the position of the feet outside the user's FoV when standing.

This chapter is based on the following publications:

B. Hatscher and C. Hansen. “Hand, Foot or Voice: Alternative Input Modalities for Touchless Interaction in the Medical Domain”. In: *Proceedings of the 20th ACM International Conference on Multimodal Interaction*. ICMI ’18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 145–153.

B. Hatscher, M. Luz, L. E. Nacke, N. Elkmann, V. Müller, and C. Hansen. “GazeTap: Towards Hands-free Interaction in the Operating Room”. In: *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. ICMI ’17. New York, NY, USA: Association for Computing Machinery, 2017, pp. 243–251.

6

Effects of Multimodal Interaction on Primary Tasks

In previous chapters, unimodal and multimodal input methods were investigated regarding their suitability and performance with an interventional setting in mind. However, interaction with medical data is an auxiliary task that must not interfere with the clinical task at hand. Further, it is crucial to understand the influence of a manual task on user performance in an HCI context to support informed decision making when designing interventional user interfaces. In this chapter, the mutual influence of a primary, manual task and a secondary interaction task is investigated. Further, hands-free interaction methods are researched regarding the workload they impose on the user.

Performing multiple tasks at the same time or in short succession is called multitasking. It can be roughly characterized by the time spent on a task before switching to another task. This multitasking continuum ranges from concurrent multitasking to sequential multitasking [173]. Multitasking happens regularly when using computers at work [57, 116], in meetings [21] or lectures [4]. During multitasking, it takes additional time to resume the primary task when interrupted [134, 200]. Interruptions occurring during phases of higher workload are more disruptive than during periods of low workload [3]. It was found that users tend to defer interruptions to phases of lower workload [172]. The degree of multitasking influences productivity and accuracy in a different way: while productivity is highest for medium multitaskers compared to high and low multitaskers, accuracy decreases with an increasing level of multitasking [5].

Compared to computer workplaces where interruptions have no severe consequences, a secondary task should influence the main task as little as possible in domains with potentially critical situations like piloting a plane, driving or performing a medical intervention.

A potential solution lies in the multiple resource theory, which assumes that tasks only conflict if they tap into the same cognitive resource [223]. The multiple resource model differentiates between four dimensions (stages, perceptual modalities, visual channels, processing codes) with two levels each [223]. Using different resources for both tasks, therefore, should result in better performance during concurrent interaction than employing the same one.

In the context of aviation, scanning the instrument panel and monitoring the outside world demand the pilot's visual attention. The comparison of auditory and visual delivery of air traffic information generally suggests a note of caution for intramodality (visual-visual) representation over mixed-modality (audio-visual) [224].

In the context of driving, using a smartphone or a GPS navigation system requires divided attention. Voice entry methods performed significantly better and yielded a lower standard deviation of the lateral lane position than touch-based smartphone input for destination entry in a driving simulator [20]. Further, faster reaction times and lower miss rates in a Detection-Response Task, which assesses the attentional effects of cognitive load for secondary tasks in a driving environment [87], were found for voice entry [20]. In line with these findings, a smaller negative influence on driving and higher texting performance while driving was found for conditions that used voice input, compared to the touch keyboard of a smartphone in a texting-while-driving scenario [199].

6.1 Multitasking for Multimodal Augmented Reality Input

Recent advances in virtual reality (VR) and augmented reality (AR) hardware such as the HTC Vive (HTC Corp., Taoyuan City, Taiwan) or the Microsoft HoloLens (Microsoft Corp., Redmond, WA, USA) led to a range of applications in the medical context [205, 203].

In the OR, the amount of information presented using AR can exceed mental processing abilities, which might lead to unusable or distracting assistance systems [101]. Investigated in the context of an endoscopic navigation exercise, AR led to increased inattentional blindness (i.e., the failure to notice unexpected objects or events when focused on a task) [38]. To avoid distraction, methods that allow switching between different information sets are required [205]. Interaction with virtual content in an interventional setting, on the other hand, easily interferes with the medical, manual task. In this context, the influence of direct, hands-free input methods on primary task performance must be kept to a minimum. Therefore, the questions (1) how well perform different input modalities for different input tasks, (2) which influence do they have on a manual task and in turn, (3) which influence has a manual task on interaction performance are pursued in an AR setting.

Two abstract interaction tasks for manipulating virtual content are explained and mapped to a range of different touchless input modalities. An overview of the proposed interaction techniques can be found in Figure 6.1 and 6.3. Further, a manual task simulating the workload of simultaneous interaction during an intervention is described. A user study is conducted to understand the interaction between interaction and manual tasks.

6.1.1 Interaction Tasks

To allow manipulation of virtual objects in the user's FoV, methods for pointing, selection and manipulation of parameters are provided. When using head-mounted displays, a straightforward approach for pointing is using the head direction. To confirm the selection of an object while pointing, a mid-air hand gesture, a foot gesture and a voice command are used. The hand gesture is called air tap and is part of the HoloLens default gesture set: It is performed by holding the hand in the FoV and connecting thumb and index finger shortly. Confirmation by foot requires a toe tap: The foot has to be lifted more than one degree and lowered within a time window of 0.8 s. Confirmation by voice command is done by uttering one of four key phrases: ”choose item”, ”okay”, ”pick out” or ”select”. The described input methods are implemented and used to perform a task modeled after the EN ISO 9241-420:2011 multi-directional tapping test [86, Annex F] (see Figure 6.2).

Manipulation of a parameter can be achieved by hand gesture, foot gesture or voice command. Foot and voice additionally require the user to turn the head in the desired direction. Scrolling via hand gesture is done by dragging: performing the air-tap gesture but keeping the thumb and index finger connected and moving the hand left or right. The foot gesture for continuous input first requires the user to rotate the head to the left or the right. After that, the ball of the foot can be lifted and rotated on the heel in the same direction as the head points to trigger scrolling. Turning the foot or the head back to a central position stops the interaction. Voice is used in combination with head movement. To activate, one of the key phrases ”begin”, ”move”, ”slide” or ”start” must be uttered. After that, manipulation is done by rotating the head to the left or right. When the head is rotated to the left or to the right only for a short amount of time (<1.3 s), the system scrolls to the next image in the corresponding direction. When the head position is held in a rotated position, the images scroll continuously with a rate of two images per second. One of the voice commands ”break”, ”exit” or ”stop action” deactivates the manipulation mode.

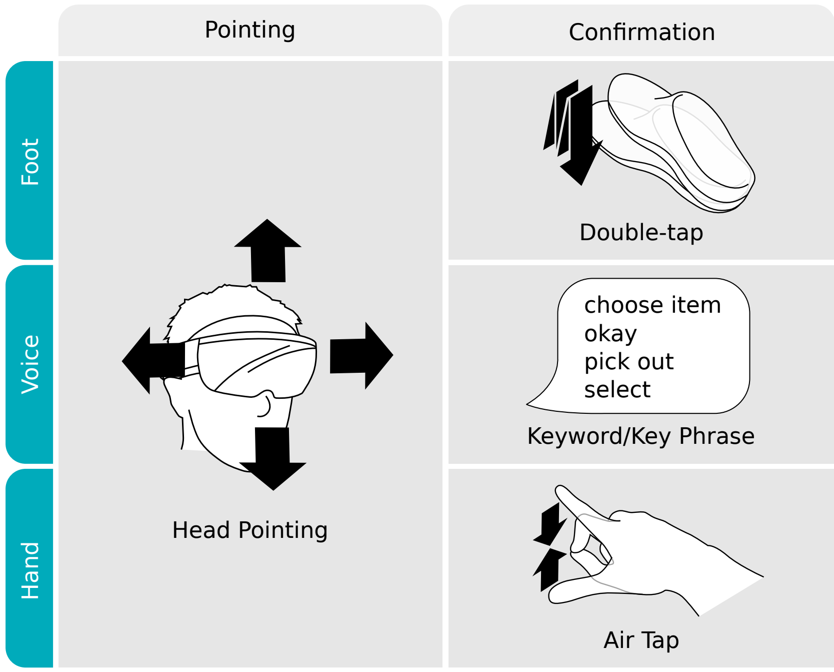


Figure 6.1: Overview of the input techniques for the confirmation task.

6.1.2 Simulating a Manual Task

A medical task such as guiding a needle or handling an ultrasound transducer requires the physician to correct the position of the instruments, for example, to compensate for respiratory motion. This kind of task is simulated in a way that does not require any prior medical knowledge yet forces the user to divide attention. A red, horizontal bar contains a green, moving area that indicates the target range and a smaller, white vertical bar. The vertical bar is controlled by the user and is to be held in the green range using a hand-held input device. The visual representation can be seen in Figure 6.4.

6.1.3 Evaluation

A user study was conducted to investigate the influence of the interaction tasks on the manual task and vice versa, as well as the performance of the proposed input methods. The study consists of two blocks. The first one targets pointing and confirmation. As pointing was always done by head movement and only differed in the way the selection was confirmed, this block is referred to as *confirmation* block. The second block investigates the *continuous input* task. Due to technical issues, two participants had to be

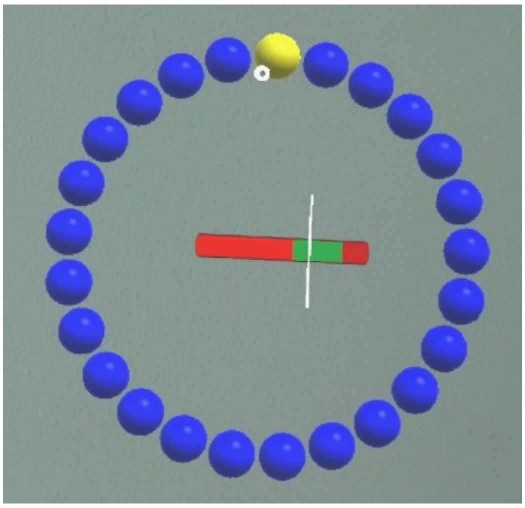


Figure 6.2: *Confirmation* task together with the manual task. The head-controlled cursor has to be pointed at the yellow sphere. Then, it has to be selected via hand, foot or voice command. Image from Solovjova, Hatscher, and Hansen [191].

excluded from the *continuous input* pool. Two additional participants were recruited for this block. This leads to 12 participants for each block but slight differences in demographics. As no comparison between both blocks is performed, this does not influence the validity of the results.

Participants

Twelve students (2 female, 10 male) between 20 and 29 years (M = 25 years, SD = 3.28) took part in the *confirmation* block. The professional backgrounds were majors in computer science (4), water management (3), mechanical engineering (2), computational visualistics (2) and environmental and energy process technology (1). 50% stated previous experience with augmented reality or virtual reality. All of them were right-footed.

In the *continuous input* block, two participants were replaced. This resulted in the same number of participants with the same distribution of sexes in the same range of ages. The mean age was 24.83 years (SD = 3.21) in this block. The professional backgrounds were majors in computer science (4), water management (3), mechanical engineering (2), computational visualistics (1), biomedical engineering (1) and environmental and energy process technology (1). 42% stated previous experience with augmented reality or virtual reality. As in the *confirmation* block, all of them were right-footed.

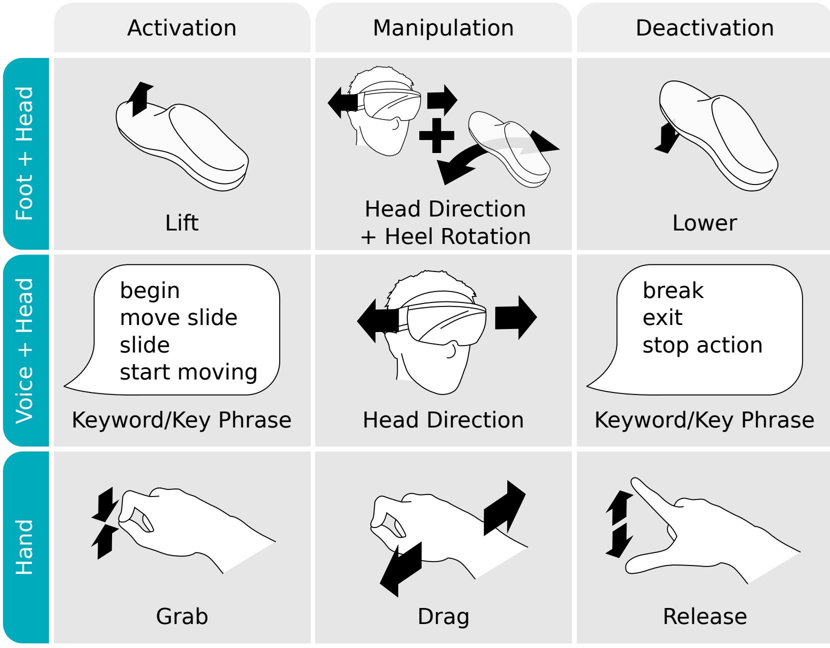


Figure 6.3: Overview of the input techniques for the manipulation task.

Tasks

The *confirmation* task is modeled after EN ISO 9241-420:2011 [86], where selectable spheres are arranged in a big circle (see Figure 6.2). The sphere to be selected is highlighted. When the head-controlled cursor points at the target sphere, the color changes and the selection can be confirmed using the air-tap hand gesture, tapping with the foot or the voice commands ”choose item”, ”okay”, ”pick out” or ”select”.

For evaluation purposes, *continuous input* is used to navigate an image gallery. The images are arranged horizontally with animated transitions. The gallery works in a circular fashion, meaning that it can be scrolled endlessly to the left or right as it loops at the end. The task to perform is given as the number of times in a particular direction. This number is displayed under the corresponding arrow left or right of the current image. The gallery including the task can be found in Figure 6.5.

Measures

Subjective workload was assessed using the RTLX variant of the NASA-TLX questionnaire [67]. Physical strain was assessed by marking body parts on a whole-body diagram, which is based on a questionnaire to assess muscu-

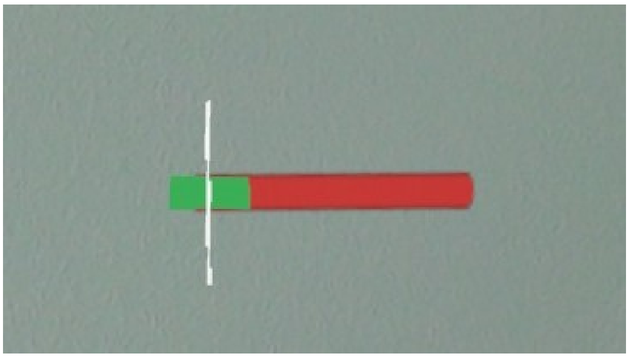


Figure 6.4: The primary task consists of a red slider with a moving green area and a white marker which is operated by the user. Image from Solovjova, Hatscher, and Hansen [191].

loskeletal disorders in stage rallying [115]. Subjective rating of the interaction approaches was gathered by asking the participants to rank the presented input methods. As an indicator of performance, task completion time was recorded. The number of overshoots, which is leaving the correct target without confirming the selection, was recorded to assess accuracy.

Apparatus

The Microsoft HoloLens Development Edition was used for augmented reality visualization, hand gesture recognition and speech recognition. Foot movement data was gathered with an MTw Awinda Wireless Motion Tracker. Input for the manual task was done using the right joystick of an Xbox wireless controller (Microsoft Corp., Redmond, WA, USA).

Procedure

The user study was carried out in a computer laboratory. It consisted of two independent blocks corresponding to the *confirmation* task and a *continuous input* task. At first, a general introduction to the topic was given, a demographic questionnaire was filled out and the system was calibrated for the participant. Next, the manual task was performed for 60 seconds without the influence of any simultaneous interaction to gather baseline data. The *confirmation* block was introduced at first as well as the first of the input modalities *hand*, *foot* and *voice*. After that, a training phase took place. The training lasted until the participant felt comfortable with the interaction technique, which never took longer than 4 minutes. The order of input modality and condition (with and without the simultaneously performed manual task) were

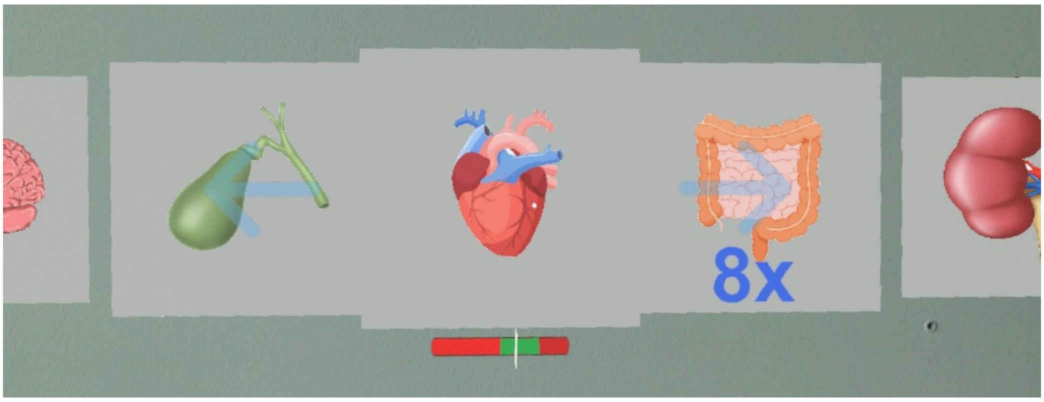


Figure 6.5: *continuous input* task together with the manual task. The current task is shown as the number of images the participant has to scroll in the corresponding direction. Image from Solovjova, Hatscher, and Hansen [191].

counterbalanced over all participants. The RTLX questionnaire was filled out after each task, and subjective physical strain was assessed after each modality. Ranking of the interaction approaches took place after each block. The same procedure was followed for the *continuous input* block.

6.1.4 Results

For the *confirmation* task, when performed without an additional manual task, hand gesture input had the lowest task completion time (M = 83.6 s, SD = 31.4 s) compared to foot input (M = 92.8 s, SD = 21.4 s) and voice commands (M = 95.6 s, SD = 24.3 s). In terms of accuracy, hand gesture interaction had the fewest overshoots (M = 2.6, SD = 2.1), followed by voice commands (M = 2.9, SD = 1.3) and foot input (M = 4.0, SD = 1.5). Overall, all *confirmation* methods performed worse in terms of task completion time an accuracy when a manual task had to be performed simultaneously (see Table 6.1 and Figure 6.7). The rise in overshots for foot input is less than for hand and voice input. A higher deviation can be observed with a manual task, especially for voice commands. During the *confirmation* task, the least influence on the primary task performance was achieved when interacting via voice commands, followed by foot and hand input (see Figure 6.6).
The physical strain questionnaire showed that 6 out of 12 participants had a strained shoulder when performing touchless hand gestures. This aligns with the observation that some participants paused interaction shortly to shake their arm. A strained shin and calf after performing the toe tap were reported by 4 out of 12 participants. The subjective rankings of the input methods

Table 6.1: Task completion times for the *confirmation* block

	without primary task		primary task present	
	Mean	SD	Mean	SD
foot input (tap)	92.8 s	21.4 s	117.3 s	26.7 s
hand gesture (air-tap)	83.6 s	31.4 s	123.1 s	41.1 s
voice command	95.6 s	24.3 s	132.4 s	38.8 s

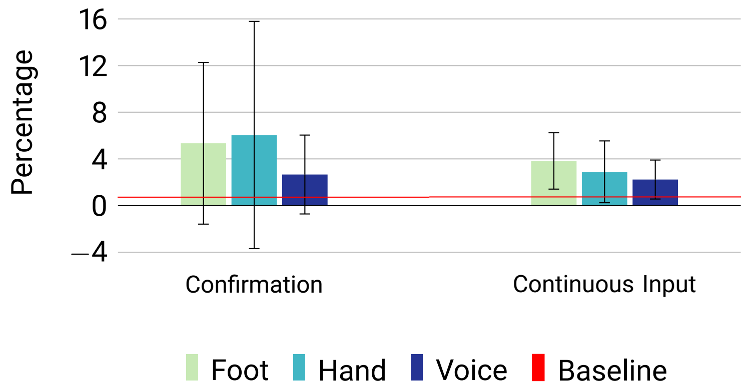


Figure 6.6: Influence of interaction tasks on primary task performance. The plot shows the mean percentage of task completion time spent outside the green target area (see Figure 6.4). The baseline (primary task without secondary interaction tasks performed for 60 s) is depicted in red.

revealed different results for the runs with and without the additional manual task. For selection as the only task, 42% preferred *confirmation* via voice command the most, 33% via foot input and 25% via hand gesture. When a manual task had to be performed simultaneously, the preference shifted to 67% interaction via voice command and 33% via foot. Hand gesture input was not considered the first choice by any of the participants with the manual task.

Results of the *continuous input* task without additional manual task reveal the shortest task completion times for hand gesture input (M = 68.8 s, SD = 10.9 s) followed by foot input (M = 90.2 s, SD = 20.2 s) and voice commands (M = 106.5 s, SD = 32.6 s) (see Figure 6.2). Accuracy for *continuous input* is shown separately for both input directions in Figure 6.9. For this task, voice commands performed worst in terms of accuracy with and without a manual task. For foot and hand input, however, the difference is quite low. Lower accuracy for input with the right hand than with the left

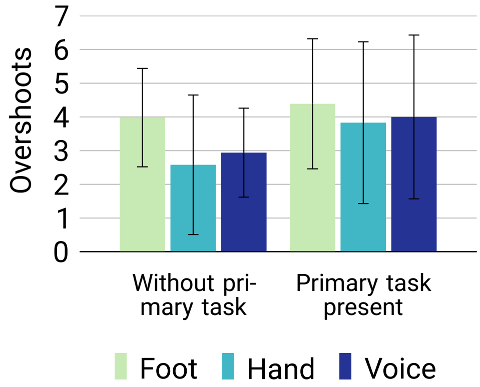


Figure 6.7: Number of overshoots (i.e. leaving the target sphere with the cursor after entering but before confirmation) during the *confirmation* task.

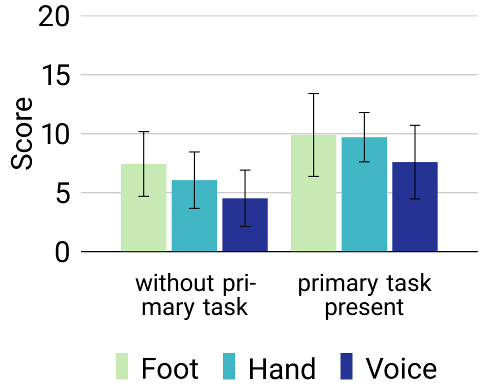


Figure 6.8: RTLX score for the *confirmation* task on a scale from 0 to 20. (0 = low perceived workload, 20 = high perceived workload)

Table 6.2: Task completion times for the *continuous input* block

	without primary task		primary task present	
	Mean	SD	Mean	SD
Foot (Heel Rotation) + Head	90.2 s	20.2 s	98.6 s	17.9 s
Hand Gesture (Drag)	68.8 s	10.9 s	75.9 s	16.0 s
Voice + Head	106.5 s	31.6 s	104.8 s	8.4 s

hand can be observed when the primary task has to be performed. When a manual task is performed during *continuous input*, it is influenced the least by a secondary task when using voice commands, but the difference between all input modalities is smaller than for *confirmation* (see Figure 6.6). Task completion time for hand and voice input increased when the manual task was performed simultaneously. Voice input yielded comparable to interaction without the manual task (see Table 6.2). The number of overshots increased except for scrolling in the right direction with voice (see Figure 6.9). In the following, only body parts that were marked as strained by three or more participants are reported. When using hand gestures, a strained shoulder was reported by six participants, a strained upper arm by four and a strained hand by three participants fot both tasks. *Confirmation* by foot yielded a strained shin and calf four times and a strained foot three times. *Continuous input*

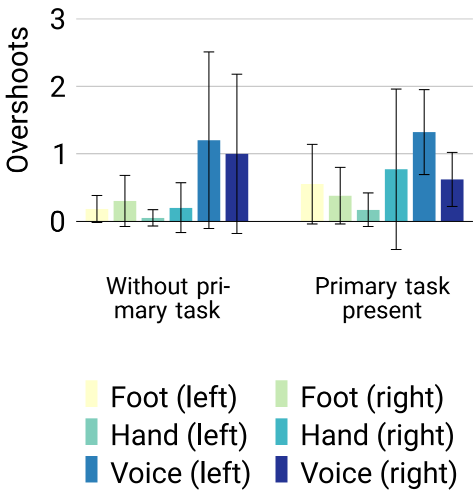


Figure 6.9: Number of overshoots (i.e. a navigation step in the wrong direction) during the *continuous input* task, separated by scrolling direction.

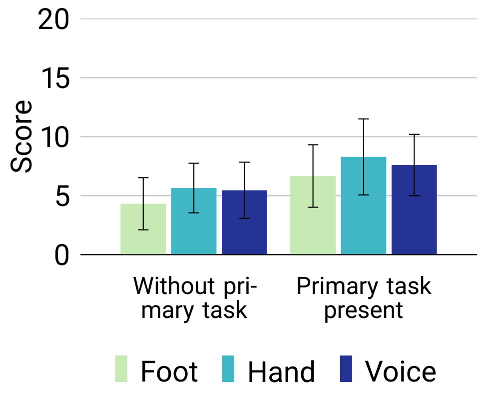


Figure 6.10: RTLX score for the *continuous input* task on a scale from 0 to 20. (0 = low perceived workload, 20 = high perceived workload)

caused a strained neck three times. Two participants reported a strained shin and calf, and a strained neck was reported three times. The RTLX scores reveal a high subjective workload for hand gesture input (see Figure 6.10). Voice commands caused the lowest subjective workload. Despite this, hand input was preferred by 67% of the participants for *continuous input* when it is the only task to be performed, 33% preferred foot input. For interaction when performing a manual task, 50% preferred foot input and 25% preferred hand input and voice commands each.

6.1.5 Discussion

In this section, three input techniques for multimodal interaction with virtual augmented reality content were compared to understand the interplay between input modalities and primary task accuracy. Input via voice commands had the least influence on primary task performance and caused the least subjective workload but was not the fastest input modality. Voice control even caused the most overshoots during the *continuous input* task, indicating the lowest secondary task accuracy. Users preferred voice for *confirmation* but not for *continuous input*.

Hand gesture input was the fastest method except for *confirmation* with a manual task present and caused the least overshoots except for right-handed interaction with a manual task. Both results might be explained by the double-role of the right hand for the interaction and the manipulation task. It also aligns with Wickens' multiple resource model [223], which proposes shared resources in case the same modality is used. In terms of user preference, an interesting shift during conditions could be observed: For *continuous input* without a manual task, the hand input method was rated as the first choice by 67% of the participants but dropped to 25% with a manual task. Even though foot input scored only slightly lower on the subjective workload than both the other modalities during the *continuous input* task and performed mediocrely in terms of task completion times, 50% of the users preferred this method.

A point of discussion is the level of realism for the primary task. The task used in this section was chosen to minimize the influence of expert knowledge on task performance. Compared to radiological interventions or surgery, the tasks at hand might be significantly more complex and thus require more attention. In future work, the role and influence of task complexity need to be taken into account. Another aspect that requires further investigation is the design of gestures. Foot input and voice input both required head movements for continuous input, which might not be necessary. Voice commands might be composed of several words such as "scroll right" to eliminate the need for an additional input modality. Foot input is possible without relying on the position of the head or a foot tap can be used for confirmation.

Overall, the results presented in this section suggest that there is no single superior input modality for multimodal input tasks in augmented reality. Instead, each modality has its advantages and downsides. Therefore, input modalities should be chosen whether the priority lies in speed, accuracy, workload or in meeting personal preferences. These findings align with results obtained during a study on modality output choices while driving: when instructed to steer safely, people tended to use an audio interface even though it was slower, while a visual interface was preferred when the secondary task was prioritized [28].

Further, the results presented in this section show that interfaces that are intended to be used in multitasking scenarios need to be evaluated with a primary task as it influences the outcome in different ways.

6.2 Multitasking during Touchless Pan and Zoom in Radiology

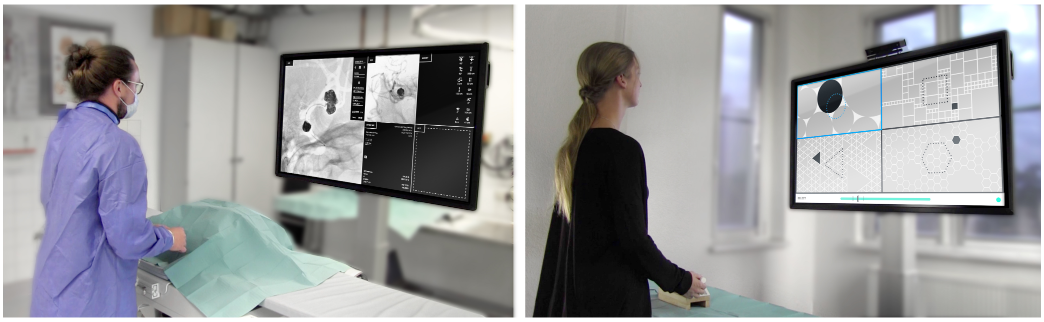


Figure 6.11: Simulated interventional setting for a radiological catheter intervention (left) and study setup (right). Image adapted from Schott et al. [180].

This section describes an approach to reduce the influence of a secondary input task on a primary manual task. The investigated scenario consists of catheter navigation as a high-priority manual task and panning and zooming of radiological images as the secondary task. It is based on the multiple resource theory [223] and passive input modes, which refer to behavior of the user that occurs naturally [146]. Nielsen proposed *Noncommand User Interfaces*, which interpret passive signals and use them to derive the user's intention [140].

6.2.1 Requirement Analysis

A radiological intervention was observed, and a semi-structured interview with an experienced radiologist was conducted. As the main tasks, catheter navigation and image manipulation were identified. Further, challenges in human-computer interaction that emerge when both tasks are performed in an interventional scenario are described.

Catheter navigation

The primary task during radiological catheter interventions is the guidance of a catheter inside the blood vessels to reach the desired structure inside the patient's body. It is done by rotating and pushing the exposed part of a guiding wire (see Figure 6.12). A high level of concentration is required, as this task can be difficult because of respiratory movements or complicated vascular configurations. This task has the highest priority for the radiologist as it serves the primary goal of optimal patient care.

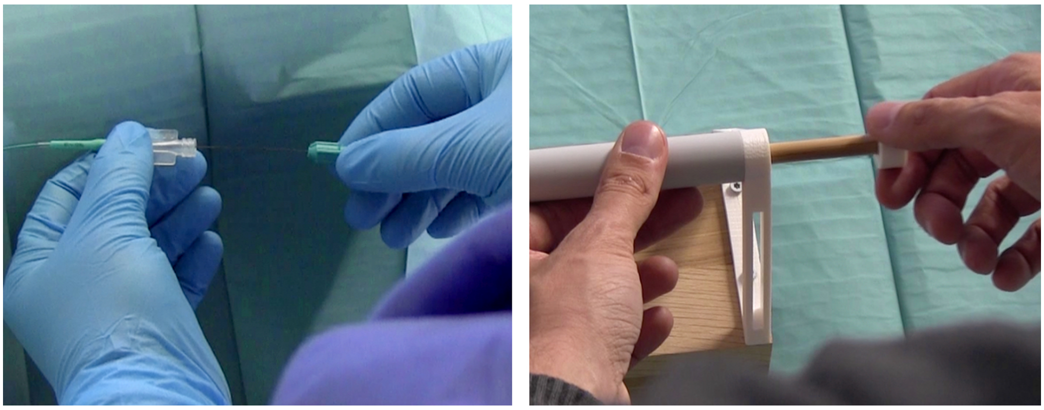


Figure 6.12: Left: Catheter manipulation in a simulated setting. Right: Prototypical input device to simulate a catheter navigation task. Image from Schott et al. [180].

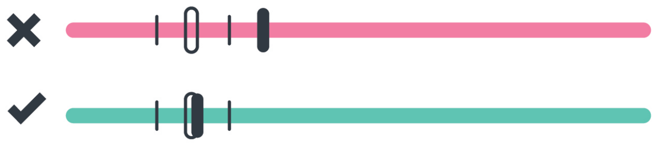


Figure 6.13: Visual feedback for the user for the manual task. The bold line indicates the current position of the input device, the thin lines represent tolerance range and center of the target position. The horizontal bar is green (top) when the current position of the bold bar is inside the tolerance and turns red when not (bottom). Image from Schott et al. [180].

Image Manipulation

Catheter navigation relies on live images because there is no line of sight between the radiologist and the instrument. Fluoroscopy, which is real-time X-ray imaging and, more specifically, angiography, which makes blood vessels visible by injection of an ionide-containing fluid, is used today for guidance of minimally-invasive interventions [198]. One or more screens are usually located in front of the physician, showing fluoroscopic images, previously acquired reference images, patient data and system parameters. The screen shows different viewports in a configurable layout. Interacting with such a system includes scrolling through image data sets, windowing or changing the magnification factor. In current systems, a control panel on the side of the patient table provides joysticks, keys and touchscreens as input devices. The proposed approach achieves almost all of the functionalities above by combining two fundamental interactions: First, interactive elements are selected by pointing at the desired target and confirm the selection. Second, depending on the element selected, one or two dimensions can be manipulated. Frequently used functions with one dimension are scrolling through a stack of images or magnification while two dimensions are required for panning, windowing or rotating 3D volume data sets. For the radiologist, however, image manipulation is a supporting task to gather the required information from the fluoroscopy images. Therefore, this task is not as important as catheter navigation hence can be considered secondary.

Challenges

During the observation, the radiologist regularly wanted to take a closer look at X-ray images, which were displayed on a moveable, ceiling-mounted screen on the other side of the patient table. A range of strategies was observed to achieve this goal:

- **Leaning closer:** The radiologist leans towards the display and over the patient table.
- **Moving the display:** An assistant is instructed to move the display closer.
- **Pointing gesture:** The radiologist points at a specific spot on the screen and instructs an assistant to magnify the corresponding part of the image.
- **Verbal delegation:** In case both hands are occupied, verbal instructions on how to adjust the displayed image section are given to an assistant.

To gain a deeper understanding of the observed behavior and situations, an expert interview was conducted. Overall, the operation of the angiography system was assessed very positively, but restrictions in the operation of the control panel were perceived as disturbing. Navigating through images with the joystick can be done without looking at the input device, which suggests perfected hand-eye coordination due to experience. Different layouts for the large screen can be used, according to the user's preference. The radiologist interviewed prefers a permanent full-screen view in order to get a more detailed view of the images. Even though zooming images is considered as definitely necessary, current problems with this function were emphasized: The radiologist described zooming as possible in principle, but a lot of manual actions are required, which makes it not popular among radiologists. To zoom in an image using the system in the observed setting, the desired image segment must be selected, navigation to image settings must take place by means of a touch interface, and magnification needs to be set to 150%. The image then can be panned using a joystick. The interviewee mentioned that in case zooming is required, all the steps are performed by the radiologist because the verbal description of the target would be too inefficient. Further, one level of magnification seems to be insufficient as continuously variable magnification and the pan was described as indispensable for a useful zooming function.

6.2.2 Interaction Methods

According to the requirement analysis, direct, practical panning and zooming for interventional scenarios seem to be an unresolved HCI challenge. This is especially challenging because the hands are occupied with catheter navigation. Apart from not being available all the time, according to the multiple resources model [223], performing multiple tasks with the same input modality share resources. For this reason, interaction methods that employ other input modalities for image manipulation than the hands to lower the workload when concurrent multitasking is required are proposed. To further lower the cognitive workload, passive input methods inspired by observed user behavior are chosen for panning and zooming. System activation and confirmation of selections, on the other hand, should only be triggered using explicit commands to avoid unintentional activation.

Passive Input for Panning and Zooming

Leaning to the front was observed as a strategy to get a closer look at radiological images. This natural user behavior is utilized to realize a zoom method that does not require explicit input. A direct mapping of the leaning angle to the zoom factor is applied to create an experience that matches the natural behavior. Correspondingly, zooming out is achieved by returning into an upright posture. As the zoom factor is controlled dynamically, a maximal magnification of 300% is set to allow exploration of details.

The second approach to passive zooming involves raising or lowering the eyebrows. Even though this was not observed during the observation, squinting

Table 6.3: hands-free manipulation techniques

Manipulation	Full body movement	Facial expressions
Zoom in	Lean to front	Lower eyebrows
Zoom out	Lean to back	Raise eyebrows
Pan	Head pointing	Head pointing

when observing details might be a natural behavior. However, the range of eyebrow movement is relatively small, which makes it unsuitable for direct mapping on the zoom factor. Therefore, a fixed rate for zooming in is applied when the eyebrows are lowered below a certain threshold. Zooming out is triggered by raising the eyebrows.

Panning is controlled with head motions. Specifically, the image is translated in a way that the area the user points its head at moves towards the center of the screen. Passive panning and zooming can be used simultaneously, with the goal to allow natural exploration of image data without interfering with a manual task.

Active Input for System Control

Image manipulation should not be triggered involuntarily to avoid confusion and prevent delays due to reverting undesired changes. System activation and confirmation of selections, therefore, require active input methods, meaning explicit commands expressed by the user [146]. Two approaches are proposed: head motions such as shaking and nodding and voice commands.

The system can be activated by voice or head gestures. For voice activation, the keyword ”Start” must be uttered. To deactivate the system, the user has to say ”Stop”. When using head gestures, nodding activates the system while shaking the head deactivates it.

Selecting elements such as viewports without using the hands combines already described techniques: Head pointing allows to control a cursor while head gestures or voice commands are used to confirm a selection. The same head gestures as for system activation allow selection (nod) and deselection (shake). Alternatively, ”Select” can be uttered for selection. Deselection is done using the keyword ”Exit”.

6.2.3 Evaluation

A user study following a between-subject design was conducted in a laboratory setting. Two selection methods and two image manipulation methods were

Table 6.4: hands-free selection techniques

Selection	**Voice commands**	**Head gestures**
System activation	"Start"	Nod
Selection (point + confirm)	Head pointing + "Select"	Head pointing + Nod
Deselection	"Exit"	Shake
System deactivation	"Stop"	Shake

compared in terms of task completion time, subjective workload and accuracy. For selection, voice commands were compared to head gestures (nodding and shaking). For Image manipulation, leaning was compared to lifting and lowering the eyebrows. Three independent variables were present. The presence of a primary task was varied between the two subject groups A and B. The selection method and the image manipulation method were varied within subjects. The tasks were designed based on observations but abstracted to eliminate the influence of varying domain knowledge.

Participants

Sixteen participants (10 female, 6 male) between 22 and 38 years of age M = 26.9, SD = 4.3) were recruited via mailing list from the local university. Seven of them were students of human medicine, and the remaining nine participants had a technical or creative background. Among the eight participants in group A who performed the manual task, the right hand was stated as dominant seven times, and no speech disorder was reported. As a spectacle frame might impede eyebrow recognition, only people with low defective vision or normal vision were recruited. Visual impairment was reported six times, while color and vision impairment was reported once. Participants were remunerated with 15 € or 30 €. The remuneration had to be increased due to recruitment problems. Limited previous knowledge in the areas of gesture control, tracking and voice control was reported.

Measures

Three dependent variables were gathered during the study. Task completion time was recorded for each trial. As a measure for primary task performance, the time spent outside the primary task target range was recorded and expressed as a percentage of total task completion time, hereafter called error rate. Subjective workload was assessed using the RTLX questionnaire.

Tasks

Two common tasks during radiological interventions served as a basis for the study task: navigating a catheter and navigating radiological image data. To limit the influence of medical domain knowledge, the tasks were abstracted so that no particular background was required.

The manual, primary task mimics guiding a catheter through a vascular structure. In the real setting, the radiologist achieves this by pulling, pushing or rotating the guiding wire (see left image in Figure 6.12). This is simplified and reduced to pulling and pushing a prototypical input device to match a given target position. The input device consists of a tube with a distance sensor and a rod which can be moved inside the tube (see the right image in Figure 6.12). A slider visualizes the target depth, a tolerance range and the current device position (see Figure 6.13).

Concurrently, a secondary image manipulation task required the participant to (1) activate the image manipulation system, (2) select one of the viewports displayed, (3) pan and zoom an image until it matches a predefined position and size and (4) deactivate the system. Similar to the manual task, the image is abstracted to eliminate the need for the medical background. Figure 6.14 shows the abstraction process starting with the angiogram of an arteriovenous malformation (AVM), which needs to be magnified and centered on a geometric composition in which the dark element needs to match the depicted contour.

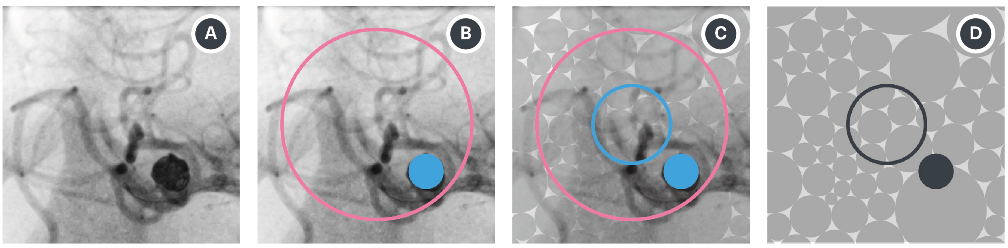


Figure 6.14: Task design based ob medical image data: A radiological image (A) with a target area depicted in blue (B) and the desired location after panning an zooming (C) is reduced to an abstract visualization (D). Image adapted from Schott et al. [180].

Apparatus

The setting consisted of a 75” screen, a manual input device in a table in front of the participant and a Microsoft Kinect v2 (Microsoft Corp., Redmond, WA, USA) mounted on top of the screen and a computer (see Figure 6.16). The Kinect sensor contains a microphone, a depth camera and a color camera. It was used to gather voice commands, detect the user's posture, head direction and facial features.

The prototypical input device consists of a hollow tube and a rod that can be moved in and out of the tube 100 mm. A distance sensor at the end of the tube connected to an Arduino board (Uno R3) measures the insertion depth at a resolution of 1 mm. The data is transmitted to a computer over USB. 3D-printed mounts connect the tube to a wooden stand which houses the Arduino board. This construction allows a comfortable hold on the tube with the non-dominant hand, as it is the case when guiding a catheter wire (see Figure 6.12).

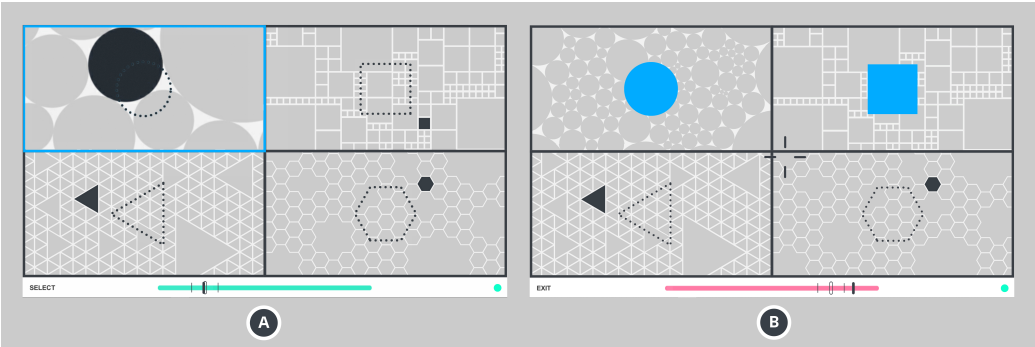


Figure 6.15: The interface consists of four segments, containing an abstract task each. Visual feedback for the primary task is provided at the bottom. A blue border marks an active viewport (A), the cursor for viewport selection is shown in (B). Image from Schott et al. [180].

A prototypical software was implemented using the Unity engine (Unity Technologies, San Francisco, USA) and the Microsoft Kinect v2 SDK (Software Development Kit) (Microsoft Corp., Redmond, WA, USA) to read and interpret the sensor data, display the user interface and to log data during the user study. Head movements control the cursor position for viewport selection and image panning when a viewport is selected. Exponential smoothing is applied to head movement data. The head position is mapped directly on the cursor or image position. Further, head movements are analyzed to detect a nod or shake as head gestures. Nodding activates the system or selects a viewport, depending on the system state. Shaking deselects a viewport or deactivates the system. When the zoom factor is controlled by leaning, the current torso position is considered as the initial position when a viewport is selected. Leaning forward to a maximum of 40 cm (about 20°) is mapped directly to a maximum zoom factor of 300 %. Facial expressions are determined using three facial landmarks. The distance between the midpoint of the eyebrows and the tip of the nose triggers zooming in or out when crossing corresponding thresholds. Voice commands are analyzed using the Kinect SDK language package for keyword detection. The list of keywords is ”start” for system activation, ”select” for viewport selection, ”exit” for viewport deselection and ”stop” for system deactivation.

Further, the software logs user input, task completion times and duration outside the tolerance range of the manual task. The graphical interface which was presented to the user consisted of four viewports, each containing one panning and zooming task and the visual feedback for the manual task (Figure 6.15). A status icon indicates whether the Kinect sensor recognizes a user or not and a text field displays the last input command recognized.

Procedure

The user study took place in a computer laboratory. At first, the participant gave written consent to collect and publish the data gathered in the study and demographic data in anonymized form. After that, a demographic questionnaire was filled out. The participant was assigned to one of both test groups at random. Group A performed image manipulation while concurrently executing the primary manual task while group B performed only the image manipulation tasks.

A table was located in front of the screen at a distance of 180 cm to match the spacial relations of an OR. The height of the screen was adjusted to the participant’s height. The prototypical input device was placed on the table in front of the monitor. Head movements and eyebrow thresholds were calibrated for each participant individually. In case the participant was part of group A and had to perform the manual task concurrently, the position of the prototypical input device was adapted in height and rotated 180° to match the handiness when necessary. The visual feedback for the manual task could be set to match the orientation of the device.

A short introduction to radiological catheter interventions and the tasks of the radiologist was given. The image interaction tasks to be performed by the participant, the input modalities and the graphical user interface were explained by the investigator. All participants were instructed to perform the image manipulation task as fast as possible.

For members of group A, the prototypical input device for the manual task and the corresponding element on the graphical user interface were explained. The participant was instructed to perform the manual task as accurately as possible. Members of group B did not receive this explanation and the visual feedback for the manual task was hidden. For group A participants, a baseline for the manual task was recorded. Only the slider for manual input feedback was visible on the screen during that phase. The baseline acquisition phase took 90 seconds.

At the beginning of the image interaction task, a dimmed user interface indicated the inactive state of the system. After activation with the given input modality, the grid of four equally sized viewports was displayed and the primary task started for group A. At first, a head-controlled cursor had to be used to select one of the viewports using a selection input method (voice or head gestures). The viewport's border turned blue upon selection and panning and zooming mode was entered immediately. Each viewport accommodated geometric patterns in light gray, a small filled black shape and a corresponding, more prominent outline at the center. The small shape had to be translated and zoomed in such a way that it matches the outline. This had to be done using an image manipulation technique (facial expression or leaning). During panning and zooming, the head cursor was not displayed anymore as the head movement was used for panning the image at this point. When the filled shape matched the contour, it turned blue, indicating completion of this subtask. The viewport had to be exited using the current selection method. After exiting a viewport, the head-controlled cursor was displayed again to select the next viewport. After panning and zooming all of the viewport contents correctly, the system had to be deactivated. The user interface during these states can be seen in Figure 6.15.

Each participant performed the image interaction task four times with different combinations of input modalities for selection and image manipulation. The combinations were randomized and arranged in a way that the same input modality was not used two times in a row. Cards showing the input modalities currently in use were placed on the table in front of the participant to avoid confusion during the trials. For each combination, training runs were carried out before the measured runs until the participants stated to be ready. Three runs per condition were performed and task completion time and time outside the tolerance range for the primary task were recorded. After the repeated trials, an RTLX questionnaire was filled out by the participant. The remaining three conditions followed the same procedure. After all four conditions, a structured interview was conducted, the general impression while performing the tasks was assessed and feedback was collected. The overall procedure lasted between one and one and a half hours per participant.

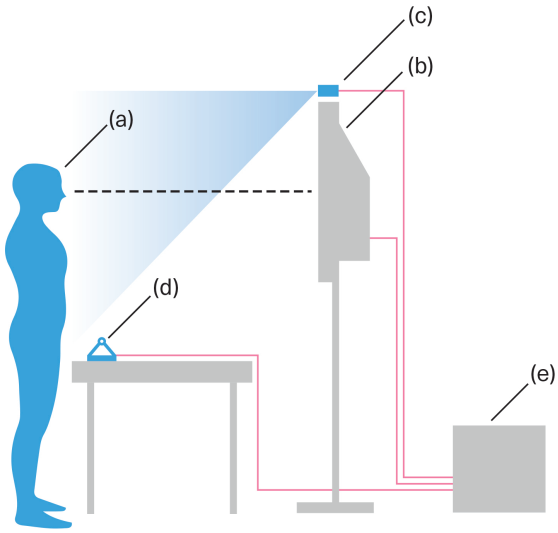


Figure 6.16: Technical setup during the user study: A participant (a) stands in front of the monitor (b) with a Microsoft Kinect v2 (c) mounted on top. The input device for the primary task (d) is connected to a computer (e) that processes the data. Image from Schott et al. [180].

6.2.4 Results

Measures of dependent variables gathered during multiple trials with identical conditions were averaged. Task completion time and subjective workload were analyzed using a three-way ANOVA, and error time was analyzed using a two-way ANOVA. Observations and participants' comments were qualitatively reviewed and clustered by two investigators. In the following, only factors and factor interactions that show significant effects are reported. An overview of the effects found can be seen in Table 6.5. The main effect on task completion time was found for the presence of a primary task (see Figure 6.19) and for the manipulation method (see Figure 6.17). A main effect on the overall RTLX rating was found for the manipulation method (Figure 6.18). Within the sample of the study, the primary task showed a considerable potential effect size ($\eta^2 = 0.043$), although this main effect was not significant.

Table 6.5: Overview of the effects found in the evaluation study

Dependent variable / effect type	Factor	df	F	p	η^2
Task Completion Time					
Main effects	Primary task	1	8.5	0.005*	0.115
	Manipulation method	1	7.97	0.007*	0.108
RTLX rating					
Main effects	Primary task	1	2.9	0.094	0.043
	Manipulation method	1	5.08	0.028*	0.075
Primary Task Timeout					
Main effect	Selection method	1	2.59	0.12	0.084

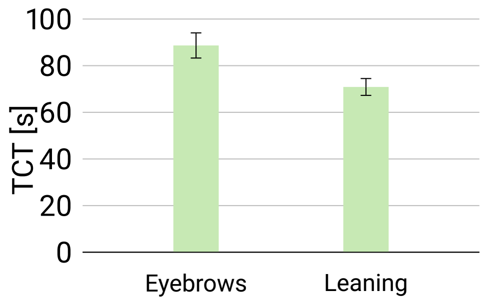


Figure 6.17: Influence of the manipulation method on task completion time. Error bars show standard error

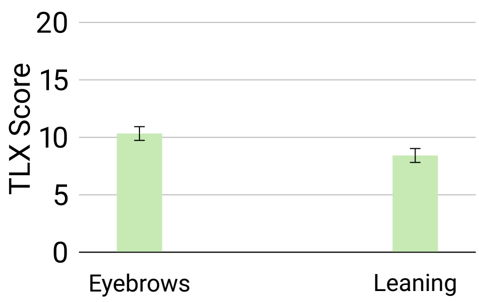


Figure 6.18: Influence of the manipulation method on subjective workload. Error bars show standard error

6.2.5 Discussion

The task completion times show that the secondary interaction takes longer when a primary task has to be fulfilled simultaneously. This is an expected result as it requires additional time to switch focus between both tasks. A nonsignificant main effect between the selection methods and the time spent outside the target range of the primary task was found. Head movements as input methods seem to influence the main task accuracy slightly less than voice commands. The comments of the participants might explain this advantage as head gestures are more comfortable to distinguish and to remember than voice commands. On the other hand, voice commands were described as intuitive compared to shaking the head, which was found uncomfortable and

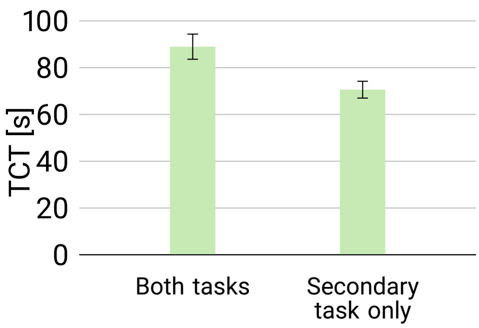


Figure 6.19: Influence of the presence of a primary task on task completion time. Error bars show standard error

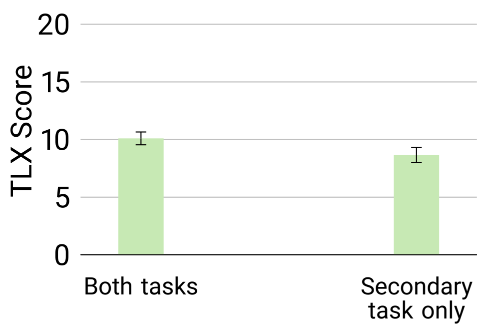


Figure 6.20: Influence of the presence of a primary task on subjective workload. Error bars show standard error

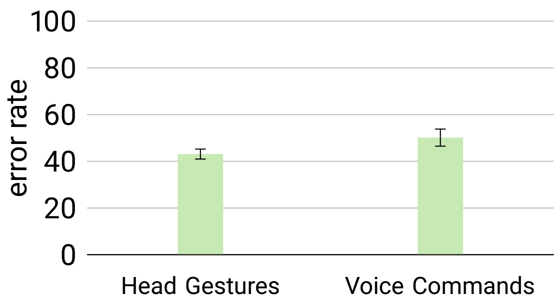


Figure 6.21: Influence of the selection method on the primary task error rate. Error bars show standard error

imprecise. For zooming, leaning performed significantly better than moving the eyebrows in terms of task completion time and subjective workload. Moving the eyebrows was perceived as physically exhausting and can be triggered unintentionally as eyebrow movement may take place for different reasons such as mood changes.

The user study took place in a computer laboratory under ideal conditions. Sources of distraction that might appear in the OR, such as interpersonal communication, audio-visual signals, or noise, are not taken into account. Even though this work deliberately focuses on interaction methods, these factors have to be included when the presented approaches are brought towards clinical application. The abstracted primary task is simplified, meaning that the complexity of the corresponding medical task is not fully reproduced.

This might be counterbalanced by the added difficulty of the visual feedback at the lower edge of the screen. In reality, the correct catheter movements can be seen in the X-ray image. The radiologist, therefore, focuses on the image and does not concurrently shift attention to another part of the screen as the study setup demands. Even though there was a training phase, measuring the primary task baseline before the actual tasks might have been influenced by learning effects. This is ,however, only relevant for the presented study with inexperienced participants. Domain experts most likely will not experience a learning curve when performing familiar tasks.

Overall, the proposed approach for finding suitable, natural input methods seems to produce promising results. Derived from observations, leaning allows natural, direct zooming when the hands are occupied and is less exhausting than moving the eyebrows. Head gestures are easy to remember than voice commands even though they might be too physically demanding for more prolonged use. However, valuable alternatives to hand gestures or voice commands were presented, which extend the available input vocabulary for demanding scenarios.

6.3 Conclusion

This chapter explored the role of the interaction tasks when a more critical, manual task is present, as it is the case in medical scenarios. In section 6.1, the mutual influence of different interaction methods and a manual task were investigated. The results suggest that voice input increases the error rate for the manual task the least. In terms of user preference, voice commands were favored by most participants for confirmation tasks regardless of the presence of a manual task, while continuous input preferences shifted from hand input to foot input during a manual task. It can be said that multimodal systems need to be tested with a primary task as it may influence the outcome. In the presented setting, no single input modality excelled in all the measured dimensions. This means that the choice of modalities requires prioritized goals for the desired interaction method. Section 6.2 presented an approach aimed at creating a lower mental workload by employing natural user movements for secondary tasks. Derived from observations in the OR, a zooming method based on leaning was implemented and compared to eyebrow movements in a user study. Further, touchless methods for panning and system activation were employed. Leaning performed significantly better than facial expressions for zooming.

This chapter is based on the following publications:

D. Schott, B. Hatscher, F. Joeres, M. Gabele, S. Hußlein, and C. Hansen. "Lean-Interaction: passive image manipulation in concurrent multitasking". In: *Proceedings of Graphics Interface 2020*. GI 2020. Canadian Human-Computer Communications Society / Société canadienne du dialogue humain-machine, 2020, pp. 404–412.

A. Solovjova, B. Hatscher, and C. Hansen. "Influence of augmented reality interaction on a primary task for the medical domain". In: *Mensch und Computer 2019 - Workshopband*. Bonn: Gesellschaft für Informatik e.V, 2019, pp. 325–330.

The multimodal input concepts for augmented reality described in Section 6.1 were implemented and evaluated by Alina Solovjova within her master's thesis [190]. The interaction concepts for touchless paning and zooming described in Section 6.2 were implemented and first evaluated by Danny Schott within his master's thesis [179]. Both works were supervised by the author of this thesis.

7

Conclusion

This thesis presented methods that allow direct, hands-free control of medical image data to overcome HCI limitations and workarounds during minimally-invasive interventions. For this purpose, a variety of alternative input channels such as the feet, body and head movements, eye gaze and voice commands were investigated regarding their suitability for basic image manipulation tasks.

To understand the differences between direct input and proxy user input in a medical setting, a comparative study in the context of an MRI-guided minimally invasive percutaneous needle intervention was presented. Hand gestures yielded a higher subjective usability rating than gestural and verbal communication with an assistant at comparable task completion times. A higher level of control, the potential to save time and independence from an assistant were mentioned as advantages in post-test interviews. Besides improvements in gesture design, feedback and sensor placement, possible negative implications on supporting staff and observers outside the scan room were mentioned as downsides as following an intervention is more difficult with direct interaction.

Foot-based input methods were investigated as foot pedals are a familiar input device in clinical practice and using the feet keeps the hands free for medical tasks. Feet, in general, are suitable for coarse selection tasks. In the context of hands-free image manipulation, tactile floors allow input without body-worn hardware and are relatively easy to learn. However, they rely on clear, descriptive visual feedback and hardware with an adequate resolution. Pressure-sensing inlays and wearable sensors are advantageous over conventional foot pedals as they cannot get lost. Plantar pressure distribution as input method was found to be difficult to extend beyond directional input as overlapping pressure zones are hard to distinguish due to physiological differences and preferences between users. Rotating the foot on the heel was investigated as an input method as it causes less fatigue than moving the foot [207]. In this thesis, a method that required sliding over the floor with the ball of the foot was found to perform best in an image browsing task with one foot.

Multiple input modalities were combined to overcome the availability issues and limitations of single modalities. In the context of simple image manipulation tasks, medical students performed better with foot gestures than with voice commands and hand gestures. Further, foot input was rated as more suitable for the desired use case. As users of multimodal systems might focus on finding one optimal modality instead of using multiple modalities or switching [9], this result highlights the potential of foot-based input methods in interventional scenarios even though alternatives might be valuable in demanding situations.

The combination of eye-tracking and virtual pedals on a tactile floor extended the foot input methods by coarse pointing. However, visual checks on the foot position easily inhibited multimodal interaction, which indicates the need for more elaborate visual feedback for multimodal approaches.

Concerning the overall goal of this dissertation, to provide direct interaction methods while performing a medical intervention, the interplay of secondary human-computer interaction tasks and primary manual tasks was researched. Results suggest that voice commands influence primary task performance the least while user preference favors voice commands for discrete input but foot-based heel rotation for continuous manipulation. In the context of a catheter intervention, leaning towards a screen was observed when image details need to be assessed. Leaning performed faster and caused less workload than eyebrow movements, suggesting natural movements as a promising approach for secondary interactions.

Overall, several approaches for hands-free interaction are proposed in this thesis and their feasibility is demonstrated. While foot-based input seems capable of providing more functionality than traditional pedals and may be a suitable alternative for voice commands or hand gestures, the available space and possible variations in foot movement and placement should be considered. Secondary input methods intended to be performed concurrently to a primary task may lower the subjective workload when derived from natural behavior.

Limitations Even though the presented approaches provide valuable insights towards direct physician-computer interaction, some limitations have to be considered. The proposed methods are designed with specific use-cases from the medical domain in mind. This leads to a range of boundary conditions and restrictions such as a standing position, sterility, concurrent manual tasks or environmental influences that might not necessarily play a role in a wide range of applications. The importance of minimally-invasive procedures in the light of today's demographic change, however, may justify the effort.

Evaluation of the presented input approaches took place in a controlled setting with a simplified set of tasks and no environmental sources of distraction. The investigated discrete, one DoF and two DoF interaction tasks are relevant in clinical practice but do not cover all functionalities required during an intervention. The adjustment of interventional planning or validation of instrument positions might require measuring distances and angles. Accurate pointing techniques are not covered within this thesis. Interventional scenarios bear several sources of distraction and interruption. Monitoring of medical equipment, communication with the medical team, or consultation with colleagues requires shared attention. As the workflow might be influenced by personal preferences, available hardware and experience, it is difficult to model these factors in a way that yields generalizable results.

The complexity of secondary interaction tasks and primary manual tasks is one factor that can be adjusted to match clinical levels more closely in future studies. In the long run, interaction techniques need to be evaluated in simulated interventional scenarios concerning time pressure, demands with accuracy, sources of distraction and duration of interventions. Virtual reality settings, which have already been used to evaluate input methods [162], can be extended to meet these requirements.

The proposed interaction methods and techniques were not compared to proxy-user interaction or state-of-the-art input methods such as touchscreens or joysticks. This might limit the direct influence of this research on medical practice. However, keeping the hands free requires unfamiliar approaches whose advantages and pitfalls need to be understood before similar or better performance than established methods can be expected. This thesis provides basic findings to inform the development of prototypes, which is the next logical step towards hands-free interaction.

Outlook Given the complexity of modern operating rooms and medical settings, future work needs to improve hands-free interaction methods, evaluate them in clinical settings and look in-depth on how to integrate them into existing environments and workflows in a meaningful way.

Hands-free interaction might not necessarily replace proxy-user interaction completely but can serve as a valuable alternative when required. Compared to tailor-made interventional user interfaces, assistants can provide support in complex, unanticipated situations, which is hard to account for in software. Situations where assistance is not available or lacks the required expertise; however benefit from a fast and easy direct input method that is immediately available. This further requires concepts for handing over control so that radiologists or surgeons can seamlessly switch between direct and mediated interaction while focusing on their main task.

In the OR of the future, context-aware systems will play an essential role by providing services that are relevant for the task at hand [152]. Direct clinician-computer interfaces can adapt based on additional contextual information in two ways: input channels with a smaller range of expressions than the hands or verbal task delegation can be sufficient when only functions relevant in the current workflow step or situation are provided. On the other hand, workflow information can be used to automatically determine and communicate which input channels are expected to be reliable at the current point in time. Voice commands, for example, can be suggested during quiet phases.

Research projects towards interconnected, intelligent operating rooms such as SCOT [143], MD PnP [220], and OR.NET [99, 221] bear the potential to integrate different devices and systems into one interconnected OR system [156]. However, usability has to be discussed critically in this context [156]. Sterile, direct, potentially hands-free input methods might provide a flexible user interface that allows clinicians to leverage the advantages of intelligent operating rooms regardless of the circumstances.

In the long run, this work hopefully supports the development of systems and environments that reduce interruptions, improve minimally-invasive workflows, allow more efficient treatments and contribute to the challenges demographic changes might hold in the future.

Bibliography

[1] A. Zaman, A. Roy, K. Fatema, N. J. Farin, T. Doring, and R. Malaka. "Explore Voice and Foot-based Interaction Techniques to Navigate 2D Radiological Images in the Virtual Reality Operation Theatre". In: *2019 22nd International Conference on Computer and Information Technology (ICCIT)*. 2019, pp. 1–7 (cit. on p. 89).

[2] N. Abolhassani, R. Patel, and M. Moallem. "Needle insertion into soft tissue: A survey". In: *Medical engineering & physics* 29.4 (2007), pp. 413–431 (cit. on p. 5).

[3] P. D. Adamczyk and B. P. Bailey. "If not now, when?: the effects of interruption at different moments within task execution". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2004, pp. 271–278 (cit. on p. 109).

[4] D. Adams. "Wireless laptops in the classroom (and the Sesame Street syndrome)". In: *Communications of the ACM* 49.9 (2006), pp. 25–27 (cit. on p. 109).

[5] R. F. Adler and R. Benbunan-Fich. "Juggling on a high wire: Multitasking effects on performance". In: *International Journal of Human-Computer Studies* 70.2 (2012), pp. 156–168 (cit. on p. 109).

[6] A. Alapetite. "Impact of noise and other factors on speech recognition in anaesthesia". In: *International journal of medical informatics* 77.1 (2008), pp. 68–77 (cit. on p. 16).

[7] J. Alexander, T. Han, W. Judd, P. Irani, and S. Subramanian. "Putting your best foot forward: investigating real-world mappings for foot-based gestures". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2012, pp. 1229–1238 (cit. on pp. 11, 64, 71).

[8] F. Alvarez-Lopez, M. F. Maina, and F. Saigí-Rubió. "Use of Commercial Off-The-Shelf Devices for the Detection of Manual Gestures in Surgery: Systematic Literature Review". In: *Journal of medical Internet research* 21.5 (2019), e11925 (cit. on pp. 1, 8, 10).

[9] S. Arevalo, S. Miller, M. Janka, and J. Gerken. “What’s behind a choice? Understanding Modality Choices under Changing Environmental Conditions”. In: *2019 International Conference on Multimodal Interaction*. New York, New York, USA: ACM / Association for Computing Machinery, 2019, pp. 291–301 (cit. on p. 138).

[10] J. A. Argoty and P. Figueroa. “Design and development of a prototype of an interactive hospital room with Kinect”. In: *Proceedings of the XV International Conference on Human Computer Interaction*. New York, NY: ACM, 2014, pp. 1–4 (cit. on p. 16).

[11] M. Ashmore, A. T. Duchowski, and G. Shoemaker. “Efficient eye pointing with a fisheye lens”. In: *Proceedings of Graphics interface 2005*. 2005, pp. 203–210 (cit. on p. 15).

[12] T. Augsten, K. Kaefer, R. Meusel, C. Fetzer, D. Kanitz, T. Stoff, T. Becker, C. Holz, and P. Baudisch. “Multitoe: high-precision interaction with back-projected floors based on high-resolution multi-touch input”. In: *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. 2010, pp. 209–218 (cit. on pp. 11, 12).

[13] D. Bachmann, F. Weichert, and G. Rinkenauer. “Evaluation of the Leap Motion Controller as a New Contact-Free Pointing Device”. In: *Sensors* 15.1 (2015), pp. 214–233 (cit. on p. 29).

[14] R. Bade, F. Ritter, and B. Preim. “Usability comparison of mouse-based interaction techniques for predictable 3d rotation”. In: *Smart graphics*. Vol. 5. 2005 (cit. on p. 94).

[15] A. Bangor, P. Kortum, and J. Miller. “Determining what individual SUS scores mean: Adding an adjective rating scale”. In: *Journal of usability studies* 4.3 (2009), pp. 114–123 (cit. on pp. 21, 38).

[16] A. Bangor, P. T. Kortum, and J. T. Miller. “An Empirical Evaluation of the System Usability Scale”. In: *International Journal of Human–Computer Interaction* 24.6 (2008), pp. 574–594 (cit. on pp. 21, 38).

[17] J. Barkhausen, T. Kahn, G. A. Krombach, C. K. Kuhl, J. Lotz, D. Maintz, J. Ricke, S. O. Schönberg, T. J. Vogl, and F. K. Wacker. “White Paper: Interventionelle MRT: Status Quo und Entwicklungspotenzial unter ökonomischen Perspektiven, Teil 2: Therapeutische und onkologische Anwendungen”. In: *RoFo : Fortschritte auf dem Gebiete der Rontgenstrahlen und der Nuklearmedizin* 189.11 (2017), pp. 1047–1054 (cit. on p. 23).

[18] R. L. Barnett. “Foot controls: Riding the pedal”. In: *The Ergonomics Open Journal* 2 (2009), pp. 13–16 (cit. on pp. 13, 62).

[19] T. Baudel and M. Beaudouin-Lafon. “Charade: remote control of objects using free-hand gestures”. In: *Commun. ACM* 36.7 (1993), pp. 28–35 (cit. on p. 9).

[20] N. Beckers, S. Schreiner, P. Bertrand, B. Reimer, B. Mehler, D. Munger, and J. Dobres. “Comparing the Demands of Destination Entry using Google Glass and the Samsung Galaxy S4”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 58.1 (2014), pp. 2156–2160 (cit. on p. 110).

[21] R. Benbunan-Fich and G. E. Truman. “Technical Opinion: Multitasking with Laptops During Meetings”. In: *Commun. ACM* 52.2 (2009), pp. 139–141 (cit. on p. 109).

[22] D. Black, M. Unger, N. Fischer, R. Kikinis, H. Hahn, T. Neumuth, and B. Glaser. “Auditory display as feedback for a novel eye-tracking system for sterile operating room interaction”. In: *International Journal of Computer Assisted Radiology and Surgery* 13.1 (2018), pp. 37–45 (cit. on p. 15).

[23] R. A. Bolt. *“Put-that-there”: Voice and gesture at the graphics interface*. Vol. 14. ACM, 1980 (cit. on pp. 9, 17).

[24] S. Boring, M. Jurmu, and A. Butz. “Scroll, tilt or move it”. In: *OZCHI 2009*. ACM international conference proceeding series. New York, NY: ACM Press, 2009, p. 161 (cit. on p. 10).

[25] A. F. Botero-Ospina, S. I. Duque-Vallejo, J. F. Ochoa-Gómez, and A. M. Hernández-Valdivieso. “Touchless control module for diagnostic images at the surgery room using the Leap Motion system and 3D Slicer Software”. In: *Revista Facultad de Ingeniería Universidad de Antioquia* 82 (2017), p. 40 (cit. on p. 78).

[26] J. Bressem and S. H. Ladewig. “Rethinking gesture phases: Articulatory features of gestural movement?” In: *Semiotica* 2011.184 (2011), pp. 53–91 (cit. on p. 8).

[27] J. Brooke et al. “SUS-A quick and dirty usability scale”. In: *Usability evaluation in industry* 189.194 (1996), pp. 4–7 (cit. on pp. 21, 34).

[28] D. P. Brumby, S. C. Davies, C. P. Janssen, and J. J. Grace. “Fast or safe?” In: *29th Annual CHI conference on human factors in computing systems*. New York: Association for Computing Machinery, 2011, p. 473 (cit. on p. 120).

[29] M. Carter, E. Velloso, J. Downs, A. Sellen, K. O'Hara, and F. Vetere. "PathSync: Multi-User Gestural Interaction with Touchless Rhythmic Path Mimicry". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2016, pp. 3415–3427 (cit. on p. 15).

[30] I. Chatterjee, R. Xiao, and C. Harrison. "Gaze+Gesture". In: *the 2015 ACM*. 2015, pp. 131–138 (cit. on p. 106).

[31] E. Choke and R. Sayers. "adaptation of vascular surgery in the interventional era". In: *Heart* 101.5 (2015), pp. 342–345 (cit. on p. 5).

[32] E. Christoforou, E. Akbudak, A. Ozcan, M. Karanikolas, and N. V. Tsekos. "Performance of interventions with manipulator-driven real-time MR guidance: implementation and initial in vitro tests". In: *Magnetic resonance imaging* 25.1 (2007), pp. 69–77 (cit. on p. 24).

[33] K. F. Chu and D. E. Dupuy. "Thermal ablation of tumours: biological mechanisms and advances in therapy". In: *Nature Reviews Cancer* 14.3 (2014), pp. 199–208 (cit. on p. 5).

[34] N. T. Clancy, G. P. Mylonas, G.-Z. Yang, and D. S. Elson. "Gaze-contingent autofocus system for robotic-assisted minimally invasive surgery". In: *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference* 2011 (2011), pp. 5396–5399 (cit. on p. 15).

[35] A. Çöltekin, J. Hempel, A. Brychtova, I. Giannopoulos, S. Stellmach, and R. Dachselt. "Gaze and feet as additional input modalities for interacting with geospatial interfaces". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Prague, Czech Republic, 2016 (cit. on p. 90).

[36] S. Cronin and G. Doherty. "Touchless computer interfaces in hospitals: A review". In: *Health informatics journal* (2018), pp. 1325–1342 (cit. on pp. 1, 7, 8, 9, 10, 16, 18, 38).

[37] Díaz-Oreiro, López, Quesada, and Guerrero. "Standardized Questionnaires for User Experience Evaluation: A Systematic Literature Review". In: *Proceedings* 31.1 (2019), p. 14 (cit. on p. 22).

[38] B. J. Dixon, M. J. Daly, H. Chan, A. D. Vescan, I. J. Witterick, and J. C. Irish. "Surgeons blinded by enhanced navigation: the effect of augmented reality on attention". In: *Surgical Endoscopy* 27.2 (2013), pp. 454–461 (cit. on p. 110).

[39] L. C. Ebert, P. M. Flach, M. J. Thali, and S. Ross. "Out of touch–A plugin for controlling OsiriX with gestures using the leap controller". In: *Journal of Forensic Radiology and Imaging* 2.3 (2014), pp. 126–128 (cit. on p. 9).

[40] L. C. Ebert, G. Hatch, G. Ampanozi, M. J. Thali, and S. Ross. "You can't touch this touch-free navigation through radiological images". In: *Surgical innovation* 19.3 (2012), pp. 301–307 (cit. on pp. 8, 17, 78, 79).

[41] W. K. English, D. C. Engelbart, and M. L. Berman. "Display-selection techniques for text manipulation". In: *IEEE Transactions on Human Factors in Electronics* 1 (1967), pp. 5–15 (cit. on p. 11).

[42] A. Esteves, E. Velloso, A. Bulling, and H. Gellersen. "Orbits". In: *the 28th Annual ACM Symposium*, pp. 457–466 (cit. on p. 15).

[43] A. Évain, F. Argelaguet, G. Casiez, N. Roussel, and A. Lécuyer. "Design and Evaluation of Fusion Approach for Combining Brain and Gaze Inputs for Target Selection". In: *Frontiers in neuroscience* 10 (2016), p. 454 (cit. on pp. 13, 15, 93).

[44] F. Fischbach, J. Bunke, M. Thormann, G. Gaffke, K. Jungnickel, J. Smink, and J. Ricke. "MR-Guided Freehand Biopsy of Liver Lesions With Fast Continuous Imaging Using a 1.0-T Open MRI Scanner: Experience in 50 Patients". In: *CardioVascular and Interventional Radiology* 34.1 (2011), pp. 188–192 (cit. on pp. 5, 23, 26, 29).

[45] G. S. Fischer, A. Deguet, C. Csoma, R. H. Taylor, L. Fayad, J. A. Carrino, S. J. Zinreich, and G. Fichtinger. "MRI image overlay: application to arthrography needle insertion". In: *Computer aided surgery : official journal of the International Society for Computer Aided Surgery* 12.1 (2007), pp. 2–14 (cit. on p. 24).

[46] P. M. Fitts. "The information capacity of the human motor system in controlling the amplitude of movement". In: *Journal of experimental psychology* 47 6 (1954), pp. 381–391 (cit. on p. 16).

[47] T. Fitzke, N. Krail, F. Kroll, L. Ohlrogge, F. Schröder, L. Spillner, A.-K. Voll, F. Dylla, M. Herrlich, and R. Malaka. "Fußbasierte Interaktion mit Computersystemen im Operationssaal". In: *CURAC 2015 - Tagungsband*. Bremen: Hahn, Horst K. et al., 2015, pp. 49–54 (cit. on p. 57).

[48] J. Fritz, N. Tzaribachev, C. Thomas, J. A. Carrino, C. D. Claussen, J. S. Lewin, and P. L. Pereira. "Evaluation of MR imaging guided steroid injection of the sacroiliac joints for the treatment of children with refractory enthesitis-related arthritis". In: *European radiology* 21.5 (2011), pp. 1050–1057 (cit. on p. 29).

[49] M. Fritzsche, J. Saenz, and F. Penzlin. "A large scale tactile sensor for safe mobile robot manipulation". In: *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. 2016, pp. 427–428 (cit. on p. 43).

[50] K. Fujii, G. Gras, A. Salerno, and G.-Z. Yang. "Gaze gesture based human robot interaction for laparoscopic surgery". In: *Medical image analysis* 44 (2018), pp. 196–214 (cit. on p. 15).

[51] K. Fukahori, D. Sakamoto, and T. Igarashi. "Exploring Subtle Foot Plantar-based Gestures with Sock-placed Pressure Sensors". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. New York, NY, USA: ACM, 2015, pp. 3019–3028 (cit. on pp. 11, 57, 64).

[52] P. Gailloud. "A large display is a powerful tool to reduce radiation exposure during single-plane fluoroscopically guided procedures". In: *AJR. American journal of roentgenology* 204.4 (2015), W483–5 (cit. on p. 90).

[53] L. Gallo, A. P. Placitelli, and M. Ciampi. "Controller-free exploration of medical image data: Experiencing the Kinect". In: *Computer-based medical systems (CBMS), 2011 24th international symposium on*. 2011, pp. 1–6 (cit. on pp. 7, 8, 17, 79).

[54] C. P. Gerba, A. L. Wuollet, P. Raisanen, and G. U. Lopez. "Bacterial contamination of computer touch screens". In: *American journal of infection control* 44.3 (2016), pp. 358–360 (cit. on pp. 1, 6).

[55] D. Gergle and D. S. Tan. "Experimental Research in HCI". In: *Ways of Knowing in HCI*. New York, NY: Springer New York, 2014, pp. 191–227 (cit. on p. 19).

[56] F. Göbel, K. Klamka, A. Siegel, S. Vogt, S. Stellmach, and R. Dachselt. "Gaze-supported Foot Interaction in Zoomable Information Spaces". In: (2013) (cit. on p. 90).

[57] V. M. González and G. Mark. ""Constant, constant, multi-tasking craziness"". In: *CHI 2004*. CHI letters. New York, NY: Association for Computing Machinery, 2004, pp. 113–120 (cit. on p. 109).

[58] C. Graetzel, T. Fong, S. Grange, and C. Baur. "A non-contact mouse for surgeon-computer interaction". In: *Technology and Health Care* 12.3 (2004), pp. 245–257 (cit. on p. 6).

[59] S. Grange, T. Fong, and C. Baur. "M/ORIS: a medical/operating room interaction system". In: *Proceedings of the 6th international conference on Multimodal interfaces*. 2004, pp. 159–166 (cit. on pp. 6, 7).

[60] F. Güttler, J. C. Rump, and U. Teichgraeber. "MR-compatible wireless communication system for the interventional open high-field MRI". In: *Proc. Intl. Soc. Mag. Reson. Med.* Vol. 17. 2009, p. 2576 (cit. on p. 27).

[61] F. Güttler. *Gesture control: The hand for the contactless control of the MRI during interventional procedures*. 2018 (cit. on p. 24).

[62] T. Han, J. Alexander, A. Karnik, P. Irani, and S. Subramanian. "Kick". In: *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*. New York, NY: ACM, 2011, p. 29 (cit. on p. 11).

[63] C. Hansen, A. Köhn, S. Schlichting, F. Weiler, S. Zidowitz, M. Kleemann, and H.-O. Peitgen. "Intraoperative modification of resection plans for liver surgery". In: *International Journal of Computer Assisted Radiology and Surgery* 3.3-4 (2008), p. 291 (cit. on p. 1).

[64] D. W. Hansen and Q. Ji. "In the eye of the beholder: A survey of models for eyes and gaze". In: *IEEE transactions on pattern analysis and machine intelligence* 32.3 (2009), pp. 478–500 (cit. on p. 14).

[65] S. Harada, J. A. Landay, J. Malkin, X. Li, and J. A. Bilmes. "The vocal joystick:: evaluation of voice-based cursor control techniques". In: *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*. 2006, pp. 197–204 (cit. on pp. 16, 79).

[66] C. Harrison, S. Ramamurthy, and S. E. Hudson. "On-body interaction". In: *Proceedings of the sixth International Conference on Tangible, Embedded and Embedded Interaction : February 19-22, 2012, Queens Human Media Lab, Kingston, ON Canada*. [Place of publication not identified]: ACM, 2012, p. 69 (cit. on p. 10).

[67] S. G. Hart. "NASA-task load index (NASA-TLX); 20 years later". In: *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 50. 2006, pp. 904–908 (cit. on pp. 20, 21, 33, 51, 59, 68, 95, 114).

[68] S. G. Hart and L. E. Staveland. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research". In: *Advances in psychology*. Vol. 52. Elsevier, 1988, pp. 139–183 (cit. on pp. 20, 95).

[69] M. Hassenzahl, F. Koller, and M. Burmester. "Der User Experience (UX) auf der Spur: Zum Einsatz von www.attrakdiff.de". In: *Tagungsband UP08*. Stuttgart: Fraunhofer Verlag, 2008, pp. 78–82 (cit. on p. 22).

[70] B. Hatscher and C. Hansen. "Hand, Foot or Voice: Alternative Input Modalities for Touchless Interaction in the Medical Domain". In: *Proceedings of the 20th ACM International Conference on Multimodal Interaction*. ICMI '18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 145–153 (cit. on pp. 76, 77, 80, 83).

[71] B. Hatscher, M. Luz, and C. Hansen. "Foot Interaction Concepts to Support Radiological Interventions". In: *Mensch und Computer 2017 - Tagungsband*. Regensburg: Gesellschaft für Informatik e.V, 2017, pp. 93–104 (cit. on p. 74).

[72] B. Hatscher, M. Luz, and C. Hansen. "Foot Interaction Concepts to Support Radiological Interventions". In: *i-com* 17.1 (2018), pp. 3–13 (cit. on pp. 11, 65, 67, 74).

[73] B. Hatscher, M. Luz, L. E. Nacke, N. Elkmann, V. Müller, and C. Hansen. "GazeTap: Towards Hands-free Interaction in the Operating Room". In: *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. ICMI '17. New York, NY, USA: Association for Computing Machinery, 2017, pp. 243–251 (cit. on p. 96).

[74] B. Hatscher, A. Mewes, E. Pannicke, U. Kägebein, F. Wacker, C. Hansen, and B. Hensen. "Touchless scanner control to support MRI-guided interventions". In: *International Journal of Computer Assisted Radiology and Surgery* 15 (2020), pp. 545–553 (cit. on pp. 28, 29, 31, 32, 36, 37, 40).

[75] B. Hatscher, S. Wagner, L. Grimaldi, M. Fritzsche, N. Elkmann, and C. Hansen. "Navigation in medizinischen Bilddaten mittels eines taktilen Fussbodens". In: *Proceedings of the Annual Meeting of the German Society of Computer- and Robot-Assisted Surgery (CURAC)*. Bern, Schweiz, 2016, pp. 33–38.

[76] S. Heo, M. Annett, B. Lafreniere, T. Grossman, and G. Fitzmaurice. “No Need to Stop What You’re Doing: Exploring No-Handed Smartwatch Interaction”. In: *Proceedings of Graphics Interface*. 2017, pp. 107–116 (cit. on p. 78).

[77] J. Hettig, A. Mewes, O. Riabikin, M. Skalej, B. Preim, and C. Hansen. “Exploration of 3D Medical Image Data for Interventional Radiology using Myoelectric Gesture Control”. In: *Proceedings of Eurographics Workshop on Visual Computing for Biology and Medicine*. Chester, UK, 2015, pp. 177–185 (cit. on pp. 6, 8, 24).

[78] J. Hettig, P. Saalfeld, M. Luz, M. Becker, M. Skalej, and C. Hansen. “Comparison of gesture and conventional interaction techniques for interventional neuroradiology”. In: *International Journal of Computer Assisted Radiology and Surgery* (2017), pp. 1643–1653 (cit. on pp. 6, 38, 77, 79, 90, 91).

[79] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani. “Consumed endurance: a metric to quantify arm fatigue of mid-air interactions”. In: *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. 2014, pp. 1063–1072 (cit. on p. 10).

[80] E. R. Hoffmann. “A comparison of hand and foot movement times”. In: *Ergonomics* 34.4 (1991), pp. 397–406 (cit. on p. 12).

[81] M. B. Hoy. “Alexa, Siri, Cortana, and more: an introduction to voice assistants”. In: *Medical reference services quarterly* 37.1 (2018), pp. 81–88 (cit. on pp. 16, 89).

[82] A. Hübler, C. Hansen, O. Beuing, M. Skalej, and B. Preim. “Workflow Analysis for Interventional Neuroradiology using Frequent Pattern Mining”. In: *Proceedings of the Annual Meeting of the German Society of Computer- and Robot-Assisted Surgery*. Munich, 2014, pp. 165–168 (cit. on pp. 41, 62).

[83] D. Huggins-Daines, M. Kumar, A. Chan, A. W. Black, M. Ravishankar, and A. I. Rudnicky. “Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices”. In: *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*. Vol. 1. 2006, p. I (cit. on p. 81).

[84] T. Igarashi and J. F. Hughes. “Voice as sound: using non-verbal voice input for interactive control”. In: *Proceedings of the 14th annual ACM symposium on User interface software and technology*. 2001, pp. 155–156 (cit. on pp. 16, 79).

[85] International Organization for Standardization. *Ergonomics of human-system interaction: Part 210: Human-centred design for interactive systems*. Tech. rep. 9241-210:2019(en). Geneva, Switzerland (cit. on pp. 21, 22).

[86] International Organization for Standardization. *Ergonomics of human-system interaction: Part 420: Selection of physical input devices*. Tech. rep. 9241-420:2011(en). Geneva, Switzerland, 2011 (cit. on pp. 111, 114).

[87] International Organization for Standardization. *Road vehicles — Transport information and control systems — Detection-response task (DRT) for assessing attentional effects of cognitive load in driving*. Tech. rep. 17488:2016(en). Geneva, Switzerland (cit. on p. 110).

[88] A. Ionescu. "A mouse in the OR Ambidextrous". In: *Stanford Univ Journal of Design* 30 (2006), p. 2 (cit. on pp. 1, 6).

[89] S. Ivar Seldinger. "Catheter replacement of the needle in percutaneous arteriography: a new technique". In: *Acta radiologica* 49.suppl_434 (2008), pp. 47–52 (cit. on p. 5).

[90] J. Liu, Z. Wang, L. Zhong, J. Wickramasuriya, and V. Vasudevan. "uWave: Accelerometer-based personalized gesture recognition and its applications". In: *2009 IEEE International Conference on Pervasive Computing and Communications*. 2009, pp. 1–9 (cit. on p. 57).

[91] R. Jacob and S. Stellmach. "What you look at is what you get: gaze-based user interfaces". In: *interactions* 23.5 (2016), pp. 62–65 (cit. on p. 15).

[92] R. J. K. Jacob. "The use of eye movements in human-computer interaction techniques: what you look at is what you get". In: *ACM Transactions on Information Systems (TOIS)* 9.2 (1991), pp. 152–169 (cit. on pp. 15, 90).

[93] A. K. Jain. *Fundamentals of digital image processing*. Englewood Cliffs, NJ: Prentice Hall, 1989 (cit. on p. 8).

[94] S. Jalaliniya, J. Smith, M. Sousa, L. Büthe, and T. Pederson. "Touchless interaction with medical images using hand & foot gestures". In: *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*. 2013, pp. 1265–1274 (cit. on pp. 1, 13, 48).

[95] R. Johnson, K. O'Hara, A. Sellen, C. Cousins, and A. Criminisi. "Exploring the Potential for Touchless Interaction in Image-guided Interventional Radiology". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11. New York, NY, USA: ACM, 2011, pp. 3323–3332 (cit. on pp. 1, 6, 7, 8, 41, 42, 75, 77, 91).

[96] U. Kägebein, F. Godenschweger, B. S. R. Armstrong, G. Rose, F. K. Wacker, O. Speck, and B. Hensen. "Percutaneous MR-guided interventions using an optical Moiré Phase tracking system: Initial results". In: *PLOS ONE* 13.10 (2018), pp. 1–12 (cit. on p. 24).

[97] Y. Kammerer, K. Scheiter, and W. Beinhauer. "Looking my way through the menu". In: *Proceedings of ETRA 2008 eye tracking research and applications*. New York: ACM, 2008, p. 213 (cit. on pp. 15, 93).

[98] M. Karam et al. "A taxonomy of gestures in human computer interactions". In: (2005) (cit. on p. 9).

[99] M. Kasparick, M. Schmitz, B. Andersen, M. Rockstroh, S. Franke, S. Schlichting, F. Golatowski, and D. Timmermann. "OR.NET: a service-oriented architecture for safe and dynamic medical device interoperability". In: *Biomedizinische Technik. Biomedical engineering* 63.1 (2018), pp. 11–30 (cit. on p. 140).

[100] M. Kassner, W. Patera, and A. Bulling. "Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction". In: *Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp '14 Adjunct. New York, NY, USA: ACM, 2014, pp. 1151–1160 (cit. on p. 95).

[101] D. Katić, P. Spengler, S. Bodenstedt, G. Castrillon-Oberndorfer, R. Seeberger, J. Hoffmann, R. Dillmann, and S. Speidel. "A system for context-aware intraoperative augmented reality in dental implant surgery". In: *International Journal of Computer Assisted Radiology and Surgery* 10.1 (2015), pp. 101–108 (cit. on p. 110).

[102] M. Khamis, C. Oechsner, F. Alt, and A. Bulling. *VRpursuits: interaction in virtual reality using smooth pursuit eye movements*. 2018 (cit. on p. 15).

[103] K. Klamka, A. Siegel, S. Vogt, F. Göbel, S. Stellmach, and R. Dachselt. "Look & Pedal: Hands-free Navigation in Zoomable Information Spaces through Gaze-supported Foot Input". In: *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. 2015, pp. 123–130 (cit. on pp. 64, 90).

[104] O. V. Komogortsev and A. Karpov. “Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades”. In: *Behavior Research Methods* 45.1 (2013), pp. 203–215 (cit. on p. 15).

[105] P. Koutsabasis and P. Vogiatzidakis. “Empirical Research in Mid-Air Interaction: A Systematic Review”. In: *International Journal of Human–Computer Interaction* 35.18 (2019), pp. 1747–1768 (cit. on p. 9).

[106] K. H. Kroemer. “Foot operation of controls”. In: *Ergonomics* 14.3 (1971), pp. 333–361 (cit. on p. 11).

[107] B. Laugwitz, M. Schrepp, and T. Held. “Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten”. In: *Mensch und Computer 2006: Mensch und Computer im Strukturwandel*. München: Oldenbourg Verlag, 2006, pp. 125–134 (cit. on p. 22).

[108] J. Lazar, J. H. Feng, and H. Hochheiser. *Research methods in human-computer interaction*. Morgan Kaufmann, 2017 (cit. on pp. 19, 20).

[109] J. C. Lee. “In search of a natural gesture”. In: *XRDS: Crossroads, The ACM Magazine for Students* 16.4 (2010), p. 9 (cit. on p. 17).

[110] M. Sousa, A. Techmer, A. Steinhage, C. Lauterbach, and P. Lukowicz. “Human tracking and identification using a sensitive floor and wearable accelerometers”. In: *2013 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 2013, pp. 166–171 (cit. on p. 56).

[111] P. Majaranta, U.-K. Ahola, and O. Špakov. “Fast gaze typing with an adjustable dwell time”. In: *The 27th annual CHI conference on Human Factors in Computing Systems*. New York, N.Y.: Association for Computing Machinery, 2009, p. 357 (cit. on p. 15).

[112] P. Majaranta and K.-J. Räihä. “Text entry by gaze: Utilizing eye-tracking”. In: *Text entry systems: Mobility, accessibility, universality* (2007), pp. 175–187 (cit. on p. 93).

[113] P. Majaranta and K.-J. Räihä. “Twenty years of eye typing”. In: *ETRA 2002*. New York: ACM, 2002, p. 15 (cit. on pp. 15, 93).

[114] M. Malaterre. *GDCM Reference Manual*. 2008. (Visited on 03/06/2020) (cit. on p. 31).

[115] N. J. Mansfield and J. M. Marshall. “Symptoms of musculoskeletal disorders in stage rally drivers and co-drivers”. In: *British journal of sports medicine* 35.5 (2001), pp. 314–320 (cit. on pp. 115, 186).

[116] G. Mark, S. T. Iqbal, M. Czerwinski, P. Johns, and A. Sano. “Neurotics can’t focus: An in situ study of online multitasking in the workplace”. In: *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2016, pp. 1739–1744 (cit. on p. 109).

[117] D. J. C. Matthies, T. Roumen, A. Kuijper, and B. Urban. “CapSoles: who is walking on what kind of floor?” In: *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 2017, p. 9 (cit. on p. 57).

[118] M. Mauderer, F. Daiber, and A. Krueger. “Combining Touch and Gaze for Distant Selection in a Tabletop Setting”. In: (2013) (cit. on pp. 93, 106).

[119] S. Mauser and O. Burgert. “Touch-free, gesture-based control of medical devices and software based on the leap motion controller”. In: *MMVR*. 2014, pp. 265–270 (cit. on p. 78).

[120] M. T. Mazilu, A. Z. Faranesh, J. A. Derbyshire, R. J. Lederman, and M. S. Hansen. “Low-cost MRI compatible interface device for interactive scan plane control”. In: *Proceedings of the 19th Annual Meeting of ISMRM, Montreal, Canada*. 2011, p. 3752 (cit. on p. 24).

[121] M. McJury and F. G. Shellock. “Auditory Noise Associated With MR Procedures: A Review”. In: *Journal of Magnetic Resonance Imaging* 12.1 (2000), pp. 37–45 (cit. on p. 23).

[122] R. D. McKendrick and E. Cherry. “A Deeper Look at the NASA TLX and Where It Falls Short”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 62. 2018, pp. 44–48 (cit. on p. 21).

[123] D. McNeill. *Hand and mind: What gestures reveal about thought*. University of Chicago press, 1992 (cit. on pp. 9, 10).

[124] *meCUE 2.0 Website*. URL: `http://mecue.de/english/index.html` (visited on 03/06/2020) (cit. on p. 22).

[125] H. M. Mentis, K. O’Hara, G. Gonzalez, A. Sellen, R. Corish, A. Criminisi, R. Trivedi, and P. Theodore. “Voice or Gesture in the Operating Room”. In: *the 33rd Annual ACM Conference Extended Abstracts*. 2015, pp. 773–780 (cit. on pp. 16, 75, 78, 79).

[126] H. M. Mentis, K. O’Hara, A. Sellen, and R. Trivedi. “Interaction proxemics and image use in neurosurgery”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2012, pp. 927–936 (cit. on p. 6).

[127] A. Mewes, P. Saalfeld, O. Riabikin, M. Skalej, and C. Hansen. “A Gesture-controlled Projection Display for CT-guided Interventions”. In: *International Journal of Computer Assisted Radiology and Surgery* 11.1 (2016), pp. 157–164 (cit. on pp. 9, 10, 17, 28, 39).

[128] A. Mewes, B. Hensen, F. Wacker, and C. Hansen. “Touchless interaction with software in interventional radiology and surgery: a systematic literature review”. In: *International Journal of Computer Assisted Radiology and Surgery* 12.2 (2017), pp. 291–305 (cit. on pp. 1, 8, 9, 10, 18).

[129] B. Meyers, A. J. Brush, S. Drucker, M. A. Smith, and M. Czerwinski. “Dance your work away: exploring step user interfaces”. In: *CHI’06 extended abstracts on Human factors in computing systems*. 2006, pp. 387–392 (cit. on p. 92).

[130] J. Miehle, D. Ostler, N. Gerstenlauer, and W. Minker. “The next step: Intelligent digital assistance for clinical operating rooms”. In: *Innovative Surgical Sciences* 2.3 (2017), p. 33 (cit. on p. 75).

[131] E. Miller, Z. Li, H. Mentis, A. Park, T. Zhu, and N. Banerjee. “RadSense: Enabling one hand and no hands interaction for sterile manipulation of medical images using Doppler radar”. In: *Smart Health* 15 (2019) (cit. on p. 1).

[132] M. Minge, M. Thüring, I. Wagner, and C. V. Kuhr. “The meCUE Questionnaire: A Modular Tool for Measuring User Experience”. In: *Advances in Ergonomics Modeling, Usability & Special Populations*. Springer, 2017, pp. 115–128 (cit. on pp. 22, 51, 68, 82).

[133] S. Mitra and T. Acharya. “Gesture recognition: A survey”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37.3 (2007), pp. 311–324 (cit. on p. 8).

[134] C. A. Monk, J. G. Trafton, and D. A. Boehm-Davis. “The effect of interruption duration and demand on resuming suspended goals”. In: *Journal of Experimental Psychology: Applied* 14.4 (2008), p. 299 (cit. on p. 109).

[135] E. Moser, A. Stadlbauer, C. Windischberger, H. H. Quick, and M. E. Ladd. “Magnetic resonance imaging methodology”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 36.1 (2009), pp. 30–41 (cit. on p. 25).

[136] F. Müller, J. McManus, S. Günther, M. Schmitz, M. Mühlhäuser, and M. Funk. “Mind the Tap: Assessing Foot-Taps for Interacting with Head-Mounted Displays”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI ’19. New York, NY, USA: ACM, 2019, 477:1–477:13 (cit. on p. 13).

[137] V. Müller, M. Fritzsche, and N. Elkmann. “Sensor design and calibration of piezoresistive composite material”. In: *SENSORS, 2015 IEEE*. 2015, pp. 1–4 (cit. on pp. 43, 96).

[138] G. P. Mylonas, K.-W. Kwok, A. Darzi, and G.-Z. Yang. “Gaze-Contingent Motor Channelling and Haptic Constraints for Minimally Invasive Robotic Surgery”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 676–683 (cit. on p. 15).

[139] P. G. Nestor and R. K. Schutt. *Research methods in psychology: Investigating human behavior*. Sage Publications, 2018 (cit. on p. 19).

[140] J. Nielsen. “Noncommand User Interface”. In: *Communication of the ACM* 36.4 (1993), pp. 83–100 (cit. on p. 121).

[141] K. O’Hara, G. Gonzalez, G. Penney, A. Sellen, R. Corish, H. Mentis, A. Varnavas, A. Criminisi, M. Rouncefield, N. Dastur, and T. Carrell. “Interactional Order and Constructed Ways of Seeing with Touchless Imaging Systems in Surgery”. In: *Computer Supported Cooperative Work (CSCW)* 23.3 (2014), pp. 299–337 (cit. on pp. 6, 7, 9, 16, 75, 78, 79).

[142] K. O’Hara, G. Gonzalez, A. Sellen, G. Penney, A. Varnavas, H. Mentis, A. Criminisi, R. Corish, M. Rouncefield, N. Dastur, et al. “Touchless interaction in surgery”. In: *Communications of the ACM* 57.1 (2014), pp. 70–77 (cit. on pp. 1, 6, 10, 63, 78, 79).

[143] J. Okamoto, K. Masamune, H. Iseki, and Y. Muragaki. “Development concepts of a Smart Cyber Operating Theater (SCOT) using ORiN technology”. In: *Biomedizinische Technik. Biomedical engineering* 63.1 (2018), pp. 31–37 (cit. on p. 140).

[144] *Ways of Knowing in HCI*. New York, NY: Springer New York, 2014 (cit. on p. 18).

[145] A. Opromolla, V. Volpi, A. Ingrosso, S. Fabri, C. Rapuano, D. Passalacqua, and C. M. Medaglia. “A Usability Study of a Gesture Recognition System Applied During the Surgical Procedures”. In: *Design, User Experience, and Usability: Interactive Experience Design*. Cham: Springer International Publishing, 2015, pp. 682–692 (cit. on pp. 9, 10, 17).

[146] S. Oviatt. “Multimodal interfaces”. In: *The human-computer interaction handbook*. CRC press, 2007, pp. 439–458 (cit. on pp. 17, 121, 125).

[147] V. Paelke, C. Reimann, and D. Stichling. “Foot-based mobile interaction with games”. In: *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*. [Place of publication not identified]: ACM, 2004, pp. 321–324 (cit. on p. 11).

[148] T. Pakkanen and R. Raisamo. “Appropriateness of foot interaction for non-accurate spatial tasks”. In: *CHI’04 extended abstracts on Human factors in computing systems*. 2004, pp. 1123–1126 (cit. on p. 12).

[149] A. V. Pandey, K. Rall, M. L. Satterthwaite, O. Nov, and E. Bertini. “How Deceptive are Deceptive Visualizations?” In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. New York, NY: ACM, 2015, pp. 1469–1478 (cit. on p. 21).

[150] E. Pannicke, B. Hatscher, B. Hensen, A. Mewes, C. Hansen, F. Wacker, and R. Vick. “MR Compatible and Sterile Gesture Interaction for Interventions”. In: *Proceedings of the 12th Interventional MRI symposium*. Boston, MA, USA, 2018, p. 43 (cit. on pp. 30, 40).

[151] G. Pearson and M. Weiser. “Of moles and men: the design of foot controls for workstations”. In: *ACM SIGCHI Bulletin*. Vol. 17. 1986, pp. 333–339 (cit. on pp. 10, 11).

[152] I. Pernek and A. Ferscha. “A survey of context recognition in surgery”. In: *Medical & Biological Engineering & Computing* 55.10 (2017), pp. 1719–1734 (cit. on p. 140).

[153] A. Perrakis, W. Hohenberger, and T. Horbach. “Integrated operation systems and voice recognition in minimally invasive surgery: comparison of two systems”. In: *Surgical Endoscopy* 27.2 (2013), pp. 575–579 (cit. on pp. 6, 16).

[154] K. Pfeuffer, J. Alexander, M. K. Chong, Y. Zhang, and H. Gellersen. “Gaze-Shifting”. In: *the 28th Annual ACM Symposium*. 2015, pp. 373–383 (cit. on p. 106).

[155] K. Pfeuffer, M. Vidal, J. Turner, A. Bulling, and H. Gellersen. “Pursuit calibration: making gaze calibration less tedious and more flexible”. In: *Proceedings of the 26th annual ACM symposium on User interface software and technology*. 2013, pp. 261–270 (cit. on p. 106).

[156] M. Pirlich, M. Stöhr, T. Neumuth, and A. Dietz. "The Intelligent ENT Operating Room of the Future". In: *Laryngo- rhino- otologie* 98.S 01 (2019), S5–S31 (cit. on p. 140).

[157] E. D. Pisano, J. Chandramouli, B. M. Hemminger, D. Glueck, R. E. Johnston, K. Muller, M. P. Braeuning, D. Puff, W. Garrett, and S. Pizer. "The effect of intensity windowing on the detection of simulated masses embedded in dense portions of digitized mammograms in a laboratory setting". In: *Journal of digital imaging* 10.4 (1997), p. 174 (cit. on p. 8).

[158] *PocketSphinx 5prealpha Website*. URL: `https://github.com/cmusphinx/pocketsphinx` (visited on 05/02/2018) (cit. on p. 81).

[159] A. Poole and L. J. Ball. *Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects. Encyclopedia of Human-Computer Interaction (2005), 211–219.* 2005 (cit. on p. 15).

[160] B. Preim and R. Dachselt. *Interaktive Systeme: Band 2: User Interface Engineering, 3D-Interaktion, Natural User Interfaces*. Springer-Verlag, 2015 (cit. on pp. 9, 10).

[161] F. Quek, D. McNeill, R. Bryll, S. Duncan, X.-F. Ma, C. Kirbas, K. E. McCullough, and R. Ansari. "Multimodal human discourse: gesture and speech". In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 9.3 (2002), pp. 171–193 (cit. on p. 9).

[162] A. V. Reinschluessel, J. Teuber, M. Herrlich, J. Bissel, M. van Eikeren, J. Ganser, F. Koeller, F. Kollasch, T. Mildner, L. Raimondo, L. Reisig, M. Ruedel, D. Thieme, T. Vahl, G. Zachmann, and R. Malaka. "Virtual Reality for User-Centered Design and Evaluation of Touch-free Interaction Techniques for Navigating Medical Images in the Operating Room". In: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. CHI EA '17. New York, NY, USA: ACM, 2017, pp. 2001–2009 (cit. on pp. 11, 72, 78, 79, 139).

[163] M. J. Riffe, S. R. Yutzy, Y. Jiang, M. D. Twieg, C. J. Blumenthal, D. P. Hsu, L. Pan, W. D. Gilson, J. L. Sunshine, C. A. Flask, J. L. Duerk, D. Nakamoto, V. Gulani, and M. A. Griswold. "Device localization and dynamic scan plane selection using a wireless magnetic resonance imaging detector array". In: *Magnetic resonance in medicine* 71.6 (2014), pp. 2243–2249 (cit. on p. 24).

[164] F. Ritter, T. Boskamp, A. Homeyer, H. Laue, M. Schwier, F. Link, and H.-O. Peitgen. "Medical image analysis". In: *IEEE pulse* 2.6 (2011), pp. 60–70 (cit. on pp. 66, 97).

[165] A. Roaas and G. B. J. Andersson. "Normal range of motion of the hip, knee and ankle joints in male subjects, 30–40 years of age". In: *Acta Orthopaedica Scandinavica* 53.2 (1982), pp. 205–208 (cit. on p. 11).

[166] D. A. Robinson. "The mechanics of human smooth pursuit eye movement". In: *The Journal of Physiology* 180.3 (1965), pp. 569–591 (cit. on p. 15).

[167] G. M. Rosa and M. L. Elizondo. "Use of a gesture user interface as a touchless image navigation system in dental surgery: Case series report". In: *Imaging science in dentistry* 44.2 (2014), pp. 155–160 (cit. on p. 9).

[168] E. Rothgang, W. D. Gilson, F. Wacker, J. Hornegger, C. H. Lorenz, and C. R. Weiss. "Rapid freehand MR-guided percutaneous needle interventions: an image-based approach to improve workflow and feasibility". In: *Journal of magnetic resonance imaging : JMRI* 37.5 (2013), pp. 1202–1212 (cit. on pp. 6, 23, 24, 26).

[169] M. A. Rube, A. B. Holbrook, B. F. Cox, R. Buciuc, and A. Melzer. "Wireless mobile technology to improve workflow and feasibility of MR-guided percutaneous interventions". In: *International Journal of Computer Assisted Radiology and Surgery* 10.5 (2015), pp. 665–676 (cit. on pp. 23, 24, 27).

[170] W. A. Rutala, M. S. White, M. F. Gergen, and D. J. Weber. "Bacterial contamination of keyboards: efficacy and functional impact of disinfectants". In: *Infection Control & Hospital Epidemiology* 27.04 (2006), pp. 372–377 (cit. on pp. 1, 6).

[171] P. Saalfeld, D. Kasper, B. Preim, and C. Hansen. "Touchless Measurement of Medical Image Data for Interventional Support". In: *Mensch und Computer 2017 - Tagungsband*. Regensburg: Gesellschaft für Informatik e.V, 2017, pp. 83–92 (cit. on p. 9).

[172] D. D. Salvucci and P. Bogunovich. "Multitasking and monotasking: the effects of mental workload on deferred task interruptions". In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2010, pp. 85–88 (cit. on p. 109).

[173] D. D. Salvucci, N. A. Taatgen, and J. P. Borst. “Toward a Unified Theory of the Multitasking Continuum: From Concurrent Performance to Task Switching, Interruption, and Resumption”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’09. New York, NY, USA: ACM, 2009, pp. 1819–1828 (cit. on p. 109).

[174] N. Sangsuriyachot and M. Sugimoto. “Novel interaction techniques based on a combination of hand and foot gestures in tabletop environments”. In: *Proceedings of the 10th asia pacific conference on Computer human interaction*. 2012, pp. 21–28 (cit. on pp. 50, 57, 64).

[175] W. Saunders and D. Vogel. “Tap-Kick-Click: Foot Interaction for a Standing Desk”. In: *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. 2016, pp. 323–333 (cit. on pp. 73, 78, 106).

[176] W. Saunders and D. Vogel. “The performance of indirect foot pointing using discrete taps and kicks while standing”. In: *Proceedings of the 41st Graphics Interface Conference*. 2015, pp. 265–272 (cit. on p. 103).

[177] J. Sauro. *A practical guide to the system usability scale: Background, benchmarks & best practices*. Measuring Usability LLC Denver, CO, 2011 (cit. on p. 95).

[178] J. Sauro and J. R. Lewis. *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann, 2016 (cit. on p. 21).

[179] D. Schott. “Entwicklung von multimodalen Interaktionstechniken zur handfreien Selektion und Manipulation von medizinischen Bilddaten”. Master’s Thesis. Breitscheidstr. 2, 39114 Magdeburg, Germany: Magdeburg-Stendal University of Applied Sciences (cit. on p. 135).

[180] D. Schott, B. Hatscher, F. Joeres, M. Gabele, S. Hußlein, and C. Hansen. “Lean-Interaction: passive image manipulation in concurrent multitasking”. In: *Proceedings of Graphics Interface 2020*. GI 2020. Canadian Human-Computer Communications Society / Société canadienne du dialogue humain-machine, 2020, pp. 404–412 (cit. on pp. 121, 122, 127, 128, 131).

[181] W. J. Schroeder, K. W. Martin, W. E. Lorensen, and L. S. Avila. *The visualization toolkit: An object-oriented approach to 3D graphics ; [visualize data in 3D - medical, engineering or scientific ; build your own applications with C++, Tcl, Java or Python ; includes source code for VTK (supports UNIX, Windows and Mac)*. 4. ed. Clifton Park, NY: Kitware, 2006 (cit. on p. 31).

[182] M. O. Schurr, G. F. Buess, F. Wieth, H.-J. Saile, and M. Botsch. "Ergonomic surgeon's chair for use during minimally invasive surgery". In: *Surgical Laparoscopy Endoscopy & Percutaneous Techniques* 9.4 (1999), p. 244 (cit. on p. 13).

[183] J. Scott, D. Dearman, K. Yatani, and K. N. Truong. "Sensing foot gestures from the pocket". In: *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. 2010, pp. 199–208 (cit. on pp. 11, 12, 13, 64).

[184] K. Seals, R. Al-Hakim, P. Mulligan, E. Lehrman, N. Fidelman, K. Kolli, R. Kohlbrenner, M. Kohi, and A. Taylor. "03:45 PM Abstract No. 38 The development of a machine learning smart speaker application for device sizing in interventional radiology". In: *Journal of Vascular and Interventional Radiology* 30.3 (2019), S20 (cit. on p. 16).

[185] S. Shimizu, K. Kondo, T. Yamazaki, H. Koizumi, T. Miyazaki, S. Osawa, T. Sagiuchi, K. Nakayama, I. Yamamoto, and K. Fujii. "Hanging Foot Switch for Bipolar Forceps: A Device for Surgeons Operating in the Standing Position". In: *Neurologia medico-chirurgica* 53.1 (2013), pp. 53–55 (cit. on p. 11).

[186] L. E. Sibert and R. J. K. Jacob. "Evaluation of eye gaze interaction". In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 2000, pp. 281–288 (cit. on pp. 13, 91).

[187] J. Sigut and S.-A. Sidha. "Iris center corneal reflection method for gaze tracking using visible light". In: *IEEE Transactions on Biomedical Engineering* 58.2 (2010), pp. 411–419 (cit. on p. 13).

[188] A. L. Simeone, E. Velloso, J. Alexander, and H. Gellersen. "Feet movement in desktop 3D interaction". In: *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. 2014, pp. 71–74 (cit. on pp. 13, 71).

[189] S. B. Solomon and S. G. Silverman. "Imaging in interventional oncology". In: *Radiology* 257.3 (2010), pp. 624–640 (cit. on p. 6).

[190] A. Solovjova. "Entwicklung und Evaluierung multimodaler berührungsloser AR-Interaktionstechniken für Multitasking-Szenarien". Master's Thesis. Universitätsplatz 2, Magdeburg, Germany: Otto von Guericke University Magdeburg, 2019 (cit. on p. 135).

[191] A. Solovjova, B. Hatscher, and C. Hansen. "Influence of augmented reality interaction on a primary task for the medical domain". In: *Mensch und Computer 2019 - Workshopband*. Bonn: Gesellschaft für Informatik e.V, 2019, pp. 325–330 (cit. on pp. 113, 115, 116).

[192] A. Solovjova, D. Labsch, B. Hatscher, M. Fritzsche, and C. Hansen. “Plantar pressure-based gestures for medical image manipulation”. In: *Mensch und Computer 2018 - Tagungsband*. Bonn: Gesellschaft für Informatik e.V, 2018, pp. 421–425 (cit. on pp. 11, 58).

[193] J. Stattaus, S. Maderwald, M. Forsting, J. Barkhausen, and M. E. Ladd. “MR-guided core biopsy with MR fluoroscopy using a short, wide-bore 1.5-Tesla scanner: Feasibility and initial results”. In: *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 27.5 (2008), pp. 1181–1187 (cit. on p. 24).

[194] A. Steinhage and C. Lauterbach. “Monitoring Movement Behavior by Means of a Large Area Proximity Sensor Array in the Floor”. In: vol. 396. 2008, pp. 15–27 (cit. on p. 48).

[195] S. Stellmach and R. Dachselt. “Still looking: investigating seamless gaze-supported selection, positioning, and manipulation of distant targets”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2013, pp. 285–294 (cit. on pp. 13, 92).

[196] S. Stellmach, M. Jüttner, C. Nywelt, J. Schneider, and R. Dachselt. “Investigating Freehand Pan and Zoom”. In: *Mensch & Computer 2012: interaktiv informiert – allgegenwärtig und allumfassend!?* München: Oldenbourg Verlag, 2012, pp. 303–312 (cit. on p. 79).

[197] D. Stoyanov, G. P. Mylonas, and G.-Z. Yang. “Gaze-Contingent 3D Control for Focused Energy Ablation in Robotic Assisted Surgery”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 347–355 (cit. on p. 15).

[198] P. Suetens. *Fundamentals of medical imaging*. Cambridge university press, 2017 (cit. on p. 122).

[199] K. G. Tippey, E. Sivaraj, and T. K. Ferris. “Driving While Interacting With Google Glass: Investigating the Combined Effect of Head-Up Display and Hands-Free Input on Driving Safety and Multitask Performance”. In: *Human factors* 59.4 (2017), pp. 671–688 (cit. on p. 110).

[200] J. G. Trafton, E. M. Altmann, D. P. Brock, and F. E. Mintz. “Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal”. In: *International Journal of Human-Computer Studies* 58.5 (2003), pp. 583–603 (cit. on p. 109).

[201] D. Tsetis, R. Uberoi, F. Fanelli, I. Roberston, M. Krokidis, O. van Delden, B. Radeleff, S. Müller-Hülsbeck, M. Szerbo-Trojanowska, M. Lee, et al. "The provision of interventional radiology services in Europe: CIRSE recommendations". In: *CardioVascular and Interventional Radiology* 39.4 (2016), pp. 500–506 (cit. on pp. 5, 6).

[202] J. Turner, J. Alexander, A. Bulling, D. Schmidt, and H. Gellersen. "Eye Pull, Eye Push: Moving Objects between Large Screens and Personal Devices with Gaze and Touch". In: *Human-Computer Interaction – INTERACT 2013*. Vol. 8118. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 170–186 (cit. on p. 106).

[203] R. N. Uppot, B. Laguna, C. J. McCarthy, G. de Novi, A. Phelps, E. Siegel, and J. Courtier. "Implementing virtual and augmented reality tools for radiology education and training, communication, and clinical care". In: *Radiology* 291.3 (2019), pp. 570–580 (cit. on p. 110).

[204] M. A. van Veelen, C. J. Snijders, E. van Leeuwen, R. H. M. Goossens, and G. Kazemier. "Improvement of foot pedals used during surgery based on new ergonomic guidelines". In: *Surgical Endoscopy And Other Interventional Techniques* 17.7 (2003), pp. 1086–1091 (cit. on pp. 13, 41, 42, 63, 72, 73).

[205] P. Vávra, J. Roman, P. Zonča, P. Ihnát, M. Němec, J. Kumar, N. Habib, and A. El-Gendi. "Recent development of augmented reality in surgery: a review". In: *Journal of healthcare engineering* 2017 (2017) (cit. on p. 110).

[206] F. J. Veith and M. L. Marin. "Endovascular technology and its impact on the relationships among vascular surgeons, interventional radiologists, and other specialists". In: *World journal of surgery* 20.6 (1996), pp. 687–691 (cit. on pp. 5, 6).

[207] E. Velloso, J. Alexander, A. Bulling, and H. Gellersen. "Interactions Under the Desk: A Characterisation of Foot Movements for Input in a Seated Position". In: *Human-Computer Interaction – INTERACT 2015*. Vol. 9296. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 384–401 (cit. on pp. 13, 137).

[208] E. Velloso, D. Schmidt, J. Alexander, H. Gellersen, and A. Bulling. "The Feet in Human–Computer Interaction: A Survey of Foot-Based Interaction". In: *ACM Computing Surveys (CSUR)* 48.2 (2015), p. 21 (cit. on pp. 11, 12, 62).

[209] E. Velloso, J. Turner, J. Alexander, A. Bulling, and H. Gellersen. "An empirical investigation of gaze selection in mid-air gestural 3D manipulation". In: *Human-Computer Interaction – INTERACT 2015*. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 315–330 (cit. on pp. 105, 106).

[210] M. Vidal, A. Bulling, and H. Gellersen. "Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets". In: *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp '13. New York, NY, USA: ACM, 2013, pp. 439–448 (cit. on pp. 15, 106).

[211] H. Visarius, J. Gong, C. Scheer, S. Haralamb, and L. P. Nolte. "Man-machine interfaces in computer assisted surgery". In: *Computer Aided Surgery: Official Journal of the International Society for Computer Aided Surgery (ISCAS)* 2.2 (1997), pp. 102–107 (cit. on p. 6).

[212] M. Visentini-Scarzanella, G. P. Mylonas, D. Stoyanov, and G.-Z. Yang. "i-BRUSH: A Gaze-Contingent Virtual Paintbrush for Dense 3D Reconstruction in Robotic Assisted Surgery". In: *Medical Image Computing and Computer-Assisted Intervention MICCAI 2009 00*. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 353–360 (cit. on p. 15).

[213] J. P. Wachs, H. I. Stern, Y. Edan, M. Gillam, J. Handler, C. Feied, and M. Smith. "A Gesture-based Tool for Sterile Browsing of Radiology Images". In: *Journal of the American Medical Informatics Association* 15.3 (2008), pp. 321–323 (cit. on p. 7).

[214] J. P. Wachs, M. Kölsch, H. Stern, and Y. Edan. "Vision-based hand-gesture applications". In: *Communications of the ACM* 54.2 (2011), pp. 60–71 (cit. on p. 10).

[215] S. Wagner. "Interaktion mit medizinischen Bilddaten unter sterilen Bedingungen mit Hilfe von Fußgesten". Master's Thesis. Universitätsplatz 2, Magdeburg, Germany: Otto von Guericke University Magdeburg, 2017 (cit. on p. 74).

[216] S. Wagner, B. Hatscher, M. Luz, B. Preim, and C. Hansen. "Konzepte mit unterschiedlichen Platzanforderungen zur Navigation in medizinischen Bilddaten mittels eines Sensorfußbodens". In: *Proceedings of the Annual Meeting of the German Society of Computer- and Robot-Assisted Surgery*. Hannover, Germany, 2017, pp. 220–225 (cit. on pp. 49, 50, 51, 52).

[217] Q. Wang, G. Kurillo, F. Ofli, and R. Bajcsy. “Evaluation of Pose Tracking Accuracy in the First and Second Generations of Microsoft Kinect”. In: *2015 IEEE International Conference on Healthcare Informatics*. Piscataway, NJ: IEEE, 2015, pp. 380–389 (cit. on p. 10).

[218] C. Ware and H. H. Mikaelian. “An evaluation of an eye tracker as a device for computer input2”. In: *ACM SIGCHI Bulletin*. Vol. 17. 1987, pp. 183–188 (cit. on pp. 13, 93).

[219] L. S. G. L. Wauben, M. A. van Veelen, D. Gossot, and R. H. M. Goossens. “Application of ergonomic guidelines during minimally invasive surgery: a questionnaire survey of 284 surgeons”. In: *Surgical Endoscopy And Other Interventional Techniques* 20.8 (2006), pp. 1268–1274 (cit. on pp. 13, 62).

[220] *Website of the MD PnP Medical Device Interoperability Program*. URL: `http://mdpnp.mgh.harvard.edu/` (visited on 03/02/2020) (cit. on p. 140).

[221] *Website of the OR.NET e.V.* URL: `http://ornet.org` (visited on 03/02/2020) (cit. on p. 140).

[222] J. Weiss, R. Hoffmann, H. Rempp, D.-E. Keßler, P. L. Pereira, K. Nikolaou, and S. Clasen. “Feasibility, efficacy, and safety of percutaneous MR-guided ablation of small (≤12 mm) hepatic malignancies”. In: *Journal of magnetic resonance imaging : JMRI* 49.2 (2019), pp. 374–381 (cit. on pp. 23, 26).

[223] C. D. Wickens. “Multiple resources and performance prediction”. In: *Theoretical issues in ergonomics science* 3.2 (2002), pp. 159–177 (cit. on pp. 109, 110, 120, 121, 124).

[224] C. D. Wickens, J. Goh, J. Helleberg, W. J. Horrey, and D. A. Talleur. “Attentional models of multitask pilot performance using advanced display technology”. In: *Human factors* 45.3 (2003), pp. 360–380 (cit. on p. 110).

[225] D. Wigdor and D. Wixon. *Brave NUI world: designing natural user interfaces for touch and gesture*. Elsevier, 2011 (cit. on p. 17).

[226] J. C. F. de Winter. “Controversy in human factors constructs and the explosive use of the NASA-TLX: a measurement perspective”. In: *Cognition, technology & work* 16.3 (2014), pp. 289–297 (cit. on pp. 20, 21).

[227] R. Wipfli, V. Dubois-Ferrière, S. Budry, P. Hoffmeyer, and C. Lovis. “Gesture-Controlled Image Management for Operating Room: A Randomized Crossover Study to Compare Interaction Using Gestures, Mouse, and Third Person Relaying”. In: *PLOS ONE* 11.4 (2016), e0153596 (cit. on p. 38).

[228] M. Wirth, S. Gradl, J. Sembdner, S. Kuhrt, and B. M. Eskofier. “Evaluation of Interaction Techniques for a Virtual Reality Reading Room in Diagnostic Radiology”. In: *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. UIST ’18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 867–876 (cit. on pp. 8, 17).

[229] J. O. Wobbrock, J. Rubinstein, M. W. Sawyer, and A. T. Duchowski. “Longitudinal evaluation of discrete consecutive gaze gestures for text entry”. In: *Proceedings of the 2008 symposium on Eye tracking research & applications*. 2008, pp. 11–18 (cit. on p. 15).

[230] A. Zaman, L. Reisig, A. V. Reinschluessel, H. Bektas, D. Weyhe, M. Herrlich, T. Döring, and R. Malaka. “An Interactive-Shoe For Surgeons: Hand-Free Interaction With Medical 2D Data”. In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI EA ’18. New York, NY, USA: ACM, 2018, LBW633:1–LBW633:6 (cit. on pp. 6, 78, 79).

[231] S. Zhai, C. Morimoto, and S. Ihde. “Manual and gaze input cascaded (MAGIC) pointing”. In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 1999, pp. 246–253 (cit. on pp. 13, 15, 91).

[232] K. Zhong, F. Tian, and H. Wang. “Foot menu: Using heel rotation information for menu selection”. In: *2011 15th Annual International Symposium on Wearable Computers (ISWC)*. 2011, pp. 115–116 (cit. on pp. 11, 64).

[233] J. Zhu and J. Yang. “Subpixel eye gaze tracking”. In: *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*. 2002, pp. 131–136 (cit. on p. 15).

[234] E. R. Zorzal, J. M. Campos Gomes, M. Sousa, P. Belchior, P. G. da Silva, N. Figueiredo, D. S. Lopes, and J. Jorge. “Laparoscopy with augmented reality adaptations”. In: *Journal of Biomedical Informatics* 107 (2020), p. 103463 (cit. on p. 18).

List of Figures

List of Tables

Appendix

Appendix to Chapter 3

German version of online survey

The online survey was provided in german and english. The english questions and answers are reported in the corresponding section. The german version of the survey questions and answer options is listed below.

User study questionnaires

The demographic questionnaire, the SUS questionnaire and the TLX questionnaire are appended. The TLX was slightly modified by replacing the vertical tick marks with checkboxes to avoid confusion.

Wie viel Erfahrung haben Sie mit Nadelinterventionen (in Jahren)?	
Wie viel Erfahrung haben Sie mit MR-gestützten Interventionen (in Jahren)?	
Wie viele MR-gestützte Nadelinterventionen haben Sie bisher durchgeführt (ca.)?	
Wie bedienen Sie den MR aktuell?	Verbale Delegation an assistierendes Personal Verwendung eines Trackball-Terminals/MR-sicheren Terminals Bedienung der Workstation im Kontrollraum Sonstiges
Wie häufig benötigen Sie folgende Funktionen präoperativ?	Windowing/Fenstern Sequenz wechseln Sequenz starten/stoppen Sequenzparameter ändern Bildebene verschieben Bildebene rotieren Wechsel zwischen paralleler oder orthogonaler Ebenenausrichtung zur Nadel Wechsel zwischen saggittaler, coronaler und axialer Anzeigen des Planungsdatensatz (inklusive Tumor in 3D, Einstichstelle, geplanter Trajektorie, Ziel)
Wie häufig benötigen Sie folgende Funktionen intraoperativ?	Windowing/Fenstern Sequenz wechseln Sequenz starten/stoppen Sequenzparameter ändern Bildebene verschieben Bildebene rotieren Wechsel zwischen paralleler oder orthogonaler Ebenenausrichtung zur Nadel Wechsel zwischen saggittaler, coronaler und axialer Anzeigen des Planungsdatensatz (inklusive Tumor in 3D, Einstichstelle, geplanter Trajektorie, Ziel)
Wie nützlich fänden Sie es, folgende Funktionen intraoperativ zur Verfügung stehen zu haben?	Windowing/Fenstern Sequenz wechseln Sequenz starten/stoppen Sequenzparameter ändern Bildebene verschieben Bildebene rotieren Wechsel zwischen paralleler oder orthogonaler Ebenenausrichtung zur Nadel Wechsel zwischen saggittaler, coronaler und axialer Anzeigen des Planungsdatensatz (inklusive Tumor in 3D, Einstichstelle, geplanter Trajektorie, Ziel)
Welche Bemerkungen und/oder Zusatzwünsche möchten Sie uns zuletzt mitteilen?	

Probanden-ID__________

Modalität __________

Hinweis: Die Auswertung der Studie und des folgenden Fragebogens erfolgt komplett **anonym.** Seien Sie bitte möglichst **objektiv** in Ihrer Beurteilung und beantworten Sie die Fragen möglichst **ehrlich.**

Demografische Angaben

Alter:	…… Jahre	
Geschlecht:	☐ weiblich	☐ männlich
Händigkeit:	☐ Linkshänder	☐ Rechtshänder

Wie viel Erfahrung haben Sie mit Nadelinterventionen (in Jahren)?	…… Jahre
Wie viel Erfahrung haben Sie mit **MRT**-gestützten Interventionen (in Jahren)?	…… Jahre
Wie viel Erfahrung haben Sie mit **CT**-gestützten Interventionen (in Jahren)?	…… Jahre
Wie viele bildgestützte Nadelinterventionen haben Sie bisher durchgeführt (ca.)?	………..

Probanden-ID__________
Modalität __________

System Usability Scale

	Stimme überhaupt nicht zu				Stimme voll zu
1. Ich denke, dass ich das System gerne häufig benutzen würde.	1	2	3	4	5
2. Ich fand das System unnötig komplex.	1	2	3	4	5
3. Ich fand das System einfach zu benutzen.	1	2	3	4	5
4. Ich glaube, ich würde die Hilfe einer technisch versierten Person benötigen, um das System benutzen zu können.	1	2	3	4	5
5. Ich fand, die verschiedenen Funktionen in diesem System waren gut integriert.	1	2	3	4	5
6. Ich denke, das System enthielt zu viele Inkonsistenzen.	1	2	3	4	5
7. Ich kann mir vorstellen, dass die meisten Menschen den Umgang mit diesem System sehr schnell lernen.	1	2	3	4	5
8. Ich fand das System sehr umständlich zu nutzen.	1	2	3	4	5
9. Ich fühlte mich bei der Benutzung des Systems sehr sicher.	1	2	3	4	5
10. Ich musste eine Menge lernen, bevor ich anfangen konnte das System zu verwenden.	1	2	3	4	5

Probanden-ID_________
Modalität _________

NASA-TLX Questionaire

Geistige Anforderung

Wie viel geistige Anforderung war bei der Informationsaufnahme und bei der Informationsverarbeitung erforderlich (z.B. Denken, Entscheiden, Rechnen, Erinnern, Hinsehen, Suchen ...)? War die Aufgabe leicht oder anspruchsvoll, einfach oder komplex, erfordert sie hohe Genauigkeit oder ist sie fehlertolerant?

Gering | | | | | | | | | | | | | | | | | | | | | Hoch

Körperliche Anforderung

Wie viel körperliche Aktivität war erforderlich (z.B. ziehen, drücken, drehen, steuern, aktivieren ...)? War die Aufgabe leicht oder schwer, einfach oder anstrengend, erholsam oder mühselig?

Gering | | | | | | | | | | | | | | | | | | | | | Hoch

Zeitliche Anforderung

Wie viel Zeitdruck empfanden Sie hinsichtlich der Häufigkeit oder dem Takt mit dem die Aufgaben oder Aufgabenelemente auftraten? War die Aufgabe langsam und geruhsam oder schnell und hektisch?

Gering | | | | | | | | | | | | | | | | | | | | | Hoch

Leistung

Wie erfolgreich haben Sie Ihrer Meinung nach die vom Versuchsleiter (oder Ihnen selbst) gesetzten Ziele erreicht? Wie zufrieden waren Sie mit Ihrer Leistung bei der Verfolgung dieser Ziele?

Gut | | | | | | | | | | | | | | | | | | | | | Schlecht

Anstrengung

Wie hart mussten Sie arbeiten, um Ihren Grad an Aufgabenerfüllung zu erreichen?

Gering | | | | | | | | | | | | | | | | | | | | | Hoch

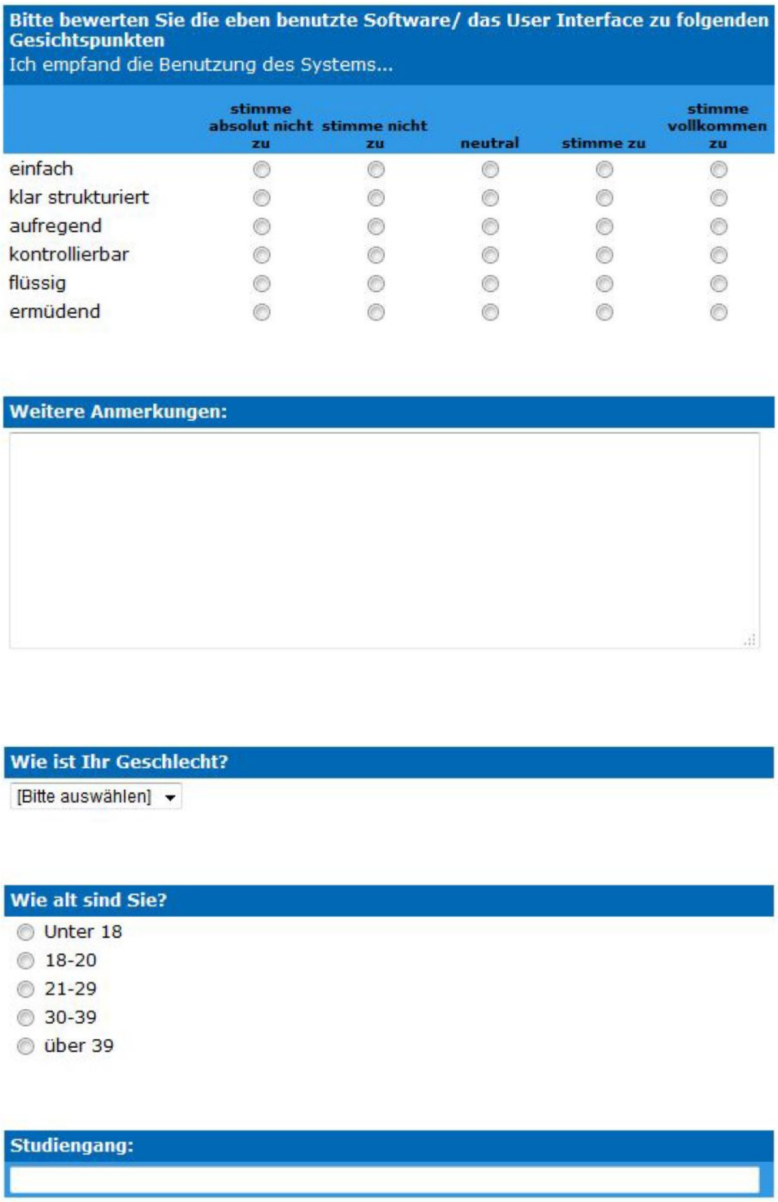
Frustration

Wie unsicher, entmutigt, irritiert, gestresst und verärgert (versus sicher, bestätigt, zufrieden, entspannt und zufrieden mit sich selbst) fühlten Sie sich während der Aufgabe?

Gering | | | | | | | | | | | | | | | | | | | | | Hoch

Appendix to Chapter 4

Questionnaire used in Section 4.1



Bitte bewerten Sie die eben benutzte Software/ das User Interface zu folgenden Gesichtspunkten
Ich empfand die Benutzung des Systems...

	stimme absolut nicht zu	stimme nicht zu	neutral	stimme zu	stimme vollkommen zu
einfach	○	○	○	○	○
klar strukturiert	○	○	○	○	○
aufregend	○	○	○	○	○
kontrollierbar	○	○	○	○	○
flüssig	○	○	○	○	○
ermüdend	○	○	○	○	○

Weitere Anmerkungen:

Wie ist Ihr Geschlecht?
[Bitte auswählen]

Wie alt sind Sie?
- ○ Unter 18
- ○ 18-20
- ○ 21-29
- ○ 30-39
- ○ über 39

Studiengang:

Appendix to Chapter 5

Modified meCUE Modules I and V used in Section 5.1

The modules I and V of the meCUE questionnaire were modified by to match the topic investigated, which was a interaction method instead of a product.

Modified SEQ and question on suitability used in Section 5.2

The questions appended were used to assess ease (SEQ) and the suitability of the presented input method for the presented task.

________ ________

Dieser Fragebogen dient dazu zu erfassen, wie Sie die Interaktionsmethode erleben.

Nachfolgend finden Sie verschiedene Aussagen, die Sie benutzen können, um Ihr Erleben zu bewerten. Bitte geben Sie den Grad Ihrer Zustimmung zu jeder Aussage an, indem Sie das entsprechende Feld ankreuzen.

Entscheiden Sie spontan und ohne langes Nachdenken, um Ihren ersten Eindruck mitzuteilen. Bitte beurteilen Sie jede Aussage, selbst wenn Sie meinen, dass sie nicht vollständig zu Ihrem Erleben passt.

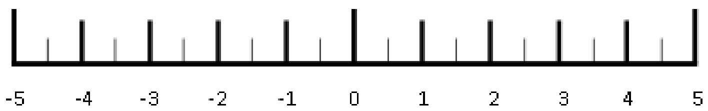
Es gibt keine "richtigen" oder "falschen" Antworten - nur Ihre persönliche Meinung zählt!

Entsperrmethode

	lehne völlig ab	**lehne ab**	**lehne eher ab**	**weder noch**	**stimme eher zu**	**stimme zu**	**stimme völlig zu**
Die Entsperrung lässt sich einfach benutzen.	O	O	O	O	O	O	O
Die Funktionen der Entsperrung sind genau richtig für meine Ziele.	O	O	O	O	O	O	O
Es wird schnell klar, wie man die Entsperrung bedienen muss.	O	O	O	O	O	O	O
Ich halte die Entsperrung für absolut nützlich.	O	O	O	O	O	O	O
Die Bedienung der Entsperrung ist verständlich.	O	O	O	O	O	O	O
Mithilfe der Entsperrung kann ich meine Ziele erreichen.	O	O	O	O	O	O	O

Wie erleben Sie die Entsperrmethode insgesamt?

als schlecht



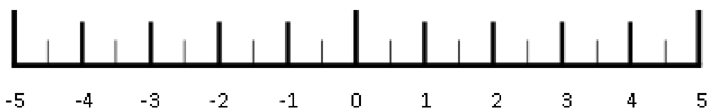
als gut

Interaktionsmethode

	lehne völlig ab	**lehne ab**	**lehne eher ab**	**weder noch**	**stimme eher zu**	**stimme zu**	**stimme völlig zu**
Die Interaktionsmethode lässt sich einfach benutzen.	O	O	O	O	O	O	O
Die Funktionen der Interaktionsmethode sind genau richtig für meine Ziele.	O	O	O	O	O	O	O
Es wird schnell klar, wie man die Interaktionsmethode bedienen muss.	O	O	O	O	O	O	O
Ich halte die Interaktionsmethode für absolut nützlich.	O	O	O	O	O	O	O
Die Bedienung der Interaktionsmethode ist verständlich.	O	O	O	O	O	O	O
Mithilfe der Interaktionsmethode kann ich meine Ziele erreichen.	O	O	O	O	O	O	O

Wie erleben Sie die Interaktionsmethode insgesamt?

als schlecht



als gut

Proband Nr._________

Wie schwierig oder einfach fanden Sie diese Aufgabe?

sehr schwierig ☐ ☐ ☐ ☐ ☐ ☐ ☐ sehr einfach

Wie geeignet fanden Sie die Eingabemethode für diese Aufgabe?
sehr ungeeignet ☐ ☐ ☐ ☐ ☐ ☐ ☐ sehr geeignet

F _________ M ________

Wie schwierig oder einfach fanden Sie diese Aufgabe?

sehr schwierig ☐ ☐ ☐ ☐ ☐ ☐ ☐ sehr einfach

Wie geeignet fanden Sie die Eingabemethode für diese Aufgabe?
sehr ungeeignet ☐ ☐ ☐ ☐ ☐ ☐ ☐ sehr geeignet

F _________ M ________

Wie schwierig oder einfach fanden Sie diese Aufgabe?

sehr schwierig ☐ ☐ ☐ ☐ ☐ ☐ ☐ sehr einfach

Wie geeignet fanden Sie die Eingabemethode für diese Aufgabe?
sehr ungeeignet ☐ ☐ ☐ ☐ ☐ ☐ ☐ sehr geeignet

F _________ M ________

Wie schwierig oder einfach fanden Sie diese Aufgabe?

sehr schwierig ☐ ☐ ☐ ☐ ☐ ☐ ☐ sehr einfach

Wie geeignet fanden Sie die Eingabemethode für diese Aufgabe?
sehr ungeeignet ☐ ☐ ☐ ☐ ☐ ☐ ☐ sehr geeignet

F _________ M ________

Wie schwierig oder einfach fanden Sie diese Aufgabe?

sehr schwierig ☐ ☐ ☐ ☐ ☐ ☐ ☐ sehr einfach

Wie geeignet fanden Sie die Eingabemethode für diese Aufgabe?
sehr ungeeignet ☐ ☐ ☐ ☐ ☐ ☐ ☐ sehr geeignet

F _________ M ________

Appendix to Chapter 6

Demographic Questionnaire and Post-Test Interview Questions used in Section 6.1

Physical Strain Diagram used in Section 6.1

The whole-body diagram to assess physical strain adapted from Mansfield and Marshall [115].

Demographic Questionnaire used in Section 6.2

Data plots for Section 6.2

Only significant measures are reported in Section 6.2. Plots for all data gathered during the user study are appended.

Demographische Daten

Alter:

Geschlecht:

Schuhgröße:

Dominante Hand:

Sehschwäche? ja / nein

Farbsehschwäche? ja / nein

Erfahrung mit AR\VR? ja / nein

Erfahrung mit der HoloLens? ja / nein

Erfahrung mit Sprachinteraktion? ja / nein

Erfahrung mit Gesteninteraktion? ja / nein

Erfahrung mit Fußinteraktion? ja / nein

Fragen des abschließenden Interviews

Wie empfandest du die Aufgabe?

Was gefiel dir gar nicht und wieso?

In welcher Reihenfolge würdest du die Interaktionsmethoden ohne Multitasking-Kontext platzieren?

Was war deine Strategie um den Marker innerhalb des grünen Bereichs zu halten?

In welcher Reihenfolge würdest du die Interaktionsmethoden mit Multitasking-Kontext platzieren?

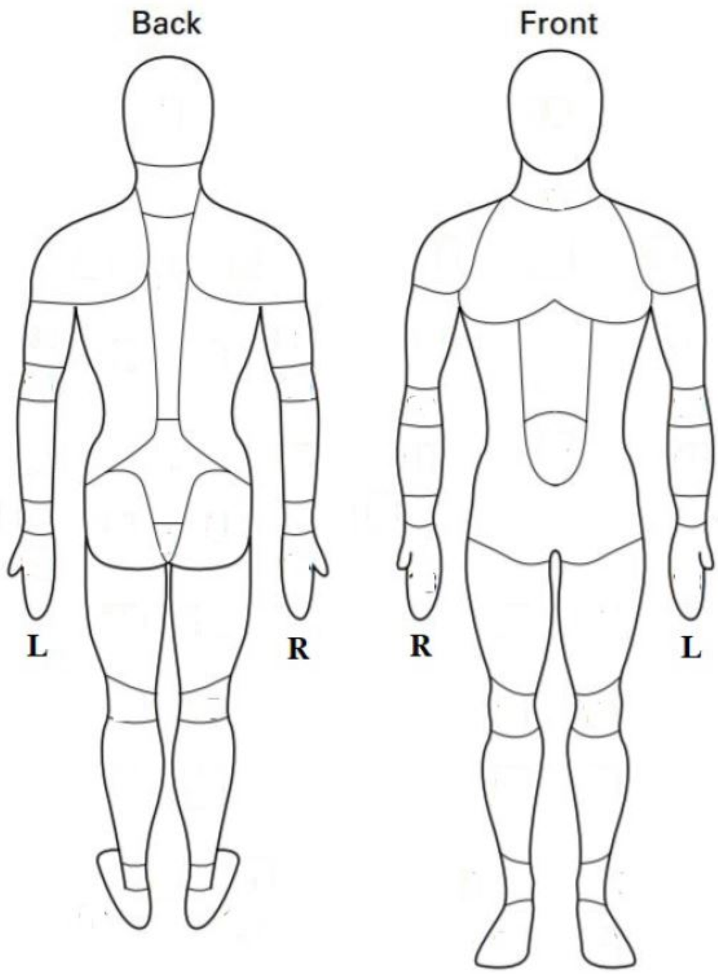


Figure .1: Adapted by permission from BMJ Publishing Group Limited. *Symptoms of musculoskeletal disorders in stage rally drivers and co-drivers*, Mansfield, N. J., Marshall, J. M., 35, 314–320, © BMJ Publishing Group Ltd. 2001

Demografische Angaben

Datum: ________ Uhrzeit: ________ ID: _____

Alter: _____ Jahre

Geschlecht: ○ weiblich ○ männlich ○ divers

Tätigkeitsstatus:
○ Schüler
○ Student, im Studienfach: ____________________
○ erwerbs- bzw. berufstätig als: ____________________
○ Sonstiges: ____________________

Dominante Hand: ○ links ○ rechts

Ist ihr Sehvermögen beinträchigt?

○ nein
○ Fehlsichtigkeit (Allgemeine Sehschwäche)
○ Farbsinnstörung (Bspw. Rot-Grün-Sehschwäche)
○ Sonstiges: ____________________

Liegt eine Sprachstörung vor?

○ nein
○ leicht
○ schwer

Wie gut beherrschen Sie die englische Sprache?

Gar nicht ○ ○ ○ ○ ○ Sehr gut ○ Muttersprache

Welche Erfahrungen haben Sie in folgenden Bereichen?

Fachwissen Mensch-Computer-Interaktion

Keine ○ ○ ○ ○ ○ Sehr erfahren

Gestensteuerung / Berührungslose Interaktion

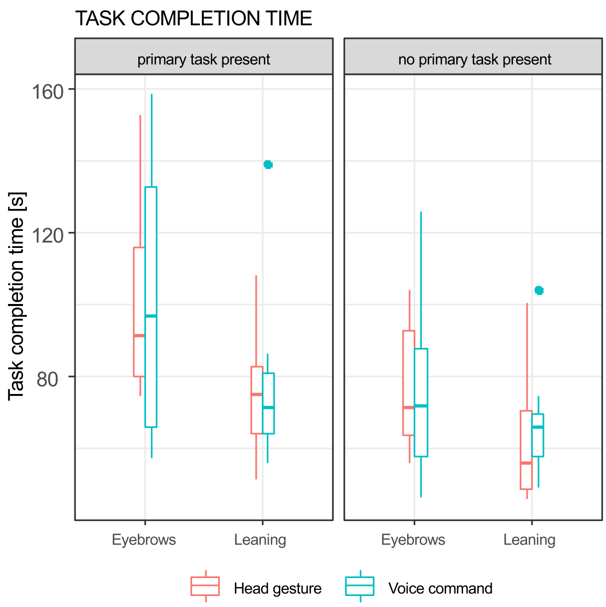
Keine ○ ○ ○ ○ ○ Sehr erfahren

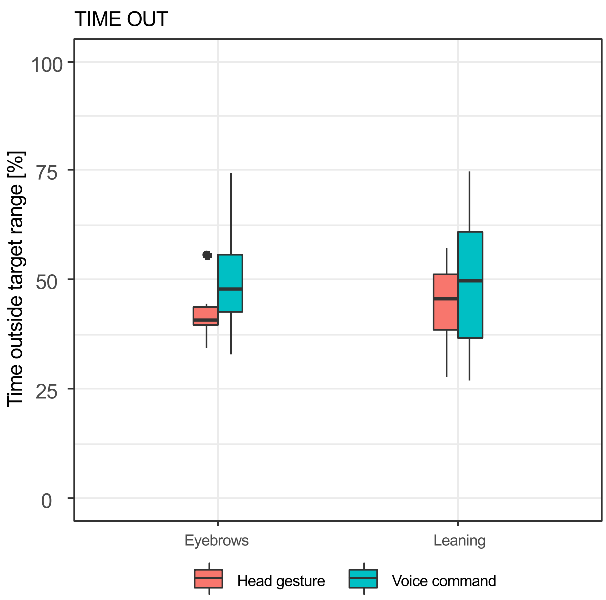
Gesichts,- Emotions oder Eyetracking

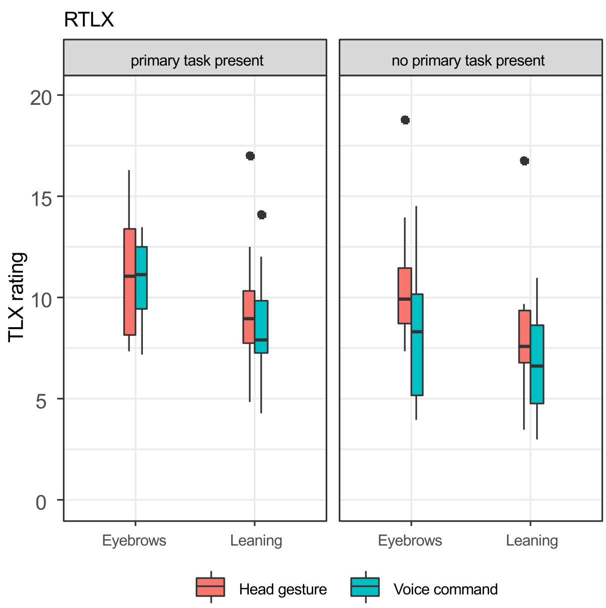
Keine ○ ○ ○ ○ ○ Sehr erfahren

Sprachsteuerung

Keine ○ ○ ○ ○ ○ Sehr erfahren


TASK COMPLETION TIME
primary task present
no primary task present
Task completion time [s]
160
120
80
Eyebrows
Leaning
Eyebrows
Leaning
Head gesture
Voice command


TIME OUT
Time outside target range [%]
100
75
50
25
0
Eyebrows
Leaning
Head gesture
Voice command


RTLX
primary task present
no primary task present
TLX rating
20
15
10
5
0
Eyebrows
Leaning
Eyebrows
Leaning
Head gesture
Voice command