



OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG



FAKULTÄT FÜR
INFORMATIK

Institut für Simulation und Graphik

Abschlussarbeit

zur Erlangung des akademischen Grades
Master of Science (M.Sc.)

Entwicklung und Evaluierung multimodaler berührungsloser AR-Interaktionstechniken für Multitasking-Szenarien

Verfasser:

Alina Solovjova

15. April 2019

Erstgutachter:

Jun.-Prof. Dr. Christian Hansen

Zweitgutachter:

Dr. Gabriel Mistelbauer

Betreuer:

M.A. Benjamin Hatscher

Inhaltsverzeichnis

Kurzfassung	3
Selbstständigkeitserklärung	5
1 Einleitung	7
1.1 Motivation und Zielsetzung	7
1.2 Aufbau der Arbeit	8
2 Grundlagen und Stand der Technik	9
2.1 Mensch-Computer-Interaktion	9
2.1.1 Definition	9
2.1.2 MCI im Operationssaal	10
2.1.3 Fußinteraktion	12
2.1.4 Handgesteninteraktion	15
2.1.5 Sprachinteraktion	16
2.1.6 Inertial-Tracking	17
2.1.7 Eye-Tracking und Head-Tracking	18
2.1.8 Multimodale Interaktion	20
2.2 Augmented Reality	21
2.2.1 Definition	21
2.2.2 Formen und Anwendungsgebiete von AR	22
2.2.3 HoloLens im Operationssaal	24
2.3 Belastung im Operationssaal	27
2.4 Zusammenfassung	28
3 Konzeption und Umsetzung	31
3.1 Verwendete Interaktionstechnologien	31
3.1.1 Head-Tracking	31
3.1.2 Fußinteraktion durch Inertialsensor	32
3.1.3 Handgesteninteraktion	33
3.1.4 Sprachinteraktion	34
3.2 Module	35
3.2.1 Modul 1: Abstrakte medizinische Primäraufgabe	35
3.2.2 Modul 2: Selektionsaufgabe	36
3.2.3 Modul 3: Manipulationsaufgabe	37
3.3 Implementierung	38
3.3.1 Programmierumgebung	39
3.3.2 Prototyp 1: Abstrakte medizinische Primäraufgabe	39
3.3.3 Prototyp 2: Selektionsaufgabe	39
3.3.4 Prototyp 3: Manipulationsaufgabe	43

4	Evaluation	47
4.1	Aufbau und Ablauf	47
4.1.1	Selektionsblock	50
4.1.2	Manipulationsblock	52
4.2	Ergebnisse und Diskussion	54
4.2.1	Ergebnisse: Selektionsblock	54
4.2.2	Diskussion: Selektionsblock	60
4.2.3	Ergebnisse: Manipulationsblock	62
4.2.4	Diskussion: Manipulationsblock	66
5	Abschluss	69
5.1	Zusammenfassung	69
5.2	Fazit und Ausblick	70
	Anhang	73
A.	Fragebögen	73
A.1	Demographische Daten	73
A.2	Fragen des abschließenden Interviews	73
A.3	NASA TLX	74
A.4	Menschliche Silhouette Fragebogen	75
B.	Quantitative Ergebnisse	76
B.1	Selektionsblock	76
B.2	Manipulationsblock	78
	Abkürzungsverzeichnis	81
	Abbildungsverzeichnis	83
	Literaturverzeichnis	87

Kurzfassung

Der Einsatz von Augmented Reality im Operationssaal soll Chirurgen helfen, die kognitive Lücke beim Übertragen von präoperativen Patientendaten auf den Patienten zu überbrücken. Mit der Verwendung von Head-Mounted Displays soll der Einsatz von Augmented Reality erleichtert und die Hand-Augen-Koordination der Chirurgen unterstützt werden. Allerdings stellt die Interaktion im Operationssaal eine besondere Herausforderung dar, die auch bei der Interaktion mit Head-Mounted Displays Untersuchungen bedarf.

Um dem Problem der Interaktion nachzugehen, werden in dieser Arbeit die multimodalen Interaktionstechniken Head-Tracking und Handgesteninteraktion, Head-Tracking und Sprachinteraktion sowie Head-Tracking und Fußinteraktion in Verbindung mit der Microsoft HoloLens untersucht. Dabei werden die Techniken innerhalb der Interaktionsaufgaben der Selektion und Manipulation eingesetzt. Der Fokus liegt nicht nur auf der Beurteilung der Interaktionstechniken innerhalb der einzelnen Aufgaben, sondern auch auf dessen Einsatz in Verbindung mit einer primären Aufgabe.

Im Rahmen einer Studie wurde gezeigt, dass im Vergleich zwischen den Durchläufen mit und ohne Primäraufgabe alle Interaktionstechniken innerhalb des Multitasking-Szenarios fehleranfälliger waren und eine längere Zeit beanspruchten. Den größten Unterschied bezüglich der subjektiven Bewertungen zeigte die Handgesteninteraktion. Diese hatte bei den Durchläufen ohne Primäraufgabe die besten und in den Durchläufen mit Primäraufgabe die schlechtesten Bewertungen. Währenddessen wurde die Sprachinteraktion für die Selektionsaufgabe am geeignetsten und für die Manipulationsaufgabe am ungeeignetsten empfunden. In zukünftigen Arbeiten sollte dieser Sachverhalt in einer konkreten medizinischen Aufgabe in einem operativen Umfeld näher untersucht werden.

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Masterarbeit selbstständig und ausschließlich unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Stellen sind als solche kenntlich gemacht.

Die Arbeit wurde bisher in gleicher oder ähnlicher Form weder einer anderen Prüfungsbehörde vorgelegt oder noch anderweitig veröffentlicht.

Unterschrift

Datum

1 Einleitung

Der Ablauf einer bildgeführten Chirurgie wird in eine präoperative und eine intraoperative Phase unterschieden. Die erste Phase umfasst die drei Arbeitsschritte der Datenerfassung, der Diagnose und der computergestützten Planung. Die zweite Phase besteht aus der bildgeführten Operation selbst. Immer mehr Technologien werden entwickelt und integriert, um die beiden Phasen effektiver, kostengünstiger und für den Patienten risikofreier zu gestalten. Die Unterstützung von Computersystemen ist nicht nur in der präoperativen, sondern auch in der intraoperativen Phase entscheidend [78]. Für die genaue Umsetzung der computergestützten Planung innerhalb des Operationsaals sind intraoperative Navigationssysteme und der Zugriff auf die Bild- und Planungsdaten des Patienten notwendig. Ein großes Problem ist das Übertragen der angezeigten Planungsdaten auf den Patienten, was dem Chirurgen hohe kognitive und visuelle Fertigkeiten abverlangt. Des Weiteren muss der Chirurg den Fokus immer weg vom Patienten auf den Navigationsbildschirm legen, was risikobehaftet ist. Augmented Reality (AR) ermöglicht es diese kognitive Lücke zu überbrücken, indem die präoperativen Daten auf die reale Anatomie des Patienten überlagert werden [33]. Mit dem Erscheinen von optischen See-Through Head-Mounted Displays (HMDs) können AR Visualisierungen direkt im Sichtfeld des Chirurgen angezeigt werden. Die Microsoft HoloLens (HoloLens) ist ein, mit eigener Recheneinheit ausgestattetes, Mixed Reality (MR) HMD. Durch ihre fortschrittliche Leistung hat die HoloLens großes Interesse geweckt, sie in den Operationsaal zu integrieren [58, 91].

1.1 Motivation und Zielsetzung

Der intraoperative Arbeitsfluss und die Zusammenarbeit des Ärzteteams kann durch den Einsatz der HoloLens oder einem ähnlichen Gerät erleichtert werden. Voraussichtlich werden in naher Zukunft solche HMDs fester Bestandteil von medizinischen Anwendungen und dem Operationsaal sein [90]. Dabei kann die intraoperative Anwendung durch Einblendung von Organmodellen und Patientendaten [67] bis hin zu holographisch geführten Navigationssystemen reichen [65]. Allerdings bleibt, wie auch bei anderen intraoperativen computergestützten Systemen, die Frage der Benutzerinteraktion offen. Welche Interaktionstechniken sind für den Chirurgen am angenehmsten in Verbindung mit der HoloLens anwendbar und welche erfüllen sein Vorhaben am effizientesten? Welche lassen sich in den Operationsablauf am besten integrieren und werden gleichzeitig den spezifischen Anforderungen eines Operationsaals gerecht?

In dieser Masterarbeit werden verschiedene multimodale Interaktionstechniken innerhalb der AR Umgebung mit Hilfe der HoloLens untersucht. In der Theorie soll der Einsatz der HoloLens unter anderem dabei helfen paralleles Arbeiten zu ermöglichen, indem die Instrumente nicht abgesetzt werden müssen oder der Chirurg seine Position nicht ändern

muss. Deshalb ist es von großem Interesse zu untersuchen, wie sich die Interaktionstechniken und das parallele Arbeiten gegenseitig beeinflussen. Dabei ist es wichtig zu erfahren, ob die Interaktionstechniken durch die gleichzeitige Handhabung von Primäraufgabe (der Operation) und Sekundäraufgabe (der Interaktion) eine höhere kognitive Belastung verursachen, wodurch die Ausführung der Primäraufgabe beeinträchtigt werden könnte. Dieser Sachverhalt wird im Rahmen dieser Arbeit als Multitasking-Kontext bezeichnet.

Hierfür werden verschiedene multimodale Interaktionstechniken bestehend aus Kopfposition und Sprachkommando, Kopfposition und Fußgeste sowie Kopfposition und Handgeste innerhalb einer Studie miteinander verglichen und evaluiert. In der Studie werden für die Untersuchung die grundlegenden Interaktionen der Selektion und Manipulation verwendet. Hierbei werden die Interaktionsaufgaben unter zwei Bedingungen durchgeführt. Das Ziel dieser Arbeit ist Erkenntnisse darüber zu erlangen, welche der genannten Interaktionstechniken zum einen die geeignetste Wahl für den Aufgabenkontext darstellen und zum anderen herauszufinden, inwieweit sie sich parallel mit der Primäraufgabe vereinen lassen.

Einen neuen Beitrag dieser Arbeit stellt die Integration von Inertialsensoren für die Fußinteraktion in Verbindung mit der HoloLens dar. Des Weiteren werden die Interaktionstechniken nicht nur selbst untersucht und untereinander verglichen, sondern ebenfalls innerhalb eines Multitasking-Kontexts evaluiert. Derzeit sind keine verwandten Arbeiten bekannt, die Fußinteraktion über Inertialsensoren in Verbindung mit der HoloLens durchführen oder multimodale Interaktionstechniken innerhalb eines Multitasking-Kontexts evaluieren.

1.2 Aufbau der Arbeit

Die Arbeit umfasst weitere folgende Kapitel:

- **Kapitel 2 *Grundlagen***

In diesem Kapitel werden die Grundlagen aus den Gebieten der Mensch-Computer-Interaktion (MCI) und AR beschrieben. Dafür werden der Stand der Technik und die Forschungsergebnisse von verwandten Arbeiten präsentiert. Ein weiterer Fokus wird auf die Belastung der Chirurgen während einer Operation gerichtet.

- **Kapitel 3 *Konzeption und Umsetzung***

Aufbauend auf dem Stand der Technik werden in diesem Kapitel die Konzeptionen der Interaktionsmechanismen vorgestellt und diskutiert. Daraufhin werden die entwickelten Prototypen beschrieben, die infolge der Evaluation untersucht werden.

- **Kapitel 4 *Evaluation***

Dieses Kapitel stellt den Aufbau, Ablauf und die Ergebnisse der durchgeführten Studie vor. Anschließend wird basierend auf den Ergebnissen eine Diskussion durchgeführt.

- **Kapitel 5 *Abschluss***

Zum Schluss wird die Arbeit zusammengefasst. Daraufhin werden die zentralen Erkenntnisse gebündelt und die zukünftigen Arbeitsschritte beschrieben.

2 Grundlagen und Stand der Technik

Im Folgenden werden die Grundlagen und der Stand der Technik behandelt, die als Argumentationsgrundlage für diese Arbeit verwendet werden. Hierfür wird dieses Kapitel in drei Abschnitte gegliedert. Der erste Abschnitt behandelt den Bereich der **MCI**, in dem auf die Entstehung des Forschungsbereiches und dessen Ziele eingegangen wird. Daraufhin wird die derzeitige Interaktion mit Computersystemen innerhalb des Operationssaals beleuchtet, um die Problematik darzustellen, die diese Arbeit unter anderem motiviert. Darauf aufbauend werden verschiedene berührungslose Interaktionstechnologien beschrieben, die für eine Interaktion innerhalb des Operationssaals in Betracht kommen könnten oder bereits eingesetzt werden. Abschließend wird auf den Begriff der multimodalen Interaktion eingegangen. Der zweite Abschnitt befasst sich mit der Definition und den Anwendungsgebieten von **AR**. Anschließend wird die HoloLens vorgestellt und dessen Einsatz im Operationssaal diskutiert. In diesem Zusammenhang werden verwandte Arbeiten mit teilweise intraoperativen Bezug vorgestellt, welche die HoloLens einsetzen. Der letzte und dritte Abschnitt behandelt die intraoperative Belastung des Ärzteteams, speziell der Chirurgen, und wie sich Belastung auf die Leistungsfähigkeit auswirkt.

2.1 Mensch-Computer-Interaktion

Der Fokus dieses Abschnittes liegt auf der Benutzerinteraktion innerhalb des Operationssaals und die damit verbundenen, situationsbedingten Probleme. Als mögliche Lösungen für die Problemstellung werden verschiedene Interaktionstechnologien vorgestellt und diskutiert, darunter Eye- und Head-Tracking, Sprach-, Fuß- und Handgesteninteraktion sowie die Verwendung von Inertialsensoren. Abschließend wird der Begriff der multimodalen Interaktion beleuchtet und Aspekte benannt, die bei der Entwicklung von multimodalen Systemen beachtet werden sollen.

2.1.1 Definition

Die Kommunikation zwischen Mensch und Computer wird fachlich als **MCI**, im Englischen Human-computer-interaction (**HCI**), bezeichnet. **MCI** ist ein multidisziplinärer und großer Forschungsbereich der Informatik und Teilgebiet der Mensch-Maschine-Interaktion. Seinen Ursprung hat **MCI** in den 1980er Jahren. Mit dem Erscheinen des Apple Macintosh 1984 wurde dem Normalnutzer der Zugang zum Computer ermöglicht. Zusätzlich wurde 1983 die erste Association for Computing Machinery (**ACM**) **SIGCHI** Konferenz veranstaltet. Die **SIGCHI** ist eine Themengruppe der **ACM** im Bereich der **MCI**.

MCI umfasst verschiedene Forschungsdisziplinen wie z.B. Psychologie, Soziologie, Kognitionswissenschaften, Informatik und Linguistik [60] und beschäftigt sich mit dem Design,

der Evaluierung und der Implementierung interaktiver Computersysteme [30]. Ziel der MCI ist die Unterstützung des Benutzers in der bestmöglichen Erfüllung der verfolgten Aufgaben durch ein technisches System [30, 46]. Ein essentielles Konzept der MCI ist die Gebrauchstauglichkeit (*engl.: usability*). Nach der DIN-Norm 9241-11 1998 besteht sie aus den Maßen der Effektivität, der Effizienz, der Zufriedenstellung und wird folgendermaßen definiert:

Das Ausmaß, in dem ein Produkt durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen.

Entnommen aus der DIN-Norm 9241-11 1998.

Das entwickelte System soll einfach zum Lernen und Nutzen sein und Fehler in seiner Häufigkeit und Schwere reduzieren [30, 46].

2.1.2 MCI im Operationssaal

Im Operationssaal gelten strenge Regeln, die viel Disziplin erfordern. Die Rollen und die Aufgaben innerhalb des Ärzteteams sind zugeordnet und klar festgelegt. Damit eine möglichst einwandfreie Operation ermöglicht werden kann und das Leben des Patienten nicht unnötig gefährdet wird, müssen sich alle an die folgenden Vorschriften halten [56]. Im Operationssaal hat Sterilität die oberste Priorität: Um diese zu gewährleisten unterliegt alleine schon der Aufbau der Operationsabteilung festgelegten Vorschriften. Dieser ist nur über entsprechende Schleusen zugänglich, welche die verschiedenen Arbeitsbereiche voneinander trennen. Um eine Erregerverschleppung zu vermeiden, werden die Räume in rein und unrein unterteilt. Die reinen Räume dürfen nur mit ordnungsgemäßer Desinfizierung und Kleidung betreten werden [56].

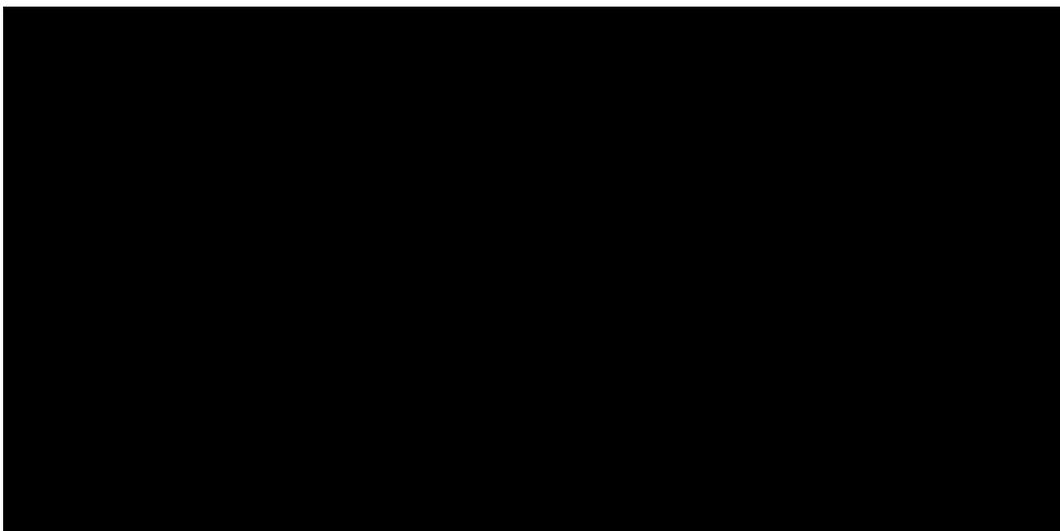


Abbildung 2.1: Skizze der Zonen um den Operationstisch. Der rote Bereich zeigt die Sterilzone. Entnommen aus [1].

Der Bereich um den OP-Tisch ist ebenfalls in drei Bereiche unterteilt: Die Sterilzone, die Springerzone und die Anästhesiezone. In **Abbildung 2.1** werden die verschiedenen Zonen

um den OP-Tisch dargestellt. Die Sterilzone umfasst den Patienten, den Instrumententisch, das steril bekleidete, operierende Team und teils steril bezogene Geräte. Weitere anwesende Personen müssen einen Sicherheitsabstand von mindestens einem Meter berücksichtigen. Die Springerzone, benannt nach der OP-Pflegekraft, befindet sich um die Sterilzone herum. Der Springer hält ebenfalls Abstand von der Sterilzone und besorgt dem operierenden Team falls nötig zusätzliche Materialien, die er dann vorsichtig in die Sterilzone überreicht. In der Anästhesiezone, die sich in dem Fall der [Abbildung 2.1](#) am Kopfende des Patienten befindet, hält sich der Anästhesist auf. Diese Zone gilt als unsteril [56].

Während der bildgeführten Chirurgie ist die Interaktion mit digitalen Bilddaten unerlässlich. Im Gegensatz zu der alltäglichen Interaktion mit Computersystemen bei Normalnutzern wird die Interaktion im Operationssaal durch die soeben genannten Einschränkungen deutlich erschwert. Der Prozess der Analyse, Implementierung, Evaluierung und des Designs in der MCI muss hier anders durchgeführt werden. Aufgrund der räumlichen Einschränkungen und der komplexen technischen Umgebung müssen vor allem Aspekte der Ergonomie berücksichtigt werden [78]. Viele Geräte sind auf Rollwägen angebracht, weshalb bei der Entwicklung von Systemen nicht von fixen Positionen ausgegangen werden darf ([Abbildung 2.2](#)).

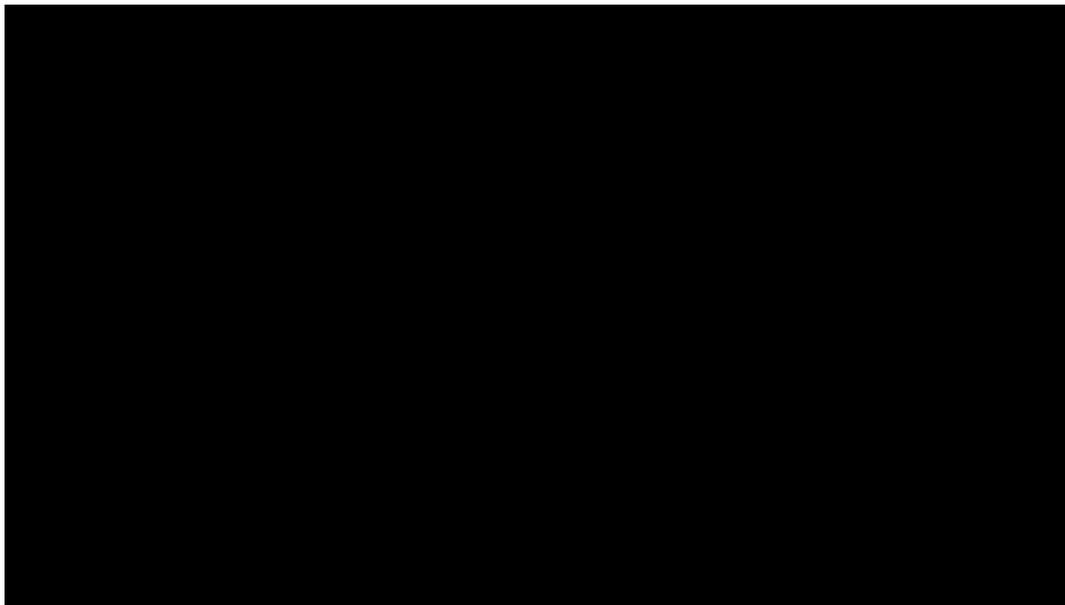


Abbildung 2.2: Skizze eines Angiographie Operationssaals mit den beweglichen Geräten in blau. Entnommen aus [51].

Je nachdem wo sich das System befindet, dürfen nur bestimmte Personen damit interagieren. Befindet sich das System außerhalb der Sterilzone, kann nur ein nicht steriler Assistent dieses bedienen. Bedient es dennoch jemand aus dem operierenden Team, muss nach der Interaktion derjenige sich wieder desinfizieren, was wiederum sehr zeitaufwändig ist. Um Zeit zu sparen fassen manche die Geräte durch die unsterile Seite ihres OP-Kittels an, was allerdings auch riskant ist [51]. Befindet sich das System innerhalb der Sterilzone, wird es mit einer sterilen Folie abgedeckt und darf nur vom operierendem Team bedient werden. Meist wird das System über einen Joystick oder einem Touchscreen mit zusätzlichen Tasten bedient [78]. Häufig hat der Chirurg jedoch nicht die Hände frei, um mit den Bilddaten zu interagieren, sodass ein Assistent die gewünschte Interaktion durchführen muss. Dies

kann zu Verzögerungen und bei missverstandenen Anweisungen zu Frustration führen [71]. Die Studie von Johnson et al. zeigte ebenfalls, dass Chirurgen ein besseres mentales Model aufbauen und somit eine bessere Einschätzung des Fortschritts der Operation erhalten, wenn sie selbst mit den Bilddaten interagieren [51].

Durch die vielen Einschränkungen gestaltet sich die Interaktion im Operationssaal problematisch. Aufgrund dessen rückten in den letzten Jahren immer mehr berührungslose Interaktionstechnologien in den Fokus. Mithilfe dieser Technologien sollen die genannten Probleme weitgehend umgangen und die Kosten gesenkt, sowie Sicherheit und Effizienz von computergestützten Chirurgien verbessert werden.

2.1.3 Fußinteraktion

Eine klassische Art von Fußinteraktion ist die Betätigung von Pedalen. Das geläufigste Beispiel hierfür ist die Bedienung eines Autos, bei der drei Pedale betätigt werden müssen. Ein weiteres Beispiel ist das Drücken von Klavierpedalen um den Klavierklang zu beeinflussen. Auch im Operationssaal werden Fußschalter verwendet (*Abbildung 2.3*), z.B. um die Höhe des Operationstisches zu verstellen [78], zur Erfassung von fluoroskopischen Bildern [51] oder für die Verwendung von zahnmedizinischen Instrumenten [100]. Die Interaktion über Füße ist eine Möglichkeit ohne Einsatz der Hände mit Systemen zu interagieren und dadurch die Sterilität besser bewahren zu können.

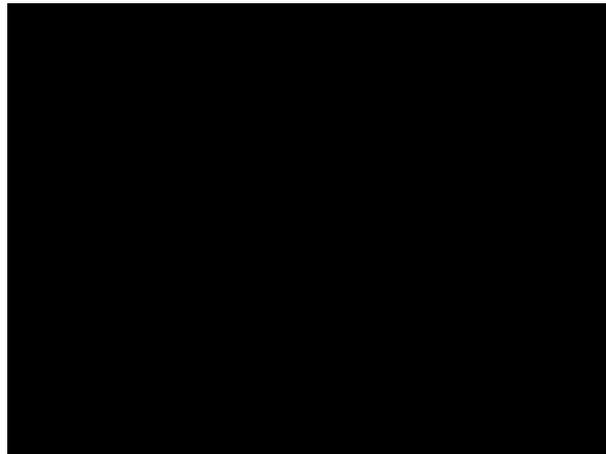


Abbildung 2.3: Im Operationssaal verwendeter Fußschalter. Entnommen aus [13].

Die genannten Beispiele gehören zu dem vermittelnden Erfassen (*engl.: mediated sensing*) von Fußeingaben. Hierbei werden nicht die Füße selbst, sondern die mechanischen Geräte, die von den Füßen bedient werden, gemessen. Weitere Möglichkeiten zur Erfassung der Fußeingabe sind das intrinsische Erfassen (*engl.: intrinsic sensing*) und das extrinsische Erfassen (*engl.: extrinsic sensing*). Ersteres beinhaltet Sensoren die direkt am Fuß angebracht sind. Meist handelt es sich um Inertialsensoren (*Unterabschnitt 2.1.6*). Letzteres erfasst die Fußeingabe von außen über Sensoren, die in der Umgebung angebracht sind. Hierzu werden meist Infrarotkameras verwendet [95]. Bei der intrinsischen und extrinsischen Methode sind die Benutzer nicht an einem festen Ort gebunden, wie es bei der *mediated* Methode der Fall ist. Jedoch erheben sie im Gegensatz dazu immer Daten, wodurch beim Design berücksichtigt werden muss, dass keine unerwünschten Interaktionen

durchgeführt werden. Ähnlich zu der Klassifikation von Gesten nach McNeill [62] unterscheiden Velloso et al. in ihrer Veröffentlichung vier Kategorien von Fußgesten in der MCI [95].

semaphorisch

Dabei handelt es sich um spezifische Fußgesten, die zu einem Verzeichnis von definierten Fußgesten gehören. Eine Liste von, in der Literatur verwendeten semaphorischen, Fußgesten ist in [Abbildung 2.4](#) zu sehen.

deiktisch

Deiktische Gesten werden üblicherweise zum Zeigen verwendet, ähnlich zum Steuern eines Mauszeigers.

manipulativ

Manipulative Fußgesten bilden z.B die Position oder Orientierung des Fußes auf Eigenschaften der Systemobjekte ab.

implizit

Beinhaltet Interaktionen, die nicht primär darauf ausgelegt sind mit einem System zu interagieren. Es handelt sich um unbewusste Signale, die von einem System erfasst und interpretiert werden, wie die Messung von Temperatur und Druck im Schuh für Diabetiker.

Die Fußsteuerung ist zwar nicht sehr präzise, jedoch eignet sie sich gut für sekundäre Aufgaben [78] oder als Unterstützung für andere Modalitäten. Für primäre Aufgaben wird Fußinteraktion dann verwendet, wenn die Hände des Benutzers gerade nicht frei sind [27, 95], was bei Chirurgen während der Operation überwiegend der Fall ist. Vor allem für die Änderung von kontinuierlichen Werten eignet sich Fußinteraktion sehr gut [27]. Aus diesen genannten Gründen kann Fußinteraktion, außerhalb von Fußschaltern, sehr gut in den intraoperativen Arbeitsfluss integriert werden.

Die Ergebnisse der Studie von Hatscher et al. zeigten, dass für die Aktivierung eines Systems Fußinteraktion in Form von einem Double-tap schneller ist, als die Aktivierung über Sprache oder Handgeste [40]. Interaktionen mit der Computermaus, wie Klick, Doppelklick oder Klicken und Halten, wurden in Rajanna & Hammond durch den Fuß ersetzt [82]. Die Fußinteraktion wurde mit Blick kombiniert und die Interaktion war in der Studie genau so schnell wie mithilfe der Computermaus, solange die Zielobjekte etwas größer waren.

Die am meisten verwendeten Fußinteraktionen in der Literatur sind die Verlagerung von Druck, auch *rocking* genannt, die Rotation und das „Sliden“ (*dt.: rutschen, gleiten*) des Fußes [54]. Resultate einer Studie zeigen, dass Rotation und Sliding schneller durchzuführen sind und von den Benutzern besser bewertet wurden, hinsichtlich der Arbeitsbelastung [54].

Verschiedene Fußinteraktionskonzepte mit Rotation der Ferse (*Heel rotation*) zur Unterstützung von radiologischen Interventionen wurden in Hatscher et al. untersucht [41]. Die Konzeptionen sollen zum Scrollen von medizinischen Bilddaten verwendet werden. Ein Konzept beinhaltet imaginäre diskrete Schaltflächen, die in einem radialen Winkel um den Fuß angebracht sind. Dabei bestimmt die Rotation des Fußes, welche Schaltfläche betätigt wird. Ein anderes Konzept nutzt die kontinuierliche Rotation des Fußes für das Scrollen. Das letzte Konzept kombiniert die zwei genannten Konzepte, indem in der Mitte

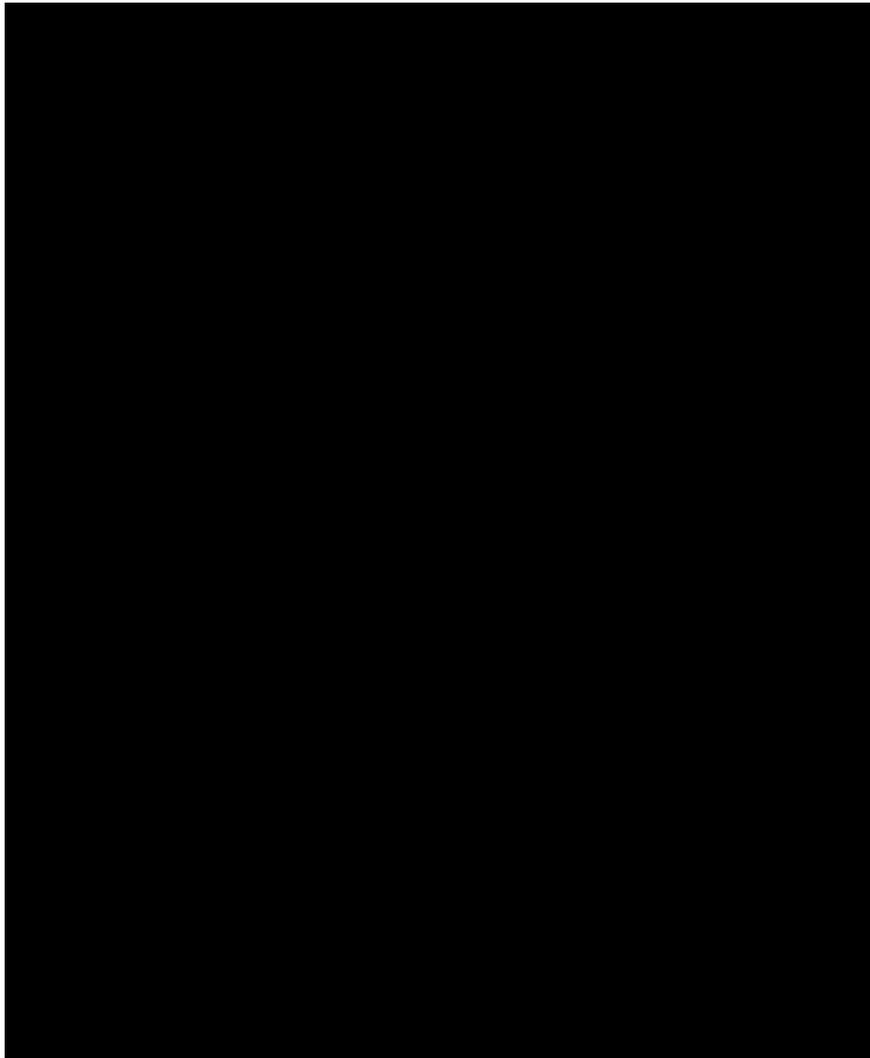


Abbildung 2.4: Eine Übersicht der in der Literatur verwendeten Fußgesten. Entnommen aus [95].

eine kontinuierliche Rotationseingabe möglich ist und am Rand diskrete Schaltflächen verwendet werden. Das Ergebnis zeigt, dass kontinuierliche Rotationen des Fußes die beste subjektive Bewertung und die schnellste Task Completion Time innerhalb der Studienaufgabe haben. Auch Velloso et al. bekräftigen die Aussage, dass die *Heel rotation* eine der angenehmsten Bewegungen für den Benutzer ist [94].

Zaman et al. entwickelten einen interaktiven Schuh für Chirurgen, die die Manipulation von 2D Bilddaten ermöglicht. Hierfür wurde der optische Sensor einer Computermaus in die Schuhsohle verarbeitet und ein Mikroschalter zum an- und ausschalten des Schuhs angebracht [100]. Für das vor und zurück Scrollen der Bilddaten muss der Fuß vor oder zurück bewegt werden. Ähnlich hierzu verwendeten Solovjova et al. eine Schuhsohle mit integrierten Drucksensoren für das Scrollen von Bilddaten [89]. Für die vor und zurück Manipulation muss entweder Druck auf den Vorder- oder Hinterfuß ausgeübt werden. Das System wird über Druckausübung auf den Zeh an und über Druckausübung auf die Ferse ausgeschaltet. Problematisch ist, dass solche Systeme unbeabsichtigte Eingaben zulassen, weswegen Fußinteraktionen besser multimodal eingesetzt werden sollten.

2.1.4 Handgesteninteraktion

Angesichts des in [Unterabschnitt 2.1.2](#) beschriebenen Problems bezüglich der Interaktion wird die Interaktion von Handgesten in der Literatur umfassend behandelt. Freihandgesten sind zum einen für den Menschen eine natürliche Interaktionsform und somit häufig sehr intuitiv, zum anderen können sie steril im Operationssaal integriert werden. Handgesten werden häufig sprachbegleitend ausgeführt [79] und lassen sich während ihrer Durchführung in folgende Phasen unterteilen:

1. Ruheposition
Dabei handelt es sich um eine stabile Position, von der aus die Geste begonnen wird.
2. Vorbereitungsphase
Hierbei wird eine Bewegung von der Ruheposition weg begonnen.
3. Gestenbewegung (*engl.: stroke*)
Die eigentliche Geste, die den meisten Informationsgehalt besitzt und dessen Bewegung den höchsten Aufwand fordert.
4. Haltephase (*engl.: stroke hold*)
In dieser Phase findet keine Bewegung statt, sie dient der Unterstützung der eigentlichen Geste. Sie kann auch vor der Ausübung der eigentlichen Geste stattfinden (*engl.: prestroke hold*).
5. Erholungsphase
Die Hände werden wieder in den Ausgangszustand, Phase 1 gebracht. Erfolgen mehrere Gesten hintereinander, entfällt diese Phase.

Mit Hilfe dieser Charakteristiken wird technischen Systemen geholfen Handgesten zu erkennen [79]. Durch die Einführung von günstigen sensing und tracking Technologien, ist auch die Integration von Handgesteninteraktion im Operationssaal möglich.

O'Hara et al. entwickelten ein berührungsloses Interaktionssystem für endovaskuläre Interventionen [70]. Es basiert auf verschiedenen Handgesten, die sprachbegleitend ausgeführt werden. Es zeigt sich, dass das System sehr nützlich für die Chirurgen ist, da sie selbst mit den Bilddaten interagieren können und nicht auf die Hilfe von Assistenten angewiesen sind. Des Weiteren wird eine neue Art der kollaborativen Zusammenarbeit während der Operation ermöglicht. Schwierigkeiten stellt jedoch die Platzierung des Sensors dar, da der Chirurg weiterhin an einen reaktiven Bereich gebunden ist. Zudem kann die Interaktion, je nach Position des Chirurgen, umständlich sein, sodass der Assistenzarzt das System weiterhin bedienen muss.

Eine weitere Schwierigkeit von Systemen, die Handgesteninteraktion verwenden, ist zum einen das Erkennen der Handgesten und zum anderen die Trennung von beabsichtigten und unbeabsichtigten Handgesten. Abgesehen von der eigentlichen Erkennung der Geste kann diese durch mehrere Personen erschwert werden. Systeme wie die Kinect erfassen durch Infrarotstrahlen Punkte am Körper des Benutzers in Form von einem abstrakten Skelett. Befinden sich mehrere Personen nah beieinander kann die Rekonstruktion des Skeletts verzerrt werden, was zu einer fehlerhaften Einschätzung der Bewegung innerhalb des Systems führt [70]. Mit der Trennung der beabsichtigten und unbeabsichtigten Handgesten beschäftigten sich Jacob et al. [47]. Hierfür wurde der zusätzliche Operationskontext hinzugezogen, z.B. die Position der Biopsie Nadel, die den Bereich von Interesse

der Bilddaten eingeschränkt oder die Körperorientierung des Benutzers zum System. Allerdings bleibt das Problem, dass für die Interaktion über Handgesten mindestens eine Hand benötigt wird, und, wie zuvor erwähnt, haben Chirurgen nicht immer eine Hand frei [51]. Zwar erforschten Rossol et al. die Möglichkeit der Ausführung von Handgesten trotz in der Hand gehaltenen Instrumenten [85], jedoch können nicht bei allen Operationen die Instrumente mitbewegt werden.

2.1.5 Sprachinteraktion

Eine der natürlichsten Interaktionen für den Menschen ist die Sprache. Deswegen wird Sprachinteraktion schon lange in dem Bereich der MCI untersucht. Sie eignet sich hervorragend für viele multimodale Anwendungen. Sprache kann zum einen sehr effizient sein, da in kurzer Zeit viele Informationen vermittelt werden können, zum anderen kann sie für die Interaktion mit entfernten und nicht im Augenkontakt stehenden Systemen verwendet werden [36]. Deshalb eignet sie sich in der Theorie gut für den Einsatz im Operationssaal, da sie wie die Fußinteraktion freihändig und steril ist.

Sprachkommandos können verwendet werden, um diskrete Zustände wie z.B. das Wechseln von Modi und Funktionalitäten zu ändern [71, 67]. Hötker et al. verglichen die Verwendung von Sprachinteraktion mit Handgesteninteraktion innerhalb der interventionellen Radiologie. Hierfür wurde eine vordefinierte Menge von sechs Handgesten und sechs Sprachkommandos verwendet, welche von Radiologen getestet wurde [45]. Das Ergebnis zeigte, dass Sprachinteraktion in 93% der Fälle erkannt wurde und weniger fehleranfällig ist als Handgesteninteraktion. In 30% der Fälle wurden Handgesten unbeabsichtigt vom System erkannt. Der Vergleich wurde allerdings nicht intraoperativ durchgeführt. Im Gegensatz dazu zeigten Mentis et al., dass es keine klare Aussage darüber gibt welche Funktionalitäten sich besser über Handgesten oder Sprache steuern lassen, sondern dass die Verwendung der Interaktionstechniken stark von der Situation abhängig ist [64]. Beide Veröffentlichungen sind sich jedoch einig, dass Sprache sich für das Ändern von kontinuierlichen Werten nicht gut eignet.

Bei dem Einsatz von Spracherkennung in Systemen muss viel Wert auf die Interaktionsweise gelegt werden. Spracherkennung hat eine höhere Erkennungsrate bei der Eingabe von Sprachkommandos als bei der Eingabe von natürlicher Sprache. Des Weiteren müssen die Sprachkommandos so gewählt werden, dass sie einfach zu merken und phonetisch unterschiedlich genug sind, um Fehlerkennung und Verwechslungen zu vermeiden [17].

Die Schwierigkeit bei der Integration von Sprachinteraktion im Operationssaal liegt darin, dass im Operationssaal laute Hintergrundgeräusche erhebliche Auswirkungen auf die Spracherkennung haben können. Zu den Hintergrundgeräuschen im Operationssaal zählen unter anderem: Diskussionen innerhalb des Ärzteteams, Töne des Anästhesiegeräts, Klimaanlage, das Klappern von metallischen Instrumenten oder das Absaugen von Speichel [17]. Des Weiteren müssen die Systeme zwischen normaler Kommunikation unter den Ärzten und der intentionalen Sprachinteraktion unterscheiden. Aus diesem Grund sollen Funktionalitäten redundant auf zusätzliche Modalitäten abgebildet werden, damit situations- und umgebungsbedingt Chirurgen die Wahl zwischen den passenden Modalitäten haben [64].

2.1.6 Inertial-Tracking

Inertialsensoren wie Beschleunigungs- und Drehratensensoren messen die Beschleunigung und Rotationsgeschwindigkeit eines Körpers. Die räumliche Kombination mehrerer Inertialsensoren bildet eine inertielle Messeinheit, auch Inertial Measurement Unit (IMU) genannt, worauf das Inertial-Tracking basiert. Es wird zwischen linearen Inertialsensoren und Gyrosensoren unterschieden. Erstere erfassen die Beschleunigung entlang einer Achse, zweitere erfassen die Winkelbeschleunigung um eine Achse. Inertial-Tracking wird für die Bestimmung der Lage eingesetzt, z.B. durch Erfassung von Gelenkstellung eines Benutzers (Abbildung 2.5), indem einzelne Inertialsensoren an den jeweiligen Gliedmaßen angebracht werden [30].



Abbildung 2.5: Links: Ein Inertialsensor (orange) mit einem Bluetooth Empfänger. Rechts: Angebrachte Sensoren zur Erfassung von Bewegungen. Entnommen aus [87] und [8].

Inertialsensoren können nicht nur zur Analyse von Bewegungen genutzt werden, sondern auch für Interaktionen. Sie sind in diversen Geräten integriert, wie z.B. in HMDs, VR/AR-Brillen, Armbänder oder Smartphones. Im Gegensatz zum optischen Tracking erfordern Inertialsensoren keine direkte Sichtverbindung (*engl.: line of sight*) und nur festgelegte Personen können mit dem System interagieren [50].

Hatscher et al. verwendeten einen IMU und Distanzsensoren, welche an einem Schuh angebracht wurden, zur Entwicklung verschiedener Konzepte zur Fußinteraktion [41]. Phienthrakul nutzte ein Myo Gestensteuerungsarmband um einen Mauszeiger zu steuern und über Handgesten Objekte zu selektieren [77]. Das Myo Armband ist mit einem EMG Sensor, welches die elektrische Aktivität der Muskeln misst, ausgestattet und einem 9-Achsen Inertialsensor. Die Steuerung wurde über den Inertialsensor und die Auswahl über den EMG Sensor getätigt. Es wurde gezeigt, dass das Myo Armband zur Steuerung des Mauszeigers gut verwendet werden kann, jedoch die Selektion noch nicht zuverlässig genug ist. Trotz angewandter Machine Learning Algorithmen ist die Erkennung der Handgesten

noch nicht ausgereift genug, um das Myo Armband in kritischen Anwendungsaufgaben zu verwenden.

In der Veröffentlichung von Young et al. wurde ein tragbares 3D Eingabegerät mit zwei integrierten Inertialsensoren verwendet, um einen Mauszeiger über Armbewegungen zu steuern [99]. Hierfür wurde es an einem Arm befestigt. Es wurden drei verschiedene Selektionsarten nach der ISO 9241-9 Aufgabe untersucht: Die Selektion über das Verweilen des Mauszeigers, über eine Drehgeste und über das Drücken eines zusätzlichen Knopfs. Es gab keinen signifikanten Unterschied zwischen den Techniken für die Task Completion Time. Die Selektion über den Knopf hatte jedoch eine signifikant höhere Anzahl an Fehltreffer verglichen mit der Selektion über die Drehgeste. Dies ist mit der „anhaftenden“ Eigenschaft des Mauszeigers bei der Drehgeste begründet. Das Steuern des Mauszeigers über Armbewegungen wird jedoch als sehr ermüdend empfunden.

Durch die Unabhängigkeit der Orientierung des Benutzers und der Möglichkeit der sterilen Interaktion, können Inertialsensoren in der Theorie gut im Operationsaal eingesetzt werden.

2.1.7 Eye-Tracking und Head-Tracking

Beim Eye-Tracking wird die Blickrichtung des menschlichen Auges verfolgt, um den Blickverlauf einer Person zu erfassen und auszuwerten. Es wird zwischen invasiven und nichtinvasiven Verfahren unterschieden. Bei invasiven Eye-Tracking Verfahren wird ein direkter Eingriff am Körper des Benutzers vorgenommen. Dies kann in Form von Kontaktlinsen sein, die mit kleinen Spulen oder Reflektoren versehen sind, durch welche sich die relative Bewegungen des Auges oder die relative Blickrichtung ableiten lassen. Im Gegensatz zu invasiven Verfahren wird bei nichtinvasiven Verfahren kein direkter Eingriff am Körper des Benutzers vorgenommen. Hierfür werden spezielle Kameras verwendet, die das Auge erfassen und durch Bildverarbeitungsalgorithmen die Blickrichtung bestimmen, auch als optisches Tracking bezeichnet (Abbildung 2.6). Die Auswertung der Blickrichtung erfolgt merkmalsbasiert, modellbasiert oder aus einer Kombination von beiden. Ersteres nutzt Merkmale der Pupille wie die Pupillengeometrie zur Auswertung des Blickes. Modellbasierte Verfahren vergleichen die Bildinformationen der aufgenommenen Augenszene mit der modellierten Augenszene [30].



Abbildung 2.6: Links: Eine Datenbrille mit Eye-Tracking. Rechts: Die Orientierungsachsen des Head-Trackings. Entnommen aus [9] und [10].

Beim Head-Tracking wird anstelle der Augen die Position und Orientierung des Kopfes verfolgt. Hierfür werden meist Inertialsensoren ([Unterabschnitt 2.1.6](#)) verwendet, die in [HMDs](#) oder einem ähnlichem Gerät integriert sind ([Abbildung 2.6](#)). Eye- und Head-Tracking werden oft für die Steuerung eines Cursors und für die Selektion verwendet, da sie sich besonders gut fürs Zeigen eignen und auch teilweise schneller sein können als die Interaktion über eine Computermaus [31, 88]. Es gibt sowohl Techniken die jeweils nur Eye- oder Head-Tracking verwenden, als auch solche, die eine Kombination beider verwenden oder sie mit anderen Modalitäten verbinden.

Für die Selektion mittels Eye-Tracking werden meist zwei Methoden verwendet. Zum einen die Selektion durch das Verweilen des Blickes (*Dwell*), zum anderen durch das Blinzeln der Augenlider. Beim Blinzeln entsteht das Problem, dass dies ein Vorgang ist, der unbewusst ausgeführt wird und somit Inhalte unbeabsichtigt selektiert werden können [23]. Auch bei dem Verweilen des Blickes für eine gewisse Zeitspanne tritt das sogenannte Midas Touch Problem auf [48]. Bei diesem Problem handelt es sich um eine aus der Literatur stammende Metapher. Wenn bei der Verwendung von Interaktionstechniken mittels Blick nicht unterschieden werden kann, ob der Benutzer visuelle Informationen wahrnehmen oder mit dem System interagieren möchte, wird dies als Midas Touch Problem bezeichnet. Sobald beim Verweilen des Blickes eine Stelle länger betrachtet oder die Augen eine ruhende Position einnehmen, wird selektiert, was für den Benutzer sehr frustrierend sein kann [101]. Afkari et al. beheben das Midas Touch Problem, indem ein zusätzlicher Selektionszustand eingeführt wird [16]. In deren Publikation wird ein drei Stufen Selektionsmechanismus vorgestellt. Dabei wird eine zu selektierende Stelle fokussiert und nach $250ms$ in den ersten Selektionszustand versetzt. Nach weiteren $250ms$ wird das Zielobjekt in den zweiten Selektionszustand versetzt und nach weiteren $250ms$ endgültig ausgewählt. Dadurch hat der Benutzer noch einen Augenblick Zeit seine Selektion zu korrigieren, jedoch nimmt die Selektion längere Zeit in Anspruch. Das Midas Touch Problem kann ebenfalls durch multimodale Interaktion ([Unterabschnitt 2.1.8](#)) umgangen werden, indem der Mauszeiger über die Augen gesteuert wird, die Selektion jedoch über eine weitere Modalität ausgeführt wird [27, 42, 83].

In der Publikation von Qian & Teather wurden drei Selektionstechniken miteinander verglichen: Die isolierte Selektion nur über Eye- und Head-Tracking sowie die Selektion durch eine Kombination beider Tracking Methoden [81]. Das Ergebnis der Studie zeigt, dass die Fehlerrate beim isolierten Eye-Tracking bei ungefähr 40% liegt. Auch die Dauer der Bewegung ist, verglichen mit der Selektion über Kopfposition, deutlich langsamer, wodurch die Leistung des isolierten Head-Trackings deutlich überlegen ist.

Hingegen berichteten Blattgerste et al. von der Überlegenheit des Eye-Trackings gegenüber dem Head-Tracking [24]. In der durchgeführten Studie war das Resultat, dass in einer VR/AR Umgebung Eye-Tracking deutlich schneller und weniger ermüdend ist als Head-Tracking, vor allem bei großen Sichtfeldern (*engl.: field of view*).

Bei der MAGIC Zeigetechnik wird eine Mischung aus Eye- und Head-Tracking verwendet. Hierbei wird durch das Eye-Tracking der Mauszeiger so nah wie möglich an das Zielobjekt bewegt und durch darauffolgende Kopfbewegungen das Zielobjekt erreicht [49]. Die Veröffentlichung von Jalaliniya et al. zeigt, dass MAGIC bei der Auswahl von kleinen Zielobjekten treffsicherer ist als Eye-Tracking. Bei kurzen Distanzen ist Head-Tracking signifikant schneller als MAGIC, bei langen Distanzen ist jedoch MAGIC schneller.

Eine Mischung aus Head-Tracking und Sprachinteraktion zur Manipulation von 3D AR Daten auf einem HMD wurden von Grinshpoon et al. verwendet [37]. Durch kleine Kopfrotationen kann der Benutzer das AR Model rotieren und skalieren. Über Sprache werden verschiedene Funktionen ausgewählt. Die Implementierung wurde so entworfen, dass der Benutzer trotz des begrenzten Sichtfeldes des HMDs immer das gesamte Model während der Transformationen sieht.

2.1.8 Multimodale Interaktion

Wie in den vorherigen Abschnitten gezeigt, gibt es verschiedene Interaktionstechnologien, mit denen der Mensch mit einem System interagieren kann. Die Interaktion über Sprache, Fuß- und Handgeste sowie Blick sind alles sogenannte Modalitäten. Die Kombination aus mindestens zwei oder mehr Modalitäten ermöglicht eine multimodale Interaktion. Eines der ersten multimodalen Demonstrationssysteme wurde von Bolt in dem Jahr 1990 in der Publikation *"Put That There"* präsentiert [26]. In dieser wurde es ermöglicht das virtuelle Zeigen über ein Touchpad parallel zur Sprache auszuführen. Seit der Veröffentlichung erschienen eine Vielzahl von multimodalen Systemen, wodurch sich multimodale Interaktion zu einem großen eigenständigen Forschungsbereich der MCI entwickelte. Ziel eines multimodalen Interfaces ist es transparent, flexibel und effizient zu sein, indem die natürliche menschliche Sprache und das menschliche Verhalten über mehrere Kanäle erkannt werden. Dadurch wird erwartet, dass multimodale Benutzeroberflächen einfacher zu lernen und zu verwenden sind, da sie von einer größeren Breite an Benutzern verwendet werden und sich vielseitigeren Bedingungen anpassen können als zuvor. Aus diesen Gründen wird ebenfalls erwartet, dass multimodale Interaktionen von Benutzern für viele Anwendungen präferiert werden [73].

Für die Entwicklung eines multimodalen Systems reicht die Intuition bezüglich dem Design allein nicht aus. Die Erkenntnisse der Kognitionswissenschaft bezüglich der koordinierten menschlichen Wahrnehmung und den natürlichen Interaktionsmustern bei kombinierten Eingabemodalitäten werden benötigt, um ein erfolgreiches multimodales System zu entwerfen. Die Idee der multimodalen Interaktion erzeugt gewisse Erwartungen bezüglich des Umgangs der Benutzer mit den Systemen. Zehn solcher Erwartungen werden in der Publikation von Oviatt aufgeführt und in einer korrigierten Form mit zentralen Ergebnissen wiedergegeben [72]. Die in diesem Rahmen interessantesten drei der 10 Mythen sind:

1. Benutzer interagieren mit multimodalen Systemen immer multimodal.
2. Multimodale Eingaben bestehen immer aus parallel ablaufenden Signalen.
3. Der Hauptvorteil multimodaler Systeme liegt in der Steigerung der Effizienz.

Obwohl Benutzer es präferieren multimodal zu interagieren, hängt die Interaktion stark von dem Aufgabenkontext und der Umgebung ab, wodurch es nicht immer zwangsläufig zu einer multimodalen Interaktion kommt. Entgegen der Erwartung von Mythos 1, interagieren Benutzer aus einer Mischung von multimodal und unimodal. Eine Studie zeigte, dass Benutzer nur 20% der Zeit wirklich multimodal interagieren [74].

Wider Erwarten von Mythos 2, verlaufen Eingabesignale meist eher sequentiell als parallel. Nimmt man das Beispiel einer Kombination von Handgeste und Sprache, z.B. „dieser

Würfel“, so ist es nicht unüblich, dass der Benutzer zwar Würfel sagt, aber den ersten Teil nur zeigt und wenn doch, dann wird der gesprochene Teil nicht parallel zum Zeigen ausgeführt. Dadurch darf beim Entwickeln von multimodalen Systemen nicht damit gerechnet werden, dass beide Signale parallel erkannt werden sollen.

Abschließend wird im Fall von Mythos 3 gezeigt, dass die Effizienz sich nicht zwangsläufig verbessert. Dies hängt stark vom Aufgabenkontext ab. Der eigentliche Hauptvorteil multimodaler Systeme liegt darin, bei erfolgreicher Integration aufgabenabhängige Fehler zu vermindern und das System robuster zu machen. Des Weiteren ist ein signifikanter Vorteil die Flexibilität für eine Vielzahl von Benutzern, Aufgaben und Umgebungen [72].

Wie zuvor gezeigt, erfordert der Entwurf von multimodalen Interfaces viel Überlegung, damit die Vorteile solcher Systeme wirklich genutzt werden können. Reeves et al. veröffentlichten Richtlinien zum Entwerfen von multimodalen Benutzeroberflächen [84]. Die wichtigsten Richtlinien lauten:

- Multimodale Systeme sollen so entworfen werden, dass sie eine Vielzahl an Benutzern und Anwendungskontexten sowie Umgebungen abdecken.
- Privatsphäre und Sicherheit muss gut durchdacht sein. Sprache sollte nicht als Modalität verwendet werden, wenn vertrauliche und persönliche Informationen an die Öffentlichkeit gelangen können.
- Die Modalitäten sollen auf eine Weise integriert werden, welche den Fähigkeiten der Modalitäten und der Präferenzen des Benutzers entsprechen.
- Multimodale Systeme sollen sich verschiedenen Benutzerprofilen, Umgebungen, Kontexten und Anwendungsbedürfnissen anpassen können.
- Ergänzende Modalitäten sollen so integriert werden, dass sie ein robusteres System ermöglichen und dem Benutzer die Möglichkeit geben aus den Modalitäten auszuwählen, um Fehler zu vermeiden.

2.2 Augmented Reality

Dieser Abschnitt erläutert zuerst den Begriff der **AR** und beschreibt die Einordnung von **AR** innerhalb des Realitäts-Virtualitäts-Kontinuums von Milgram. Danach werden mehrere Anwendungsgebiete von **AR** beleuchtet. Abschließend wird die **HoloLens** vorgestellt und diverse Publikationen präsentiert, die zum einen den Einsatz innerhalb des Operationssaals beurteilen und zum anderen mögliche medizinische Anwendungen beschreiben.

2.2.1 Definition

Augmented (*engl.: erweitert, angereichert*) Reality wird als eine „erweiterte“ Realität übersetzt. Das heißt, dass die Realität mit computergenerierten Inhalten überlagert wird, wodurch der Informationsgehalt oder die Interaktion für den Benutzer verbessert werden soll. Eine Überlagerung kann zum Beispiel eine Projektion von Inhalten auf Objekte oder

die computergenerierte Annotation von Video- oder Live-Aufnahmen sein [79]. Milgram & Koshino präsentierten in ihrer Arbeit das sogenannte „*virtuality continuum*“, welches als allgemein anerkannte Taxonomie gilt und es ermöglicht, die verschiedenen Arten von Virtualität einzuordnen [66].

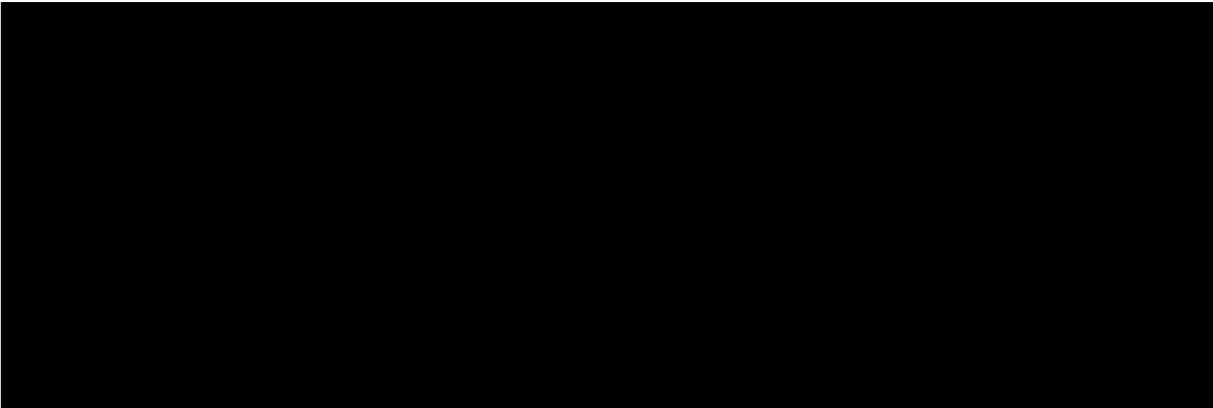


Abbildung 2.7: Das Realitäts-Virtualitäts-Kontinuum von Milgram & Koshino. Entnommen aus [7]

Abbildung 2.7 zeigt das Realitäts-Virtualitäts-Kontinuum mit entsprechenden Einordnungen verschiedener Virtualitäten. An den Enden des Strahls befinden sich die Extrempunkte der ausschließlich realen und virtuellen Umgebung. Der Strahl ohne die Endpunkte wird als MR bezeichnet und beinhaltet beliebige Zwischenstufen zwischen der Virtualität und der normalen realen Welt [79]. AR wird weiter links angeordnet, da trotz der virtuellen Inhalte der Benutzer weiterhin überwiegend die reale Umgebung wahrnimmt. Im Gegensatz dazu ist das Verhältnis bei Augmented Reality genau umgekehrt. Hier wird eine virtuelle Umgebung durch reale Inhalte angereichert, beispielsweise durch das Anzeigen der eigenen Hände in einer virtuellen Umgebung. Der Anteil an Realität nimmt von links nach rechts kontinuierlich ab, wobei der Anteil an Virtualität kontinuierlich zunimmt.

2.2.2 Formen und Anwendungsgebiete von AR

Der Einsatz von AR ist vielfältig und wird in zahlreichen Anwendungsgebieten erforscht und bereits eingesetzt. Dabei variieren die Endgeräte je nach Anwendung. Smartphones, Tablets, HMDs bis hin zu Projektionen [20]. Es werden drei Arten von AR unterschieden:

- **Video See-Through-AR**

Die reale Welt wird mit Hilfe einer Videokamera erfasst und das Videobild wird mit virtuellen Inhalten perspektivisch korrekt überlagert. Anschließend wird das Ergebnis auf dem Ausgabegerät angezeigt [30].

- **Optische See-Through-AR**

Eine Videoaufnahme ist hier nicht zwingend nötig. Die virtuellen Inhalte werden über ein semitransparentes Display angezeigt, sodass die reale Umgebung immer direkt wahrgenommen wird [30].

- **Projektionsbasierte AR**

Hierbei werden die virtuellen Inhalte auf Gegenstände der realen Umgebung projiziert [30].

Je nach Anwendungsgebiet, eignet sich eine der Formen von AR besser als die andere. Im Folgenden werden Bereiche mit entsprechenden Anwendungsbeispielen kurz vorgestellt.

Bildung

Durch die Möglichkeit des interaktiven Erlebens werden mit Hilfe von AR komplexe Lerninhalte besser vermittelt. Behmke et al. entwickelten eine AR Applikation, die es Schülern ermöglicht 2D Strukturformeln als 3D AR Objekt anzuzeigen, wodurch das chemische Verständnis der Schüler verbessert werden soll. Eine Studie zeigt dass, Schüler, die eine niedrige bis durchschnittliche Leistung haben, am meisten von einem AR unterstütztem Lernsystem profitieren [57].

Tourismus und Navigation

In Museen können AR Annotationen verwendet werden, um Exponate zu erläutern. Auch zusätzliche Visualisierungen können angezeigt werden, die das Erlebnis des Besuchers verbessern. Han et al. nutzten AR um ein Exponat eines Grabes mit zusätzlichen AR Relikten anzureichern, indem sie diese perspektivisch korrekt in die Grabstätte visualisieren [39]. Ähnlich hierzu können bei Navigationsanwendungen Texte, Bilder und graphische Symbole einblendet werden, sobald eine gewisse Orientierung des Benutzers bzw. des AR Geräts registriert wird. Hierfür werden in den meisten Fällen Tablets oder Smartphones verwendet.

Industrie

Arbeiter können mit AR Anwendungen bei der Ausführung von Arbeitsschritten in der Wartung und im Training durch das Einblenden von Hinweisen unterstützt werden [30]. Paelke entwickelte ein AR System, in dem räumlich registrierte Informationen in das Blickfeld des Benutzers eingeblendet werden [75]. Dadurch können die Arbeiter bei der Zusammensetzung bzw. Montage von unvertrauten Produkten unterstützt werden. Mithilfe von projektionsbasierter AR können in der Entwicklung und dem Design von Produktprototypen Kosten und Zeitaufwand reduziert werden. Menk et al. verwendeten projektionsbasierte AR im Design von Autos in der Industrie. Dabei wird auf ein Modellauto, bestehend aus einem kostengünstigen diffusen Material, ein 3D generiertes Bild lagerichtig projiziert [63]. Hierdurch wird ein möglichst realistischer Eindruck vom finalen Produkt gewährleistet.

Medizin

Auch in der Medizin besteht großes Interesse an der Verwendung von AR Technologien. Unter anderem befasst sich die Forschung mit der Entwicklung von AR Anwendungen für das Ausbilden von angehenden Chirurgen [18, 43]. Ein weiteres Gebiet der Forschung ist der Einsatz von AR als intraoperative Unterstützung [21, 59, 93]. Gerade bei minimalinvasiven Eingriffen sind Chirurgen auf zusätzliche Navigationssysteme und Visualisierungen angewiesen. Diese werden zurzeit auf separaten Bildschirmen angezeigt. Dadurch muss der Chirurg seinen Blick vom Patienten abwenden und die angezeigten Informationen mental vereinen, um sie in einen räumlichen Kontext zu setzen [86]. Mit Hilfe von AR soll diese kognitive Lücke überwunden, die Genauigkeit des Eingriffs und die Genesungszeit des Patienten verbessert, und die damit verbundenen Kosten gesenkt werden [20]. Ein Beispiel hierfür ist die projektionsbasierte AR. Sie kann für eine laparoskopische Operation,

ein minimalinvasiver Eingriff mit Hilfe von optischen Instrumenten innerhalb der Bauchhöhle, verwendet werden. Hierfür werden vorher angefertigte Computertomographie (CT) Aufnahmen auf den Patienten projiziert. Die Darstellung kann bearbeitet werden, sodass verschiedene wichtige Strukturen wie Knochen, Blutgefäße oder Organe angezeigt werden können [32]. Dadurch wird der Chirurg beim Finden der optimalen Position für die Platzierung des laparoskopischen Instruments unterstützt. Ein weiteres Beispiel ist das von Kersten-Oertel et al. entwickelte AR System für neurovaskuläre Eingriffe im Gehirn. Eine externe Kamera zeichnet intraoperative Aufnahmen des Patienten auf und überlagert diese lagerichtig mit präoperativ erstellten 3D Gefäßstrukturen des Patienten [53].

2.2.3 HoloLens im Operationsaal

Während einer bildgestützten Operation befinden sich die Monitore, welche die Bilddaten und Pläne präsentieren, nicht immer im Sichtbereich des Chirurgen. Die Positionierung der Bildschirme ist aufgrund der Bewahrung von sterilen Bedingungen, der Monitorhalterung und den räumlichen Einschränkungen durch die Instrumente oder des Ärzteteams, limitiert [80]. Dadurch werden die Daten häufig außerhalb des Sichtbereichs der Intervention und des Chirurgen angezeigt, was sich negativ auf seine Hand-Augen-Koordination auswirken kann [86]. Dieses Problem soll durch den Einsatz von HMDs behoben werden, indem die präoperativen Datensätze, Pläne oder Patientenmodelle mittels MR innerhalb des Sichtfeldes des Chirurgen immersiv visualisiert werden. Hierdurch können sich die Chirurgen frei im Raum bewegen und mit den virtuellen Objekten interagieren, wodurch zusätzlich das Verständnis von komplexen Daten verbessert werden soll [67]. Die HoloLens (Abbildung 2.8) vereinfacht die Entwicklung von MR Anwendungen und hat bereits eine stetig wachsende Community, welche eine Vielfalt an Anwendungen entwickelt [67].

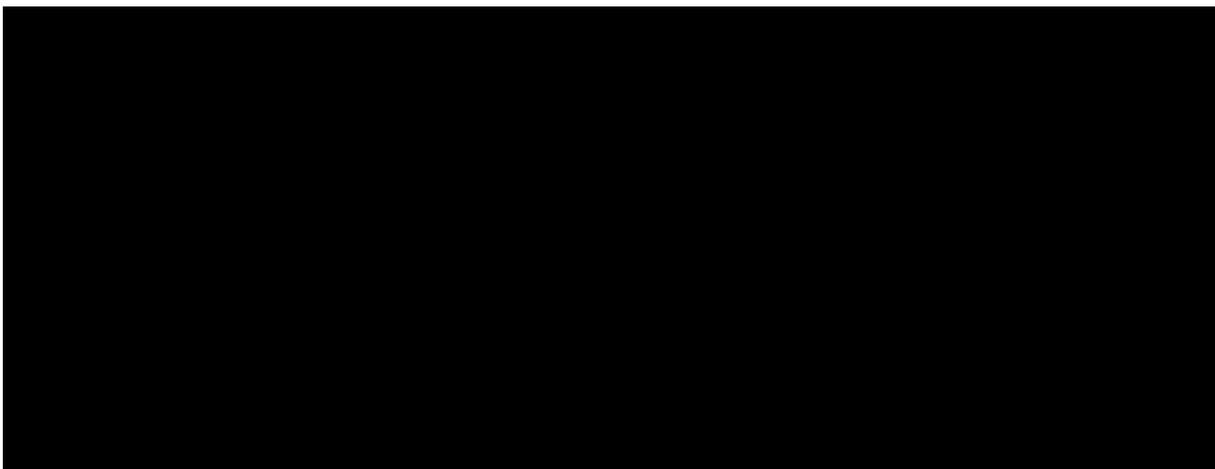


Abbildung 2.8: Komponenten der HoloLens. 1) Rechneinheit. 2) Umgebungserkennungskameras. 3) Infrarot Laser Projektor. 4) Tiefenkamera. 5) HD Videokamera. 6) Umgebungslichtsensor. 7) Holographischer Projektor. 8) See-Through Waveguide Linsen. 9) Mikrophon. 10) Eingebaute Lautsprecher. 11) Batterie. Entnommen aus [58].

Des Weiteren ist sie das erste und derzeit beliebteste Produkt auf dem Markt, welches als ein eigenständiges Gerät Handgesten- und Sprachinteraktion sowie Head-Tracking ermöglicht [90]. Ein Vergleich zwischen der HoloLens und der Meta 2 zeigt, dass die HoloLens wichtige Vorzüge bezüglich des räumlichen Trackings und der Handgestenerkennung hat und sich besser in eine operative Umgebung integrieren lässt [69].

Im Jahr 2016 wurde von Microsoft die HoloLens herausgebracht. Die HoloLens ist ein stereoskopisches MR HMD, welches es ermöglicht hochauflösende 3D Projektionen (Hologramme) in der realen Umgebung darzustellen. Bei HMDs handelt es sich um ein mobiles Visualisierungs- und Interaktionssystem, welches in Form eines Helms oder einer Datenbrille am Kopf angebracht wird. Es besteht aus einem miniaturisierten Display, einer HMD-Elektronik, welche die verschiedenen integrierten Sensoren ansteuert und einer Schnittstelle, welche das HMD mit einem Rechner, im Fall der HoloLens, über WiFi verbindet [30].

Eine Untersuchung möglicher Anwendungen der HoloLens in einer chirurgischen Umgebung wurden von Gaqsues et al. mit der Hilfe von zwei Chirurgen durchgeführt [35]. Durch mehrere Interviews und Brainstormings in einem Zeitraum von 6 Monaten wurden mögliche Anwendungsfälle und Anforderung ausgearbeitet. Neben der Echtzeitvisualisierung von Bilddaten wird auch die chirurgische Fernanweisung als möglicher Anwendungsfall genannt. Hierbei können erfahrene Chirurgen, während sie sich an einem anderen Ort befinden, in einem gemeinsamen 3D Raum Annotationen einfügen und somit dem physisch präsentem Chirurgen vor Ort assistieren. Außerdem kann die HoloLens für telemedizinische Plattformen zur operativen Fernausbildung für angehende Ärzte verwendet werden [96].

Deib et al. führten eine Studie durch, in der ein Chirurg drei verschiedene minimalinvasive Wirbelsäulenoperationen an einer Lendenwirbelsäulenpuppe durchführte [29]. Die Operationen wurden einmal mit der Unterstützung von einem üblichen Angiographie Bildschirm durchgeführt und einmal nur mit der Unterstützung der HoloLens, welche die posteroanterioren und lateralen Ansichten des Lendenbereichs in das Blickfeld des Chirurgen einblendete (Abbildung 2.9).

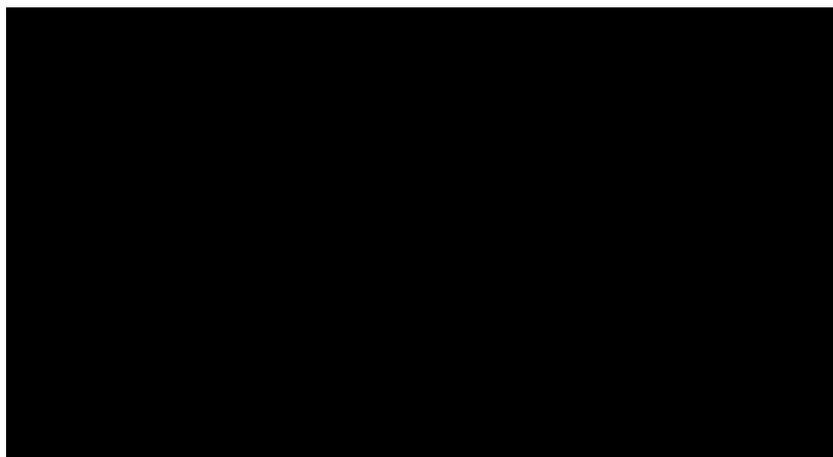


Abbildung 2.9: Eingblendete virtuelle Ansicht im Sichtbereich des Chirurgen. Entnommen aus [29].

Alle simulierten Eingriffe wurden erfolgreich durchgeführt. Die Durchführungszeiten mit der HoloLens waren ähnlich zu den typischen Durchführungszeiten für diesen Eingriff.

Die HoloLens und die konventionellen Visualisierungen waren bei der Unterstützung gleichermaßen effektiv. Sauer et al. [86] evaluierten die Anwendung der HoloLens innerhalb der Viszeralchirurgie. Dabei wurde ein 3D Modell der Lebergefäßstruktur eines Patienten während einer Operation im Sichtfeld des Chirurgen eingeblendet und evaluiert. In der Publikation wird aufgezeigt, dass die Integration der MR-Brille in den operativen Arbeitsablauf hohes Potential aufweist.

Der mögliche Einsatz der HoloLens innerhalb des Operationssaals kann einige Vorteile mit sich bringen. Die HoloLens ermöglicht eine Integrierung von verschiedensten 3D Informationen in das Sichtfeld des Benutzers [91]. Semitransparente Hologramme überdecken nur zu einem geringen Anteil die reale Szene. Dies kommt dem intraoperativen Einsatz zugute, da die zu operierende Stelle nicht überdeckt wird [96]. Die Möglichkeit, Interaktionen mit den virtuellen Daten freihändig durchzuführen ist ebenfalls ideal, um die Sterilität im Operationssaal zu bewahren. Jedoch hat der momentane Stand der Technik der HoloLens noch einige Mängel. Durch die tragbare Form und den Rechenaufwand, die die HoloLens für das Inside Out-Tracking benötigt (Positionstracking ohne externe Sensoren), bei gleichzeitig klein gehaltener Recheneinheit, kann es zu Latenzen und limitierter Auflösung der Modelle kommen [90]. Mit 597 g kann die HoloLens nach längerer Tragzeit für einige Benutzer zu schwer werden und zu Beschwerden im Nacken- oder Kopfbereich führen [28]. Des Weiteren hält die momentane Batterieleistung nur für ca. 100 Minuten bei laufender Applikation [96]. Dies ist für komplexere Operationen, bei denen sie überwiegend genutzt werden soll, zu kurz. Es kann vorkommen, dass Handbewegungen des Benutzers fälschlicherweise als die Bloom Geste erfasst werden, was zu einem Beenden der Applikation führt [96]. Die HoloLens Bloom Geste löst eine spezielle Systemaktion aus, die den Benutzer wieder zurück zum Startmenü bringt (Abbildung 2.10). Auch ist das momentane Sichtfeld der HoloLens mit ca. 35 Grad zu begrenzt [86].

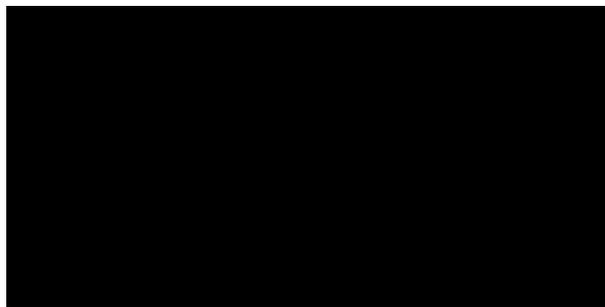


Abbildung 2.10: HoloLens Bloom Geste. Entnommen aus [4].

Zwar bietet die HoloLens einige Vorteile, indem sie es ermöglicht AR Informationen direkt in das Sichtfeld des Chirurgen einzublenden und steril damit zu interagieren, jedoch befindet sich das Gerät derzeit noch in der Anfangsphase seiner Entwicklung und ist noch nicht bereit für eine alltägliche Integrierung in den Operationssaal [69]. Nichtsdestoweniger zeigt die HoloLens hohes Potential. Zukünftige Versionen oder ähnliche Geräte mit einem größeren Sichtfeld, einer besseren Rechenleistung und besserer Ergonomie würden Anwendung finden.

2.3 Belastung im Operationssaal

Im **Unterabschnitt 2.1.2** wurde die Rollenverteilung im Operationssaal beschrieben. Jeder Anwesende trägt ein hohes Maß an Verantwortung gegenüber der Sicherheit des Patienten. Der Erfolg einer Operation hängt maßgeblich von der Leistung des Ärzteteams und vor allem des Operateurs ab. Nicht nur technische Fertigkeiten sind ausschlaggebend für die chirurgische Expertise, sondern auch kognitive Fertigkeiten, wie die Beurteilung der Situation und das Treffen von richtigen Entscheidungen [38]. Die Verantwortung und der Leistungsdruck kann unter bestimmten Umständen zu übermäßigem Stress führen. Dieser wirkt sich auf die kognitiven Prozesse, wie das Abrufen von Wissen, sowie auf die technischen Fertigkeiten aus [19, 55].

In der Publikation von Wetzell et al. wurden 16 Interviews mit Chirurgen durchgeführt. Folgende Stressfaktoren wurden dabei ausgearbeitet [97]:

1. Notfälle
2. Komplikationen
3. Fortgeschrittene Aufgaben (komplexer Eingriff, Risikopatient, Zeitdruck, Multitasking)
4. Probleme mit dem Equipment (fehlendes oder unvertrautes Equipment)
5. Teamwork Probleme (inkompetente, unerfahrene oder unaufmerksame Belegschaft, zwischenmenschliche Probleme)
6. Ablenkungen (Gesprächslärm, Kommen und Gehen von Leuten)
7. Persönliche Faktoren (emotionaler und physischer Zustand, persönliche Probleme)

Faktoren 1 und 2 wurden als sehr stressig empfunden und sind mitunter die wahrscheinlichsten Auslöser von übermäßigem Stress. Obwohl Faktor 3 eher fordernd als stressig wahrgenommen wurde, könnte eine Kombination aus Faktor 3 und weiteren Stressfaktoren ebenfalls zu sehr hohem Stress führen. Von dem gesamten Ärzteteam innerhalb des Operationssaals zeigten die primär operierenden Chirurgen die meiste Arbeitsbelastung und Stress auf [25, 98].

Moorthy et al. untersuchten den Effekt von stresserzeugenden Umständen auf die Ausführung eines simulierten laparoskopischen Eingriffs [68]. In der Aufgabenstellung sollten die Probanden ein Stück Schwamm innerhalb verschiedener Zonen ablegen. Die Aufgabe wurde zusätzlich unter vier weiteren Bedingungen getestet:

1. Hintergrund OP-Geräusche bei 80 bis 85 dB
2. Lösen von Rechenaufgaben
3. Die Aufgabe so schnell wie möglich auszuführen
4. Alle drei Stressoren vereint

Die gesamte zurückgelegte Pfadlänge der rechten Hand (alle Probanden waren Rechtshänder) war unter allen Bedingungen gleich, die Pfadlänge der linken Hand war hingegen unter den Bedingungen 2 und 4 signifikant länger. Eine signifikant höhere Anzahl an Fehlern war unter allen vier Bedingungen zu beobachten. Vor allem unter Bedingung 4, gefolgt von 2, dann 3 und zum Schluss unter Bedingung 1. Das Ergebnis der Studie zeigt, dass Stress unter anderem einen starken Einfluss auf die Motorik und letztendlich auf die chirurgische Leistung hat.

Dies wird in einer Studie von van Galen & van Huygevoort bestätigt. In dieser Studie sollten die Probanden einen Stift auf einem Grafiktablett von einer gegebenen Startposition zu einer Zielposition führen. Dabei durfte der Stift nicht angehoben werden und sollte in die kreisförmige Zielposition geführt werden. Die Aufgabe sollte einmal unter Zeitdruck durchgeführt werden, einmal mit einer zusätzlichen Multitasking Aufgabe und zuletzt unter beiden Bedingungen gemeinsam. Das Ergebnis zeigt, dass Zeitdruck einen signifikanten Effekt auf die Anzahl der Fehltreffer hat und auch die zurückgelegte Pfadlängen deutlich länger sind [92].

2.4 Zusammenfassung

In diesem Kapitel wurde der Bereich der MCI und AR behandelt, zusätzlich wurde die Belastung innerhalb des Operationssaals und dessen Auswirkungen auf die Interaktion untersucht. Während Fußinteraktion in Form von Fußschaltern bereits Bestandteil der Operationssäle ist, sind ortsunabhängige Fußeingaben noch in der Forschungsphase. Vor allem in der Kombination mit AR gibt es noch kaum Forschungsergebnisse. Hingegen wird die Integrierung von Handgesteninteraktion innerhalb des Operationssaals umfassend behandelt. Die Erkennung der Handgesten und Trennung von beabsichtigten und unbeabsichtigten Handgesten stellen allerdings ein Problem dar. Ein weiteres Problem bleibt die Tatsache, dass die operierenden Ärzte oftmals die Hände nicht frei haben, wodurch die gewünschte Interaktion in solchen Fällen wieder von den Assistenten übernommen werden muss.

Sprachinteraktion wird durch die intuitive Verwendung als vielversprechend angesehen, allerdings ist die Spracherkennung im Operationssaal aus mehreren Gründen problematisch. Hintergrundgeräusche erschweren die Erkennung, sie sind im Operationssaal jedoch aufgrund diverser Gerätschaften unvermeidbar. Ein weiteres Problem ist die Unterscheidung zwischen normaler Kommunikation und der gewollten Sprachinteraktion sowie dessen Erkennung. Aktuelle Untersuchungen innerhalb des Operationssaals mit neueren Spracherkennungssystemen wurden derzeit noch nicht durchgeführt.

Inertialsensoren können an verschiedenen Körperstellen angebracht werden und sind auf keine direkte Sichtverbindung oder Orientierung und Position des Benutzers angewiesen. Bis jetzt wurden im intraoperativen Szenario Inertialsensoren meist in Verbindung mit Handgesten [22, 44] verwendet. Erst mit der Publikation von Hatscher et al. finden Fußgesten auf Basis von Inertialsensoren Anwendung [41].

Eye- und Head-Tracking sind besonders gut für das Zeigen geeignet, vor allem bei der Nutzung von HMDs. Für die Interaktionen, wie Selektion oder Manipulation, sollte Tracking in Kombination mit anderen Methoden eingesetzt werden. Wichtige Entwicklungsaspekte,

welche bei multimodalen Systemen beachtet werden sollten, wurden im Unterabschnitt „Multimodale Interaktion“ beschrieben.

Der Einsatz von AR innerhalb des Operationssaals wird bereits umfassend behandelt. Mit dem Erscheinen der HoloLens wuchs das Interesse der intraoperativen Verwendung von optischen See-Through HMDs. In diesem Kapitel wurden einige Anwendungen hierfür vorgestellt. Zwar weist die HoloLens durchaus hohes Potential auf, jedoch bedarf sie noch weiterer Entwicklung um ein fester Bestandteil des Operationssaals zu werden. Abschließend wurde die Belastung im Operationssaal erörtert und zwei Studien aufgeführt, welche die Auswirkungen von Stress und weiteren Störfaktoren auf die Interaktion thematisieren.

3 Konzeption und Umsetzung

Um die in [Kapitel 1](#) aufgeführten Fragestellungen zu untersuchen, wurden Prototypen entwickelt. Diese werden verwendet, um die Interaktionstechniken zu evaluieren. Dabei liegt der Fokus dieses Kapitels auf der Konzeption der Prototypen und den Interaktionstechniken, d.h. auf welche Art interagiert wird, sowie der Integrierung der Interaktionstechnologien (Head-Tracking, Spracherkennungssystem und Inertialsensor).

Im ersten Abschnitt wird eine Argumentation angeführt, welche die jeweiligen Gründe für die Auswahl von Handgesteninteraktion, Head-Tracking, Sprachinteraktion und Fußinteraktion beschreibt, in Anlehnung an [Abschnitt 2.1](#). Im zweiten und dritten Abschnitt dieses Kapitels werden die Module des Prototypen vorgestellt und es wird auf die Implementierung der jeweiligen Interaktionstechniken eingegangen.

3.1 Verwendete Interaktionstechnologien

Bei der Auswahl der Interaktionstechnologien wurde sich auf die Literaturrecherche aus [Abschnitt 2.1](#) gestützt. Dort wurden diverse berührungslose Interaktionstechnologien vorgestellt sowie deren Vor- und Nachteile in Bezug auf den Einsatz im Operationssaal beleuchtet. Manche Interaktionstechnologien werden zu einem gewissen Grad bereits eingesetzt, z.B. die Fußinteraktion in Form von Fußschaltern, oder intensiv erforscht, wie im Fall von Handgesteninteraktion. Andere wurden für den Einsatz im Operationssaal weniger berücksichtigt, z.B. Inertialsensoren. Dieser Abschnitt erläutert jeweils den Hintergrund für die ausgewählten Technologien der entwickelten Prototypen und begründet, warum diese sich für die Interaktion mit der HoloLens eignen.

3.1.1 Head-Tracking

Für berührungslose und freihändige Interaktionen ist die naheliegendste Steuerung eines Mauszeigers über Eye- und/oder Head-Tracking. Im Rahmen dieser Arbeit wurde Head-Tracking ausgewählt und wird in Verbindung mit weiteren Modalitäten verwendet. Je nach zusätzlicher Modalität wird das Head-Tracking entweder nur für die Steuerung des Mauszeigers verwendet oder ist selbst Teil einer Interaktionstechnik. Wird die Selektion betrachtet, soll durch die zusätzliche Modalität zum einen das in [Unterabschnitt 2.1.7](#) beschriebene Midas Touch Problem umgangen werden, zum anderen sollen längere Selektionszeiten wie bei der *Dwell*-Methode vermieden werden. Laut der Publikation von Blattgerste et al. [24] ist in AR Umgebungen Eye-Tracking dem Head-Tracking deutlich überlegen, jedoch basiert dieses Resultat auf großen Sichtfeldern. Die HoloLens verfügt zurzeit nur über ein sehr schmales Sichtfeld, weshalb die Vorzüge des Eye-Trackings hierbei

nicht zur Geltung kommen würden. Durch das begrenzte Sichtfeld dürfen die Hologramme, wenn sie vollständig angezeigt werden sollen, nicht zu groß sein. In [Abbildung 3.1](#) ist ein Beispiel gezeigt, in dem das Hologramm durch eine geringe Senkung des Kopfes bereits nicht vollständig angezeigt wird.

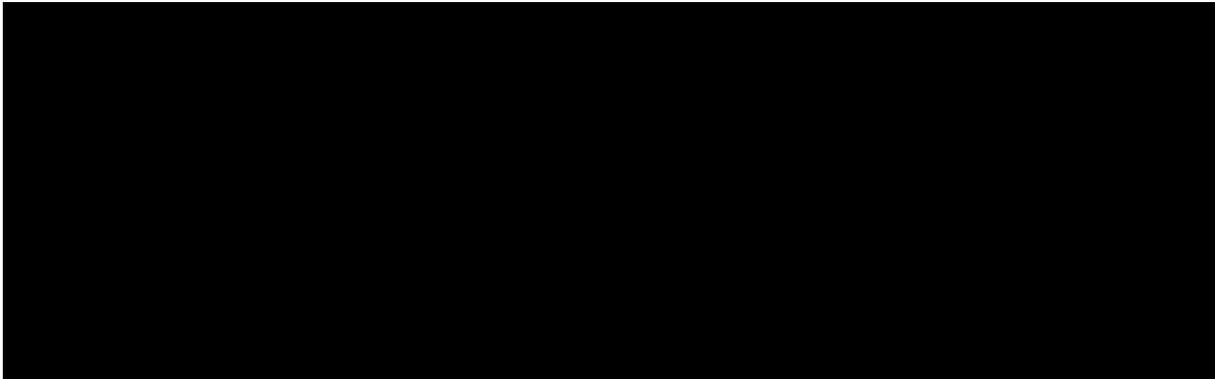


Abbildung 3.1: Beispiel eines Hologramms. Links ist das komplette Hologramm zu sehen, rechts wurde der Kopf etwas nach unten gesenkt. Entnommen aus [5].

Wie in [Unterabschnitt 2.1.7](#) beschrieben ist die Selektion kleiner Objekte über Eye-Tracking weniger treffsicher, da es unter anderem zu Abweichungen durch *Jitter* kommt. Hierbei handelt es sich um Sakkaden. Da die Hologramme in der Regel kleiner ausfallen, muss in der Theorie der Benutzer keine umfangreichen Kopfrotationen vornehmen, wodurch die Interaktion nicht ermüdend sein sollte. Des Weiteren wurde gezeigt, dass bei kleinen Distanzen Head-Tracking überlegen ist [81]. Ein weiterer, jedoch nicht ausschlaggebender, Grund für Head-Tracking ist, dass die HoloLens bereits über Head-Tracking verfügt und keine weiteren Sensoren angebracht werden müssen. Sollte ein größeres Sichtfeld verfügbar werden, wäre eine zusätzliche Untersuchung mit Eye-Tracking ebenfalls von Interesse.

3.1.2 Fußinteraktion durch Inertialsensor

Im [Unterabschnitt 2.1.3](#) wurde aufgeführt, dass sich Füße sehr gut für Aufgaben eignen, in denen keine präzise Steuerung verlangt wird, sowie für sekundäre Aufgaben. Das Auslösen einer Selektion ist keine komplexe Aufgabe und kann dementsprechend durch eine für den Benutzer leichte Bewegung ausgeführt werden. Die benötigte präzise Steuerung des Cursors sollte allerdings über eine andere Modalität durchgeführt werden. Der Vorteil von Füßen besteht darin, dass der Chirurg die Interaktion steril selbst ausführen kann und keine Instrumente abgelegt werden müssen, wodurch der Arbeitsfluss nicht unterbrochen wird. Des Weiteren eignen sich Füße sehr gut zur kontinuierlichen Veränderung von Werten. Aus diesen Gründen besteht großes Interesse daran Fußinteraktion, nicht nur in Form von Fußschaltern, im Operationssaal zu integrieren. Die meisten vorherigen Systeme verwendeten Fußinteraktion zur Bedienung von Bildbetrachtern, welche die Patientendaten anzeigen. Mit der HoloLens wäre es möglich mit den Hologrammen über die Füße zu interagieren, indem z.B. ein Hologramm rotiert oder skaliert wird, etwas ausgewählt wird oder Schwellenwerte für Visualisierungen anpasst werden.

Bei der Technologie zur Erfassung von Fußinteraktionen wurde die intrinsische Variante mit einem Inertialsensor gewählt (Unterabschnitt 2.1.6). Grund hierzu ist, dass der Benutzer sich im Raum frei bewegen kann und somit nicht, wie im Fall eines Fußschalters, an einem Ort gebunden ist. Des Weiteren muss keine Sichtlinie, wie beim optischen Tracking, gewährleistet werden, sodass die Lage der Gerätschaften nicht beachtet werden muss.

Eine weitere Möglichkeit würde die Interaktion über die Verlagerung des Fußdrucks darstellen. Hierfür würde eine Sohle mit integrierten Drucksensoren verwendet werden, jedoch wurde sich gegen diese Variante entschieden. Der Grund hierfür ist, dass sich die Druckausübung der Benutzer unterscheidet und zuerst immer ein Kalibrierungsschritt erfolgen müsste. Darüber hinaus kann das zur unerwünschter Interaktion kommen, wenn der Benutzer beim Stehen oder genauerem Betrachten sein Gewicht verlagert.

Für die Selektion und das kontinuierliche Verändern eines Wertes zeigt [Abbildung 2.4](#) eine mögliche Übersicht. Bei der Interaktion mit den entwickelten Systemen wurde die Toe tap und die Heel rotation Fußgeste gewählt. Ein Grund für den Toe tap ist, dass die Fußgeste stark einem Finger-Tipp für Touchscreens ähnelt [95]. Ein weiterer Grund ist, dass die Toe tap Fußgeste in der Literatur am häufigsten verwendet wird [95]. Wie in [Unterabschnitt 2.1.3](#) beschrieben, eignet sich laut einer Studie der Double-tap hervorragend für die Aktivierung eines Systems. Die Annahme ist, dass sich ein einfacher Tap dementsprechend auch sehr gut für die Selektion von virtuellen Objekten eignen sollte. Heel rotation wurde in Anlehnung an die Publikation von Hatscher et al. [41] gewählt, da sich diese gegenüber anderen den Konzepten durchgesetzt hat. Des Weiteren wurde auch in anderer Literatur Heel rotation favorisiert [54, 95]. Die Fußgesten können gut über die Beschleunigung sowie Orientierung, die von den Inertialsensoren gemessen wird, erfasst werden.

Als zusätzliche Absicherung, um unerwünschte Interaktionen der Benutzer zu vermeiden, wird die Fußinteraktion multimodal mit Head-Tracking kombiniert. Somit müssen zuerst bestimmte Bereiche über die Kopfposition anvisiert werden, um dann mit den Füßen interagieren zu können.

3.1.3 Handgesteninteraktion

Der Hauptfokus dieser Arbeit liegt auf den berührungslosen und freihändigen Interaktionstechniken, jedoch ist die Interaktion über Handgesten eines der am umfassendsten in der Literatur behandelten berührungslosen Interaktionen ([Unterabschnitt 2.1.4](#)). Die Handgesteninteraktion hat in der Literatur durchwegs positive Ergebnisse geliefert und positive Beurteilungen von Chirurgen erhalten. Allerdings wurde diese Interaktionstechnik noch nie in einem Multitasking-Kontext erforscht. Es besteht ein großes Interesse zu sehen, ob die Handgesteninteraktion beim gleichzeitigen Ausführen zweier Aufgaben zu einer erhöhten kognitiven Belastung und somit zu einer Beeinträchtigung der Primäraufgabe führt. Dies ist vor allem wichtig, da Handgesten den Standard für die Interaktion mit den Hologrammen der HoloLens bilden.

3.1.4 Sprachinteraktion

Im [Unterabschnitt 2.1.5](#) wurden bereits die Vorteile von Spracherkennung genannt. Sie ist für den Menschen sehr intuitiv und im Gegensatz zu anderen Interaktionsformen in der Regel besser in einem Multitasking-Kontext vereinbar, wodurch sie sich sehr gut als multimodale Ergänzung eignet. Zwar berichteten Publikationen davon, dass die Spracherkennung noch sehr fehleranfällig ist [17, 76], jedoch handelt es um einen fortlaufenden Forschungsbereich. Mit dem Erscheinen von neuen Technologien wie den Sprachassistenten Amazon Echo, Google Assistant oder Cortana und der Verbesserung derer Spracherkennungssysteme, könnten verbesserte Erkennungsraten innerhalb eines Operationssaals erzielt werden. Hierfür müssten erneute gezielte Studien durchgeführt werden. Aus diesen Gründen wird im Rahmen dieser Arbeit die Interaktion über Sprache behandelt, da sie hohes Potential bei der berührungslosen und freihändigen Interaktion aufweist, vor allem im Multitasking-Szenarien.

Für die Sprachinteraktion wird das eingebaute Spracherkennungssystem der HoloLens verwendet, welche die Microsoft Speech API nutzt. Dabei handelt es sich um eine Programmierschnittstelle, welche dem Entwickler ermöglicht Microsofts Spracherkennung zu nutzen. Eine weitere Schnittstelle ist die Google Speech API. Die dazugehörige Google Spracherkennung hat im Vergleich eine bessere Worterkennungsrate [52]. Allerdings erweist sich die Integration der Google Spracherkennung deutlich schwieriger. Gründe hierfür sind, dass im Gegensatz zu der Microsoft Speech API, diese nicht in der HoloLens eingebaut ist. Dies bedeutet, dass ein externes Programm eine Schnittstelle zwischen der Google Speech API und der HoloLens bereitstellen müsste. Des Weiteren funktioniert die Google Speech API nur über eine Internetverbindung, da die Daten an die Cloud gesendet und dort ausgewertet werden. Momentan besteht in Krankenhäusern nicht überall eine Internetverbindung, welches die Verwendung von cloudbasierten Spracherkennungssystemen erschwert. Außerdem würde es einem Konflikt bezüglich des Datenschutzes geben. Zwar kann die Spracherkennung bei der HoloLens durch eine Internetverbindung an den Benutzer angepasst werden, jedoch funktioniert sie auch ohne Internetverbindung.

Für die Verwendung von Sprachinteraktion mit der HoloLens hat Microsoft einige Richtlinien veröffentlicht. [Tabelle 3.1](#) zeigt die Empfehlungen bezüglich der Auswahl von Sprachkommandos [15].

Do	Don't
Verwendung von prägnanten mehrsilbigen Kommandos	Einsilbige Kommandos
Verwendung von einfach zu merkenden Kommandos	Verwendung von Systembefehlen, wie "remove"
Konsistenz: Gleiche Funktionalitäten sollten gleiche Kommandos haben	Verwendung von sich reimenden oder phonetisch ähnlich klingenden Kommandos

Tabelle 3.1: Empfehlungen bezüglich der Verwendung von Sprachkommandos für die Interaktion mit der HoloLens.

Liu et al. evaluierten unter anderem die Spracherkennung der HoloLens [58]. Dafür wurden systemdefinierte Sprachbefehle und benutzerdefinierte Sprachbefehle verwendet. Die verwendeten systemdefinierten Sprachbefehle waren: „adjust“, „select“, „place“, „face me“, „bigger/smaller“, „remove“, „Hey Cortana, shut down“ und „Hey Cortana, take

a picture“. Die benutzerdefinierten Sprachkommandos waren: „move“, „rotate“, „delete“, „zoom in/out“, „open“, „duplicate“, „previous“ und „help“. Benutzerdefinierte Sprachbefehle wurden zu 74,47% erkannt, während systemdefinierte Befehle zu 66,87% erkannt wurden. Des Weiteren war die Spracherkennung davon abhängig, ob der Benutzer zwischen den Wörtern eine sehr kurze oder gar keine Pause einlegte. Beides war mit Schwierigkeiten in der Erkennungsrate verbunden.

Auch wenn die Resultate derzeit nicht vollständig überzeugen, um die HoloLens mit der Sprachinteraktion direkt im Operationssaal einzusetzen, so sollten zukünftige Versionen Verbesserungen mit sich bringen. In dieser Arbeit wird Spracherkennung multimodal mit Head-Tracking kombiniert.

3.2 Module

In diesem Abschnitt werden drei Module vorgestellt, die später für die Studie zu einem Prototypen zusammengebaut werden. Es wird auf die Integration der Interaktionen für die jeweiligen Modalitäten und auf die Implementierung der Prototypen eingegangen. Die Kernidee der drei Module besteht darin, dass das erste Modul eine abstrakte medizinische Aufgabe simuliert (Modul 1), die für den Multitasking-Kontext verwendet wird. Das zweite Modul die Selektion und dritte Modul die Manipulation eines kontinuierlichen Wertes untersucht.

3.2.1 Modul 1: Abstrakte medizinische Primäraufgabe

Während einer intraoperativen Ultraschalluntersuchung oder einer Nadelintervention muss das entsprechende Instrument sowohl gehalten als auch bewegt werden. Dies erfordert vor allem bei der Nadelintervention sehr viel Konzentration, um den richtigen Einstichwinkel beizubehalten und um die Nadel nicht zu weit in den Körper einzuführen. Durch die Atmung des Patienten bewegt sich der Körper mit. Dementsprechend muss der Operateur das Instrument richtig ausrichten, wodurch eine hin und her oder eine vor und zurück Bewegung entsteht. Interagiert der Operateur dabei mit einem System werden ihm Multitasking-Fertigkeiten abverlangt.

In diesem Sinne handelt es sich bei diesem Modul um eine abstrakte Form einer primären Aufgabe. Das Modul besteht aus einem Slider (*dt.: Schieber*) der einen kleineren grünen Bereich beinhaltet (**Abbildung 3.2**). Dieser grüne Bereich bewegt sich hin und her, entlang des roten Objekts, und stellt somit die Körperbewegung des Patienten beim Atmen dar. Der weiße Marker kann frei entlang der x-Achse des Sliders bewegt werden. Die Aufgabe besteht darin, den Marker innerhalb der grünen Markierung zu behalten, trotz deren automatischen Bewegung. Die menschliche Atmung liegt im Ruhezustand bei ungefähr 10 bis 15 Atemzügen pro Minute. Die Größe des Sliders beträgt 10,8 cm x 1,2 cm und mit der Bewegungsfrequenz von 3 mal hin und her pro Minute, entlang des Sliders, soll die menschliche Atmung nachempfunden werden. Die geringere Bewegungsfrequenz wurde gewählt, da in den Vorversuchen eine Frequenz von 6 mal hin und her als viel zu schnell wahrgenommen wurde.

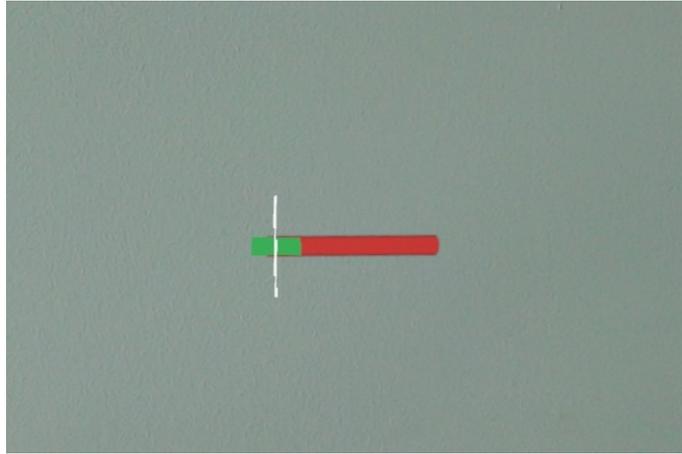


Abbildung 3.2: Modul 1. Roter Slider mit einem beweglichen grünen Bereich. Die Abbildung zeigt das Modul vor Beginn der Aufgabe.

3.2.2 Modul 2: Selektionsaufgabe

Die Selektion von Objekten, Modi oder einzelnen Einstellungen innerhalb von Patientendaten stellt eine wichtige Funktion dar und muss während einer Operation gewährleistet werden. Bei der Verwendung der HoloLens muss der Chirurg dies entsprechend auf den Hologrammen ausführen können. In diesem Zusammenhang wird die Selektion mit Hilfe des folgenden Moduls untersucht. Das Modul ist gemäß der EN ISO 9241-420:2011 konstruiert. Es handelt sich dabei um einen Tipptest in mehrere Richtungen nach dem Fitts' Gesetz. Das Gesetz beschreibt die Selektionszeit in Abhängigkeit von der Größe des Ziels und dem Abstand zwischen Cursor und Ziel. Je nach Abstand und Größe des Ziels gibt es verschiedene Schwierigkeitsstufen für die Selektion.

$$I_D = \log_2 \frac{d + w}{w} \quad (3.1)$$

mit:

$$\begin{aligned} I_D &= \text{Schwierigkeitsgrad in Bit} \\ d &= \text{Bewegungsabstand zum Zielobjekt in mm} \\ w &= \text{Breite des Zielobjekts in mm} \end{aligned}$$

Der Test sieht so aus, dass die Zielobjekte um den Umfang eines Kreises angeordnet sind. Die Bewegung des Cursors entspricht nahezu dem Durchmesser des Kreises. Das zu selektierende Zielobjekt soll hervorgehoben werden und nach jeder Auswahl wird das gegenüberliegende Zielobjekt hervorgehoben. Der Testabschnitt beginnt mit dem Anvisieren des obersten Objekts und endet mit der Auswahl des letzten Objekts.

In dem Modul sind die Zielobjekte blaue Kugeln. Es werden 25 blaue Kugeln in einem Kreis angeordnet, die zu selektierende Kugel wird in gelb angezeigt. Die Kugeln haben jeweils einen Durchmesser von 2,33 cm und der Kreis einen Radius von 23,30 cm. Dadurch ergibt sich nach Gleichung 3.1 eine Schwierigkeitsstufe von 3,4 bit. Bei der Kreis- und Kugelgröße wurde der größtmögliche Radius des Kreises und die größtmögliche Kugelgröße



Abbildung 3.3: Ablauf des Tipptests in verschiedene Richtungen nach EN ISO 9241-420:2011. Entnommen aus der deutschen Fassung von Oktober 2011.

gewählt, sodass alle Kugeln im HoloLens Sichtbereich zu sehen sind, wenn der Kopf sich in der Ausgangslage befindet. Wird die Kugel mit dem Cursor anvisiert, so färbt sie sich grün. Werden Kugeln anvisiert, die nicht ausgewählt werden sollen, färben sie sich rot (Abbildung 3.4).

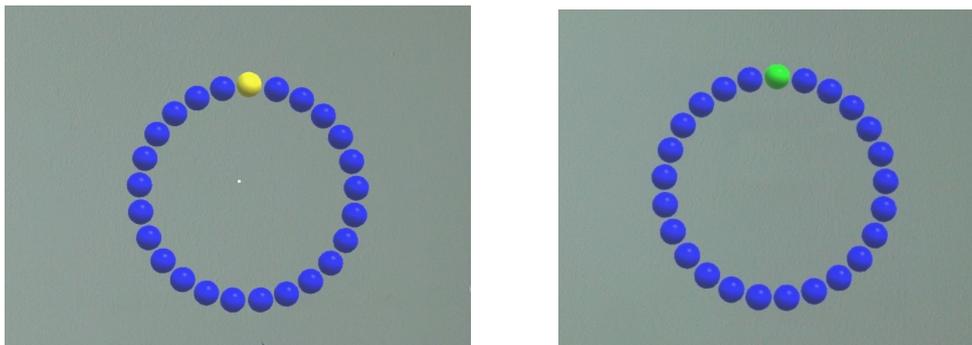


Abbildung 3.4: Modul 2. Links: Die zu selektierenden Kugel in gelb. Rechts: Die zu selektierende Kugel wird über die Kopfposition anvisiert und färbt sich grün.

3.2.3 Modul 3: Manipulationsaufgabe

Während einer bildgeführten Chirurgie ist der Zugriff auf die medizinischen Bilddaten essentiell. Die Chirurgen müssen nicht selten sich erneut ein Bild von der Situation machen, um den Fortschritt des Eingriffs besser beurteilen zu können. Aber auch die Veränderung des Bildkontrasts oder das Heranzoomen an Bildstrukturen muss gelegentlich vorgenommen werden. Deshalb soll in diesem Modul die Änderung eines kontinuierlichen Wertes simuliert werden. Dies ist in Form von Scrollen durch Bilder in einer Galerie realisiert. *Abbildung 3.5* zeigt das Modul.

Die Bildergalerie ist nach einer Drehrad Metapher entwickelt. Dies bedeutet, dass die Bilder radial angeordnet sind und das Scrollen an das Drehen eines Drehrads erinnern soll. Das Bild in der Mitte des Sichtfeldes des Betrachters ist vergrößert dargestellt, je

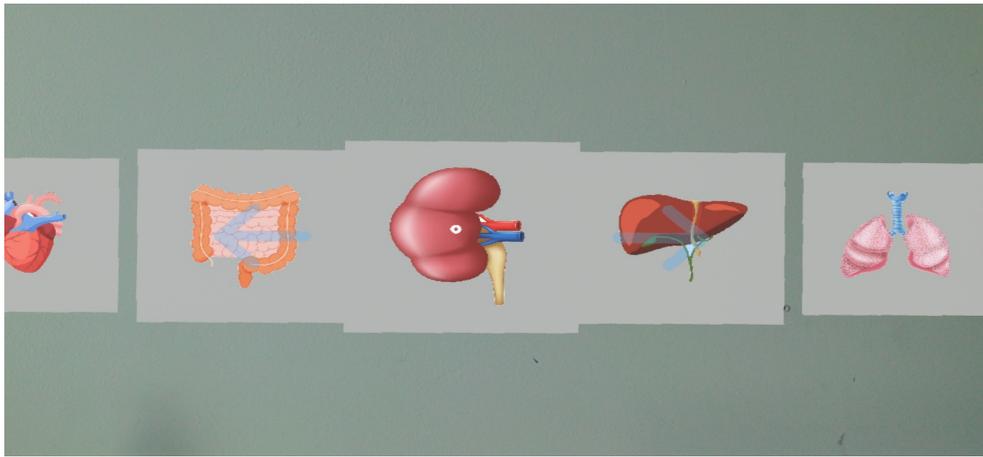


Abbildung 3.5: Modul 3. Die Bildergalerie befindet sich im nicht aktiven Scroll-Modus.

weiter links und rechts geblickt wird desto kleiner werden die Bilder dargestellt. Beim Scrollen nach links oder rechts wird das mittlere Bild in einer fließenden Animation durch das linke oder rechte Bild ersetzt. Die Richtung wird mit Hilfe von Pfeilen angezeigt. Befindet sich der Benutzer in der Scrollbewegung wird der Pfeil in die jeweilige Richtung opak blau angezeigt. Setzt ein kontinuierliches Scrollen ein, so wird der Pfeil grün dargestellt.

3.3 Implementierung

Die zuvor beschriebenen Module wurden für die Untersuchung zu drei Prototypen zusammengebaut. Der Erste besteht aus dem Modul 1 und stellt die operative Primäraufgabe, wie das Einführen einer Nadel oder der intraoperativen Ultraschalluntersuchung, dar. Hierbei wird erfasst wie genau der Prototyp ausgeführt wird.

Der zweite Prototyp wird aus dem Modul 1 und Modul 2 erstellt und soll die Selektion innerhalb der Patientenbilddaten oder Patientenmodellen, während einer Operation, simulieren und untersuchen. Hierfür wird die Selektionsaufgabe aus dem Modul 2 verwendet, in dem die Zeit für die Bearbeitung und die Genauigkeit erfasst wird. Die Durchführung findet unter zwei Bedingungen statt. Die erste untersucht nur die Selektion und die zweite untersucht die Selektion aus Modul 1 während des Ausführens der Aufgabe.

Der dritte Prototyp besteht aus Modul 1 und Modul 3 und simuliert die Manipulation von kontinuierlichen Werten, wie die Durchsuchung von Bildschichten innerhalb der Patientenbilddaten. Dafür wird die Scrollaufgabe aus Modul 3 verwendet, in der die Schnelligkeit und Genauigkeit erfasst wird. Die Durchführung findet ebenfalls unter zwei Bedingungen statt. Die Erste untersucht nur die Manipulation des kontinuierlichen Wertes und die Zweite untersucht diese, während die Aufgabe aus Modul 1 ausgeführt wird.

Im Folgenden wird detaillierter auf die Funktionsweise der einzelnen Prototypen und die Einbindung von Fuß-, Sprach- und Handgesteninteraktion sowie Head-Tracking eingegangen.

3.3.1 Programmierumgebung

Alle Prototypen wurden für die HoloLens entwickelt. Als Entwicklungsumgebung wurde Unity mit der Version 2018.1.6f1 verwendet und Visual Studio Community 2017. Bei den HoloLens Anwendungen handelt es sich um Universal Windows Plattform Apps (UWP Apps), die unter Windows 10 ausgeführt werden können. Die Verwendung von Unity ist mit dem gebotenen Support für die HoloLens begründet. Mit Hilfe von Unity und dem Mixed Reality Toolkit können Grundanwendungen schnell entwickelt und auf der HoloLens gebaut werden. Des Weiteren stellt Microsoft einige Tutorials zur Entwicklung von HoloLens Anwendungen und zur Nutzung der diversen Interaktionstechnologien zur Verfügung [2]. Die zusätzlichen Skripte für die Funktionalitäten der HoloLens wurden in C# geschrieben. Das Auslesen der Inertialsensordaten für die Fußinteraktion wurde innerhalb von Visual Studio über eine C# Windows Forms-Anwendung getätigt, die von dem Hersteller Xsens bereitgestellt wurde [6]. Die Anwendung wurde an die benötigten Funktionalitäten angepasst. Die Kommunikation zwischen der HoloLens und der Anwendung findet über das Netzwerkprotokoll UDP (User Data Protocol) statt.

3.3.2 Prototyp 1: Abstrakte medizinische Primäraufgabe

Dieser Prototyp besteht aus Modul 1 (Unterabschnitt 3.2.1). Wie beschrieben bewegt sich der grüne Bereich hin und her. Der weiße Marker, welcher innerhalb des grünen Bereichs gehalten werden soll, wird über den Thumbstick eines wireless Xbox One Controllers gesteuert. Dieser wird über die Einstellungen an die HoloLens gekoppelt. Das Erfassen der Tastensignale wird über ein Skript getätigt, welches innerhalb der Update Funktion in jedem Frame nach ausgelösten Tastensignalen abfragt. Das Skript selbst verwendet ein Skript aus dem Asset Store [11]. Für die Implementierung dieses Prototypen wurde ein übliches UI Slider Element im Unity Editor verwendet. Der ursprüngliche Griff des Sliders (*engl.: handle*) wird durch einen grünen Balken ersetzt. Dieser wird durch ein Skript, ab dessen Start, alle 100 ms um einen Wert verschoben. Der Wertebereich des Sliders umfasst dabei 0 bis 100. Durch die Collider (Box Collider) des Sliders wird gewährleistet, dass dieser innerhalb des roten Bereichs bleibt. Collider in Unity sind Komponenten, welche die gleiche Form des dazugehörigen Objektes besitzen und die Form des Objekts bei einer Kollision definieren. Der Ein- und Austritt des Markers innerhalb des grünen Bereichs wird ebenfalls registriert. Die Zeitpunkte der besagten Events werden nach Beendigung des Prototypen in eine CSV Datei gespeichert.

3.3.3 Prototyp 2: Selektionsaufgabe

Für diesen Prototyp wird das Modul 2 (Unterabschnitt 3.2.2) verwendet. Die 25 blauen Kugeln werden in einem Kreis angeordnet. Es wird eine Liste für die Reihenfolge der hervorzuhebenden Kugeln erstellt. Wird die letzte Kugel aus der Liste selektiert, wird das gesamte Objekt ausgeblendet, indem es deaktiviert wird. Die Modalitäten Sprach-, Handgesten- und Fußinteraktion werden mit Hilfe des Xbox One Controllers gewechselt. Dabei hat jede Modalität eine bestimmte Tastenbelegung, die dann innerhalb eines Scripts erfasst wird und die dann die benötigten Scripte für die gewünschte Modalität aktiviert und die nicht benötigten Scripte deaktiviert.

Head-Tracking

Bei allen drei Interaktionstechniken wird das Head-Tracking der HoloLens für das Steuern des Cursors verwendet. Es wird innerhalb der Prototypen der *DefaultCursor* aus dem Mixed Reality Toolkit verwendet. Dabei handelt es sich um einen weißen Punkt, der bei der Fokussierung von Objekten zu einem Torus wird ([Abbildung 3.8](#)). Der Benutzer bewegt den Cursor über seine Kopfposition auf das zu selektierende Objekt. Die Auswahl selbst findet über eine andere Modalität statt. Das Anvisieren wird mit Hilfe von Raycasting und den Collidern der Kugeln (*Sphere Collider*) getätigt. Beim Raycast wird ein Strahl (*engl.: ray*) vom Kopf des Nutzers bzw. vom Headset aus in die Vorwärtsrichtung projiziert und es wird überprüft, ob Kollisionen von dem Strahl mit den Collidern der Objekte vorliegen. Durch das Interface *IFocusable* des Mixed Reality Toolkits wird das Ein- sowie Austreten des Cursors in und aus den Collidern registriert. Wird eine blaue Kugel anvisiert färbt sie sich rot. Handelt es sich um die richtige Kugel in gelb wird sie grün gefärbt.

Fußinteraktion

Für die Erfassung von Fußbewegungen wird von Xsens der MTw Avinda wireless Inertialsensor verwendet. Dieser wird auf der Oberseite eines OP-Schuh angebracht ([Abbildung 3.6](#)).



Abbildung 3.6: OP-Schuh mit angebrachtem Inertialsensor.

Die Auswahl der Kugeln erfolgt nach dem Anvisieren über ein Toe tap. Die Fußgeste ist in [Abbildung 2.4](#) zu sehen. Für das Auslesen der Sensordaten wird das offizielle Xsens Beispiel Programm als Grundlage verwendet. Das Programm wurde für die Erkennung der Fußgeste und die Kommunikation mit der HoloLens erweitert.

[Abbildung 3.7](#) zeigt das veränderte Beispiel Programm aus dem MT Software Suite [6]. Über die Taste „Scan“ werden die bereitstehenden Sensoren ermittelt, mit „Enable Radio“ wird die Übertragung eingeschaltet und mit „Measure“ wird die Messung gestartet. Ist die Checkbox „Training“ nicht ausgewählt, werden die Daten an die HoloLens gesendet, ansonsten werden sie an den lokalen Rechner übertragen. Ist die Checkbox „Rotate“ ausgewählt wird die Fußgeste Heel rotation verwendet, andernfalls wird der Toe tap versucht erkannt zu werden. Die Übertragung der erkannten Fußgeste wird in einer Nachricht über UDP an den gewünschten Empfänger gesendet. Der Prototyp selbst enthält ein Skript, welches die Nachrichten empfängt. Wird eine Nachricht mit dem String „TAP“ gelesen,

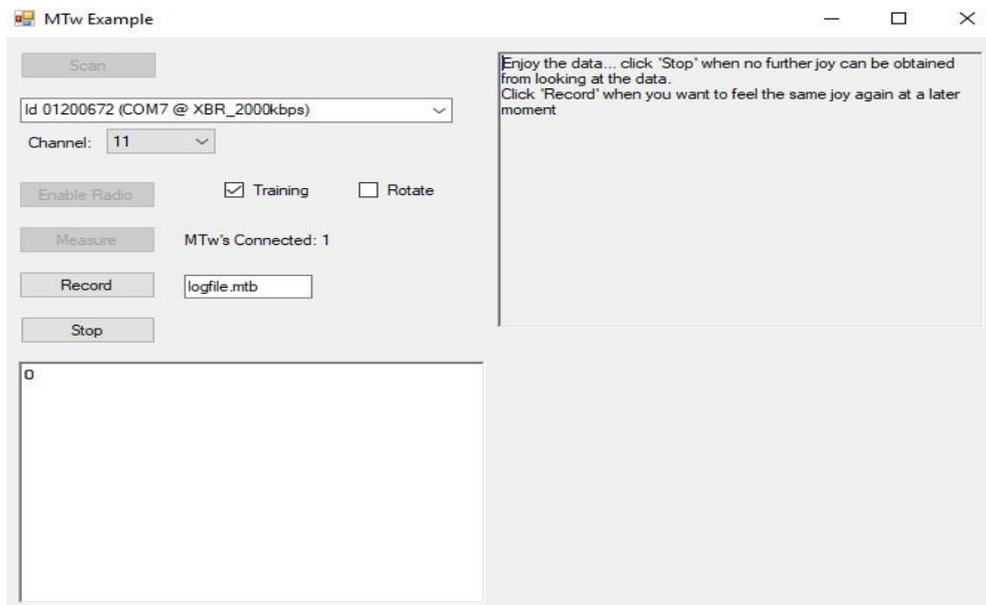


Abbildung 3.7: Ausschnitt aus der C# Windows Forms-Anwendung der MT Software Suite [6].

wird eine Funktion aufgerufen die überprüft, ob die richtige Kugel anvisiert ist und anschließend, in dem Fall, die Kugel auswählt.

Die Toe tap Erkennung besteht aus den drei Events: down, up und rest. Für die Auswertung werden die Orientierungsdaten des Sensors verwendet. Das rest Event umfasst einen nicht-reaktiven Bereich in dem der Benutzer geringe Bewegungen mit seinem Fuß ausführen kann, ohne dabei ein anderes Event auszulösen. Dieser Bereich umfasst insgesamt einen Wert von $2,0^\circ$. Dies bedeutet, dass die aufeinanderfolgenden Werte sich nicht um mehr als $1,0^\circ$ nach oben und unten unterscheiden dürfen, um weiterhin im rest Bereich zu bleiben. Unterscheidet sich der darauffolgende Wert um mehr als $1,0^\circ$ nach oben, wird ein up Event registriert. Unterscheidet sich der Wert nach unten, wird ein down Event registriert. Für das Auslösen eines Taps muss der Tap in einem gewissen Zeitfenster ausgeführt werden. Dabei registriert das Programm ein Heben und Senken des Fußes. Erfolgt innerhalb von $0,8s$ ein Richtungswechsel von up zu down einschließlich eines rest Events, wird ein Toe tap erkannt.

Die Werte wurden durch Vorversuche in einem Hallway-Testing [3] ermittelt. Es handelt sich hierbei um eine Testform bei der zufällige Probanden das System testen. Es wurden verschiedene Einstellungen mit den Zeitfenstern zwischen $0,5$ und $1,5s$ getestet. Das Zeitfenster von $0,8s$ stellte den besten Kompromiss zwischen der Höhe des anzuhebenden Fußes und der Vermeidung einer unbeabsichtigten Erkennung der Geste dar. Damit kein Tap beim normalen Gehen erkannt wird, dürfen nicht zu viele downs hintereinander erfolgen, da dies auf Gehen hindeuten würde. Bevor die Geste angesetzt wird muss vorher eine aktive Vorbereitungsphase erkannt werden. Dies bedeutet, dass vier Frames lang ein rest Event erfolgen muss, damit ein Toe tap erkannt werden kann. Diese Lösung ist robust beim kontinuierlichen Gehen. Bei sehr langsamen Gehen oder anderen Bewegungen, die das Heben und Senken eines Fußes beinhalten, könnten dennoch unbeabsichtigte Taps erkannt werden. Aus diesem Grund ist eine zusätzliche Absicherung notwendig.

Der Prototyp wird in zwei Varianten durchgeführt. Wie [Abbildung 3.8](#) zeigt, wird einmal nur die Selektionsaufgabe durchgeführt und einmal die Selektionsaufgabe parallel zu der Primäraufgabe.

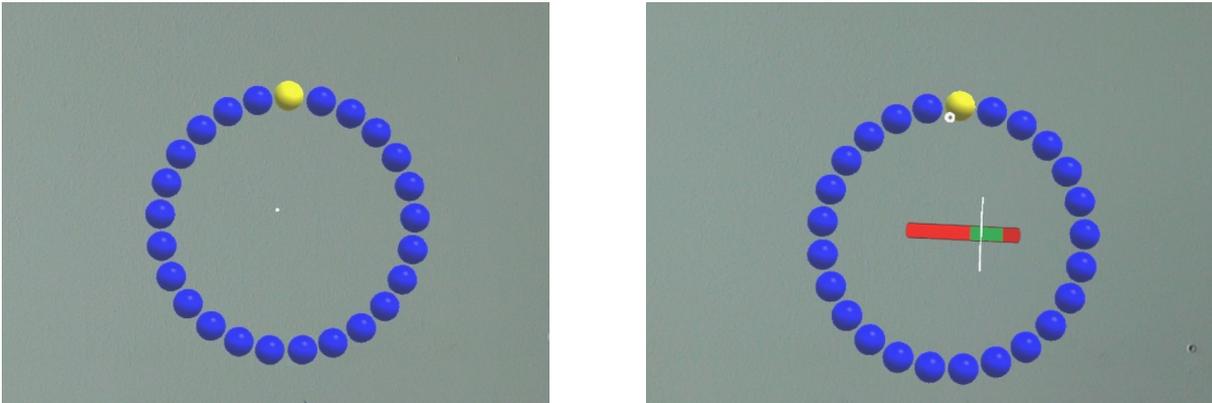


Abbildung 3.8: Selektionsprototyp. Links: Selektion, rechts: Selektion + Primäraufgabe

Handgesteninteraktion

Bei der Interaktion über Handgesten wird für die Selektion eine Handgeste aus dem Standardkatalog der HoloLens verwendet. Für die Selektion wird, ähnlich zu einem Mausklick, der sogenannte *Air tap* verwendet ([Abbildung 3.9](#)). Diese Handgeste wird auch generell bei der Nutzung der HoloLens für die Auswahl von UI Elementen verwendet.

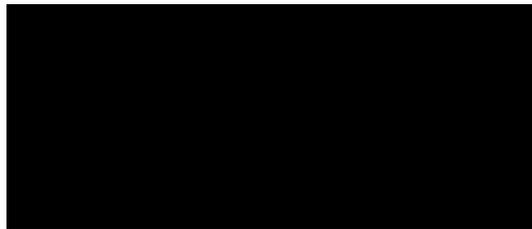


Abbildung 3.9: HoloLens Air tap Geste: zuerst werden die Finger in die vorbereitende Position gebracht (links), anschließend folgt der Tap, indem die Finger zusammengepresst werden. Entnommen aus [14].

Die Gestenerkennung wird über das Interface *IInputClickHandler* aus dem Mixed Reality Toolkit gewährleistet. Wird die hervorgehobene Kugel anvisiert und die Handgeste ausgeführt, wird die Kugel selektiert. Danach wird erneut die Funktion aufgerufen, welche die gegenüberliegende Kugel hervorhebt. Wichtig ist, dass die Handgeste innerhalb des sichtbaren Bereiches der HoloLens ausgeführt wird ([Abbildung 3.10](#)), sonst wird sie von den Sensoren nicht erfasst.



Abbildung 3.10: Der sichtbare Bereich der HoloLens in welchem Gesten erkannt werden. Entnommen aus [12].

Sprachinteraktion

Für die Auswahl über Sprache werden Sprachkommandos verwendet. Das Mixed Reality Toolkit enthält bereits die Skripte *Speech Input Source* und *Speech Input Handler*. In denen können die jeweiligen Sprachkommandos angelegt werden und die gewünschten Funktionen, nach dem Erkennen der Sprachkommandos, aufgerufen werden. Die Skripte verwenden für die Spracherkennung die *Windows Speech Recognition*. Damit die Spracherkennung nicht an der Aussprache der Kommandos scheitert, wurden mehrere Sprachkommandos angelegt von denen sich der Benutzer eins aussuchen kann:

- check item
- okay
- pick out
- select

Bei der Auswahl der Sprachkommandos wurde darauf geachtet, dass deren Funktion wiederspiegelt wird und sie leicht zum Merken sind. Um die Microsoft Richtlinien aus [Tabelle 3.1](#) zu beachten, wurden mehrsilbige Kommandos hinzugenommen.

Wird die hervorgehobene Kugel anvisiert und im Anschluss das Sprachkommando verwendet und erkannt, ist die Kugel ausgewählt. Dabei wird eine Funktion aufgerufen, welche auf die Liste mit der Reihenfolge der hervorzuhebenden Kugeln zugreift und anhand derer die nächste Kugel hervorgehoben wird.

3.3.4 Prototyp 3: Manipulationsaufgabe

Der Bildbetrachter aus Modul 3 ([Unterabschnitt 3.2.2](#)) besteht aus verschiedenen Bildern auf denen Zeichnungen von Organen abgebildet sind. Diese sind horizontal angeordnet. Aufgrund der Drehrad Metapher wird nach dem letzten Bild wieder das erste Bild angezeigt, sodass in der Theorie unendlich viel gescrollt werden kann. Beim Scrollen wird ein Bild immer in die Mitte gesetzt und beim Auslösen eines einzelnen Scrollvorgangs werden alle Bilder um eine Einheit verschoben. Dieser Scrollvorgang wird versteckt über Buttons getätigt. Das heißt, dass alle Modalitäten bei ihrer Interaktion letztendlich die Button Funktion aufrufen, welche das eigentliche Scrollen übernimmt. Für das Wechseln der Modalitäten wird erneut der Controller verwendet.

Head-Tracking

Auch bei diesem Prototypen wird das Head-Tracking bei allen drei Interaktionstechniken zum Steuern des Cursors verwendet. Jedoch erfüllt das Head-Tracking je nach Modalität noch eine zusätzliche Funktion. Bei der Sprachinteraktion wird die Kopfposition für das Scrollen der Bilder verwendet. Links und rechts vom mittleren Bild aus befinden sich zwei für den Benutzer nicht sichtbare Flächen, die das *IFocusable* Interface verwenden. Wird der Cursor in einen Collider (Mesh Collider) der jeweiligen Fläche geführt, wird nach links oder rechts gescrollt. Erfolgt nur eine kurze Kopffrotation in eine der beiden Richtungen, wobei der Benutzer seinen Kopf wieder in Ausgangslage positioniert, erfolgt nur das Scrollen eines Bildes in die jeweilige Richtung. Wird der Kopf in eine Richtung rotiert und die Position fixiert erfolgt nach 1,3s ein kontinuierliches Scrollen. Dabei wird alle 0,5s das nächste Bild angezeigt. Die Zeiten wurden durch ein Hallway Testing ermittelt [3], sodass einerseits nicht ungewollt der kontinuierliche Scroll-Modus gestartet wird und andererseits das kontinuierliche Scrollen nicht zu schnell wird. Das Feedback für das kontinuierliche Scrollen erfolgt darüber, dass der Pfeil in die Richtung anstelle von blau grün gefärbt wird. Bei der Fußinteraktion werden ebenfalls die zwei nicht sichtbaren Flächen verwendet, um die gewünschte Richtung für das Scrollen festzulegen und um in den Scroll-Modus zu gelangen. Der Scrollvorgang wird über die Füße aktiviert.

Fußinteraktion

Die Fußinteraktion besteht aus einer Kombination aus Head-Tracking und Fußgeste. Zuerst muss der Kopf in eine Richtung rotiert werden, danach muss der Vorderfuß angehoben werden, gefolgt von einer Rotation der Ferse (*Heel rotation*) in die gleiche Richtung wie der Kopf. Ist einmal eine Richtung registriert muss keine weitere Rotation des Fußes erfolgen. Es folgt ein kontinuierliches Scrollen, solange der Benutzer seinen Kopf in eine Richtung außerhalb der nicht-reaktiven Zone (Bildmitte) gerichtet hat. Der Pfeil in die Richtung färbt sich dabei wieder grün. Das Scrollen kann gestoppt werden, indem der Benutzer entweder seinen Fuß wieder oder seinen Kopf zur Bildmitte rotiert. Für die Erfassung der Fußgeste wird erneut das Programm aus [Abbildung 3.7](#) verwendet.

Für die Heel rotation muss die Checkbox „Rotate“ ausgewählt werden. Bei der Erkennung wird das up-down-rest Prinzip aus dem Selektionsprototypen verwendet. Auch hier werden erneut nur die Orientierungsdaten ausgewertet. Werden vom Programm drei hintereinander folgende up Events erkannt, wird eine Markierung mit ToesUp gesetzt. Für die nächsten zwei Frames muss ein rest Event folgen. Der Benutzer hat dabei seinen Vorderfuß angehoben und führt noch keine Bewegung durch. Das rest Event dient der Feststellung, ob die Geste beabsichtigt ist und ermittelt einen Wert für die spätere nicht-reaktive Zone. Wenn der Fuß danach nach links oder rechts rotiert wird, wird dementsprechend die Rotation durch die Änderung der Orientierungsdaten erkannt. Aufeinanderfolgende Daten müssen hierfür einen Unterschied von mehr als $1,0^\circ$ haben. Zu diesem Zeitpunkt wird eine String Nachricht mit „LINKS“ oder „RECHTS“ über UDP an die HoloLens oder die Lokale Maschine gesendet. Der Prototyp verfügt über ein Skript, welches die Nachrichten empfängt und je nach Nachricht den Button für das links oder rechts Scrollen auslöst. Wird der Fuß zur Mitte orientiert befindet er sich in einer nicht-reaktiven Zone in der keine Rotation ausgelöst wird. Sobald ein down Event registriert wird, wird dies als absetzen des Fußes gedeutet und der Heel rotation Modus wird verlassen. Wie beim Toe tap, wurden die Belegungen der Variablen durch Vorversuche ermittelt. Es musste

ein Kompromiss zwischen der Größe der Fußrotationsbewegung und der Vermeidung von unbeabsichtigter Gestenerkennung gefunden werden. Dabei war in den Vorversuchen die hier beschriebene Belegung am meisten zufriedenstellend.

Um den Prototypen mit einer speziellen Aufgabe zu versehen, muss eine gewisse Anzahl von Bildern in eine vorgesehene Richtung gescrollt werden. Hierbei wird die gewünschte zu scrollende Bildanzahl unter dem Pfeil der vorgesehenen Richtung eingeblendet. Die Aufgabe startet immer mit rechts. Sobald die vorgegebene Anzahl an Bildern gescrollt wurde, leuchtet die gleiche Zahl unter dem linken Pfeil auf. Nach dem links Scrollen, leuchtet rechts eine neue Zahl auf. Die Zahl aktualisiert sich bei den Scrollvorgängen automatisch, sodass nicht mitgezählt werden muss. Nach Beendigung aller Teilaufgaben wird der Prototyp ausgeblendet.

Wie auch bei dem Selektionsprototyp, besitzt dieser Prototyp zwei Modi. Der erste Modus besteht aus der eben genannten Aufgabe mit den beschriebenen Interaktionsmechanismen, der zweite Modus besteht aus einer Kombination aus Modul 1 und 3. Auch hier müssen dann beide Aufgaben parallel bearbeitet werden (Abbildung 3.11).

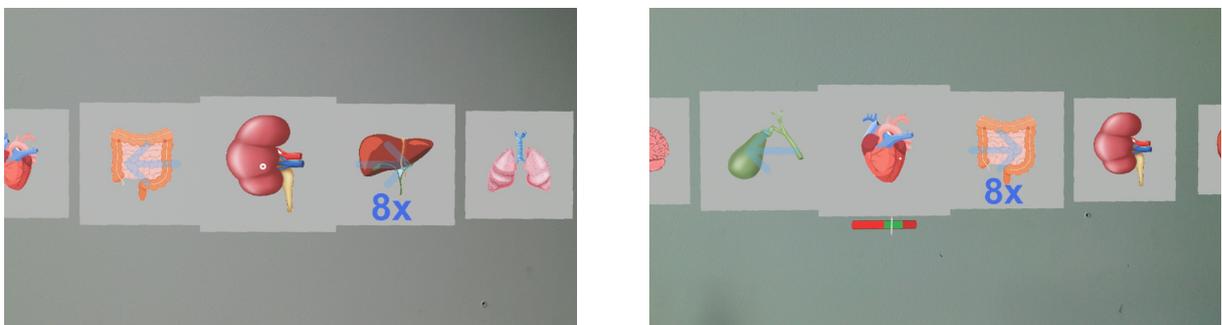


Abbildung 3.11: Prototyp 3. Links: Manipulation, rechts: Manipulation + Primäraufgabe

Handgesteninteraktion

Bei der Interaktion über Handgeste wird eine Geste aus dem Standardkatalog der HoloLens verwendet. Für das Scrollen wird die Drag (*engl.: ziehen*) Geste verwendet. Die Handgeste beginnt ähnlich zum Air tap aus [Abbildung 3.9](#). Im Gegensatz zum Air tap bleiben die Fingerspitzen in der zweiten Phase der Geste zusammen und der Benutzer zieht seine Hand nach links oder rechts.

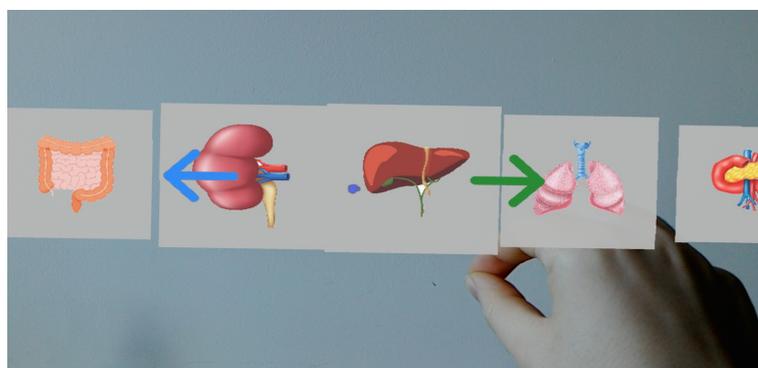


Abbildung 3.12: Scrollen durch die Gesteninteraktion. Der grüne Pfeil zeigt, dass kontinuierlich nach rechts gescrollt wird.

Sobald eine Zugrichtung erkannt wird, beginnt das Scrollen in die Richtung. Der Benutzer kann daraufhin seine Hand an einer Stelle fixieren und das kontinuierliche Scrollen setzt ein (Abbildung 3.12). Werden die Fingerspitzen wieder geöffnet wird das Scrollen beendet.

Wie auch bei der Handgesteninteraktion im Selektionsprototyp, muss die Handgeste innerhalb des sichtbaren Bereiches der HoloLens erfolgen (Abbildung 3.10). Wenn beim Ziehen die Hand den Bereich verlässt, wird die Geste automatisch abgebrochen. Als Feedback für den Benutzer wird der *CursorWithFeedback* Cursor aus dem Mixed Reality Toolkit verwendet. Wird die Hand des Benutzers innerhalb des sichtbaren Bereichs registriert, so wird eine blaue Hand als Cursor verwendet, der den Zustand zwischen den zwei Phasen des Air taps (Abbildung 3.9) anzeigt. In [Abbildung 3.12](#) ist der zweite Zustand, analog zur zweiten Phase des Air taps, zu sehen.

Sprachinteraktion

Für die Sprache werden wieder die Skripte *Speech Input Source* und *Speech Input Handler* verwendet. Bei diesem Prototypen gibt es zwei Arten von Sprachkommandos. Die erste Art ermöglicht es dem Benutzer in den Scroll-Modus zu gelangen, die zweite Art beendet den Scroll-Modus. Es wurde aus den gleichen Gründen bei diesem Prototyp mehrere Sprachkommandos zur Auswahl gestellt wie in [Unterabschnitt 3.3.3](#). Die jeweiligen Kommandos lauten:

- Beginne Scrollen
 - begin
 - move
 - slide
 - start moving
- Beende Scrollen
 - break
 - exit
 - stop action

Gelangt der Nutzer in den Scroll-Modus werden die beiden nicht sichtbaren Flächen aktiviert und somit das Scrollen selbst über Head-Tracking ausgelöst. Dafür muss der Kopf nach links oder rechts rotiert werden. Sobald eine der beiden Flächen über den Raycast getroffen wird, wird um eine Position in die gewünschte Richtung gescrollt. Das Scrollen kann auf zwei Weisen gestoppt werden. Entweder der Benutzer verwendet ein Sprachkommando, welches den Scroll-Modus beendet oder er rotiert seinen Kopf wieder in die Ausgangslage, wodurch er in den nicht-reaktiven Bereich gelangt. Dies würde aber den Scroll-Modus im aktiven Zustand beibehalten und bei erneuter Kopffrotation auslösen.

4 Evaluation

Mithilfe der in [Kapitel 3](#) vorgestellten Prototypen sollen in diesem Kapitel die verschiedenen Interaktionstechniken in Bezug auf die Selektion und Manipulation eines kontinuierlichen Wertes untersucht werden. Ziel ist es, Erkenntnisse zu erlangen, wie die Interaktionstechniken untereinander bewertet werden, wie schnell und genau die Aufgaben gelöst werden können und wie sich Multitasking auf die Leistung und subjektive Bewertung der Techniken auswirkt. Hierfür wird zuerst der Aufbau und Ablauf der durchgeführten Studie beschrieben. Danach werden die Ergebnisse präsentiert und anschließend diskutiert.

4.1 Aufbau und Ablauf

Die Studie wurde in einem Seminarraum im Gebäude der Fakultät für Informatik an der Otto-von-Guericke Universität durchgeführt. Wie in [Abbildung 4.1](#) zu sehen ist, befand sich die Versuchsperson immer zu einer weißen Wand gedreht.

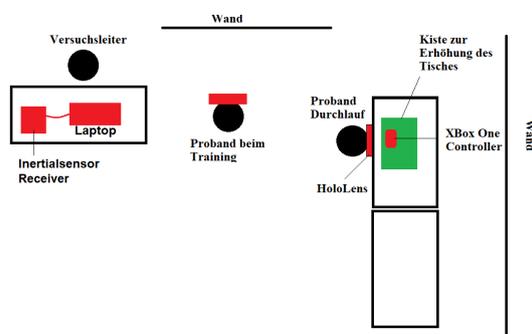


Abbildung 4.1: Links: Schematische Skizze des Aufbaus mit den Positionen der Probanden beim Testdurchlauf und beim gemessenem Durchlauf. Rechts: Bereich des gemessenen Durchlaufs.

Dies diente zum einen zur Vermeidung von schlechtem Kontrast und zum anderen sollte sich die Versuchsperson aufgrund der Lichtverhältnisse vom Fenster weg drehen. Es wurden folgende Geräte für die Studie verwendet:

- Microsoft HoloLens Entwickler Edition
- xSens MTw Awinda Inertialsensor
- Xbox One Controller

- Laptop
- Router

In **Abbildung 4.2** wird gezeigt, wie die Geräte miteinander verbunden waren.

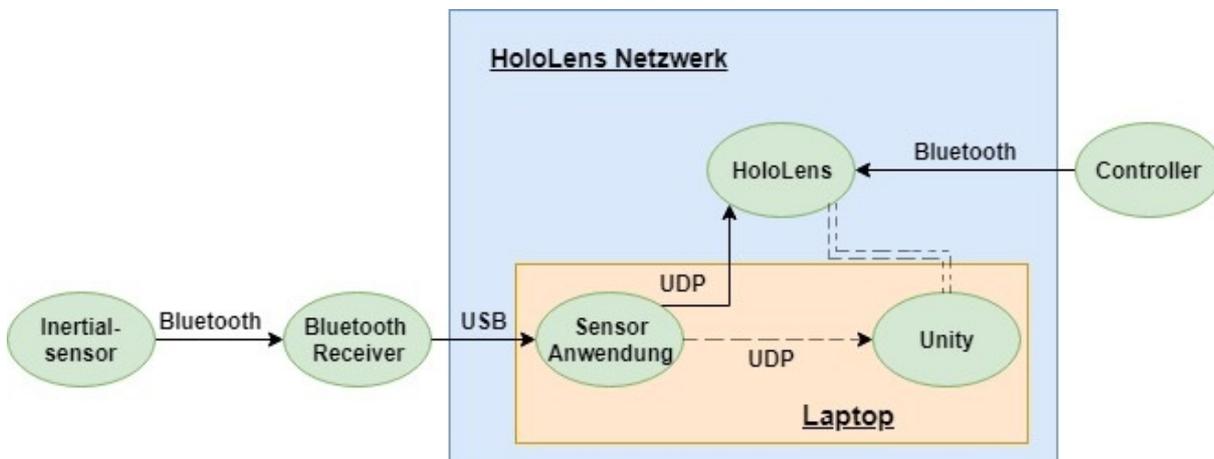


Abbildung 4.2: Der Verbindungsaufbau der verwendeten Geräte. Im Trainingsmodus wird die Verbindung zwischen Sensor Anwendung und HoloLens durch die gestrichelten Verbindungen ersetzt.

Wurde die Fußinteraktion verwendet, wurden über Bluetooth die Orientierungsdaten des Sensors an den Receiver geschickt. Der Receiver war über USB mit dem verwendeten Laptop verbunden. Dieser befand sich mit der HoloLens in einem gemeinsamen Netzwerk. Auf dem Laptop befand sich die C# Anwendung, welche die Sensordaten auslas und diese verarbeitete. Die fertig verarbeiteten Daten wurden über UDP an die HoloLens gesendet. An diese war per Bluetooth ein Controller, welcher für die Primäraufgabe verwendet wurde, gekoppelt. In den Trainingsdurchläufen wurden die verarbeiteten Sensordaten an Unity gesendet. Während dieses Vorgangs war Unity mit der HoloLens verbunden, sodass die Anwendung zusätzlich in Unity gezeigt wurde. Für die Fußinteraktion wurden OP-Schuhe in den Größen von 41/42, 43/44 und 45/46 bereitgestellt.

Die Studie bestand aus zwei Blöcken. Im ersten Block bearbeiteten die Probanden den Selektionsprototypen (im Folgenden Selektionsblock genannt) und im zweiten Block bearbeiteten sie den Manipulationsprototypen (im Folgenden Manipulationsblock genannt). Das Studiendesign sah 12 Probanden vor. Insgesamt haben 14 Probanden, darunter 2 Frauen, an der Studie teilgenommen. Aufgrund von zwei technischen Problemen im Manipulationsblock wurden die Daten zweier Probanden nicht ausgewertet. Deshalb wurden zwei zusätzliche Probanden akquiriert, die den Manipulationsblock durchführten. Da die beiden Aufgabenblöcke in der Durchführung und in den Ergebnissen unabhängig voneinander sind, wurde in den genannten zwei Fällen der Selektionsblock übersprungen. Dadurch veränderten sich die demographischen Daten. Im Selektionsblock betrug das durchschnittliche Probandenalter 25 Jahre und die durchschnittliche Schuhgröße 43,33. Das durchschnittliche Alter sank im Manipulationsblock auf 24,83 Jahre und die Schuhgröße auf 43,1. Innerhalb der demographischen Datenerhebung wurde nach Erfahrungen in Bezug auf Sprachinteraktion (z.B. Amazon Echo oder Smartphone Sprachassistent), Fußinteraktion (z.B. Interaktionen mit Sensorfußboden) oder Handgesteninteraktion (z.B. Handgesten mit Kinect) gefragt. **Abbildung 4.3** zeigt die Verteilung der demographischen Daten.

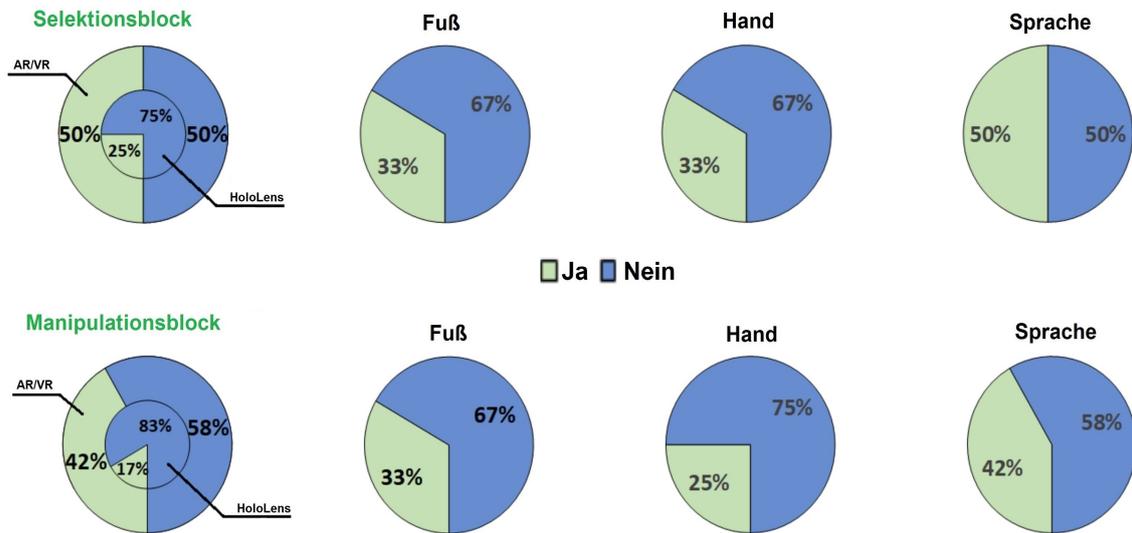


Abbildung 4.3: Das Ergebnis der Befragung bezüglich der Erfahrungen. Oben sind die Ergebnisse aus dem Selektionsblock und unten aus dem Manipulationsblock zu sehen.

Bezüglich der Erfahrungen mit AR/VR Systemen, der HoloLens, Handgesten- und Sprachinteraktion waren die Probanden im Selektionsblock erfahrener. Bei der Fußinteraktion war die Verteilung der Erfahrungen mit 33% der Probanden in beiden Blöcken gleich. Im Selektionsblock hatten vier Probanden in keiner der drei Interaktionstechnologien Erfahrungen, im Manipulationsblock waren es fünf.

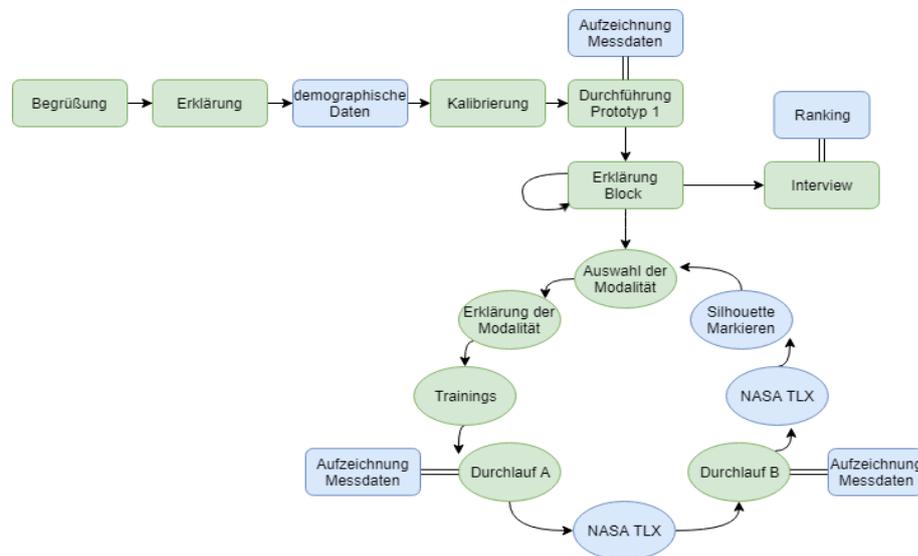


Abbildung 4.4: Der schematische Ablauf der Studie. Es wurden zwei Blöcke durchgeführt mit jeweils drei Modalitäten. Blaumarkierte Kästen zeigen Zeitpunkte an, in denen Daten aufgezeichnet wurden.

In *Abbildung 4.4* wird der schematische Ablauf der Studie skizziert. Zu Beginn erfolgte eine kurze Einführung in die Thematik und eine Beschreibung des groben Ablaufs durch die Versuchsleiterin. Anschließend wurden von der Versuchsleiterin die demographischen Daten der Versuchsperson aufgenommen (siehe Anhang *Unterabschnitt A.1*). Als

nächstes erfolgte eine Kalibrierung der HoloLens auf die Versuchsperson, bei der sie die HoloLens aufsetzen mussten. Das bewegliche Applikationsfenster der Kalibrierung musste in dem Sichtbereich der Versuchsperson mit einem Air tap fixiert werden. Je nachdem, ob die Versuchsperson Rechts- oder Linkshänder war, wurde sie bereits hier darum gebeten, den Air tap mit ihrer nicht dominanten Hand auszuführen. Der Grund hierfür war, dass die Probanden an dieser Stelle mit der Interaktion über die nicht dominante Hand vertraut werden sollten, da die dominante Hand für die Primäraufgabe gebraucht wurde. Die Kalibrierung wurde mit Hilfe der Anweisungen vom Sprachassistenten der HoloLens durchgeführt. Nach erfolgreicher Kalibrierung, wurde der Versuchsperson der Primäraufgabenprototyp ([Unterabschnitt 3.3.2](#)) mit Hilfe eines Schaubildes erklärt. Es folgte direkt danach die Durchführung der Aufgabe. Hierbei sollten die Probanden eine Minute den weißen Marker innerhalb des grünen Bereiches mit Hilfe des Controllers halten. Die Zeitspannen, in denen sich der Marker im grünen Bereich befand, wurden in einer Datei festgehalten. Nach einer Minute wurde das Objekt ausgeblendet und die HoloLens durfte wieder abgesetzt werden. Während des Durchlaufs lag der Controller immer auf der Ablagefläche und die Versuchsperson musste zur Steuerung ihre dominante Hand verwenden. Die andere Hand durfte auf der Ablagefläche oder auf dem Controller selbst platziert werden. Es folgte daraufhin die Beschreibung der Aufgabe aus dem Selektionsblock. In zwei Fällen wurde direkt der zweite Block durchgeführt, da das Problem erst im Manipulationsblock auftauchte und nur dort eine erneute Untersuchung stattfinden musste. Insgesamt dauerte die Studie im Durchschnitt 112 Minuten.

4.1.1 Selektionsblock

In diesem Block führten die Probanden die Aufgabe aus dem Selektionsprototypen durch ([Unterabschnitt 3.3.3](#)). In Abhängigkeit von der jeweiligen Bedingung, wurde die Variante mit oder ohne Multitasking verwendet ([Abbildung 3.8](#)). Vor jedem Ablauf wurde die Versuchsperson darum gebeten, die Aufgabe so schnell wie möglich, aber auch gleichzeitig so genau wie möglich, durchzuführen. Dies bedeutete, dass sie schnelle Kopfbewegungen, die zum Überfliegen des zu selektierenden Objekts (Overshoot) führen, vermeiden sollte. Bei den Abläufen im Multitasking-Kontext wurde zusätzlich darum gebeten, dass die Probanden gleichzeitig darauf achten sollen, mit der Markierung möglichst innerhalb des grünen Bereiches zu bleiben.

Nach der Einführung in die Aufgabe startete der in [Abbildung 4.4](#) aufgeführte Ablauf. Dieser Kreislauf wurde dreimal durchlaufen. Da es bei den Durchläufen zu Lerneffekten kommen kann, wurde bei jeder Versuchsperson die Reihenfolge der Bedingungen sowie der Modalitäten gewechselt. Aus drei Modalitäten und zwei Bedingungen entstanden in diesem Studiendesign sechs verschiedene gemessene Durchläufe. Die Probanden wurden in zwei Gruppen unterteilt. Die eine Hälfte startete mit der Variante ohne Multitasking und die andere Hälfte startete mit der Variante mit Multitasking. Unabhängig von der Gruppe wurde vor jeder Durchführung ein Training für die jeweilige Modalität absolviert. Die Trainingsphase wurde solange durchgeführt, bis die Versuchsperson das Gefühl hatte, die Aufgabe bewältigen zu können und die Interaktionstechnik verstanden zu haben.

Die Trainingsumgebung wurde identisch zu dem Prototypen programmiert. Der einzige Unterschied war, dass die Kugeln in zwei Reihen angeordnet wurden. Eine andere An-

ordnung wurde gewählt, damit die Probanden zuerst auf den größeren Kugeln trainieren konnten. Außerdem wurde die Anzahl auf 10 Kugeln reduziert (siehe [Abbildung 4.5](#)). Dadurch wurde die Trainingsumgebung simpler und der Durchlauf kürzer gestaltet. Fast alle Probanden haben nach der Selektion der 10 Kugeln den Trainingsdurchlauf beendet.

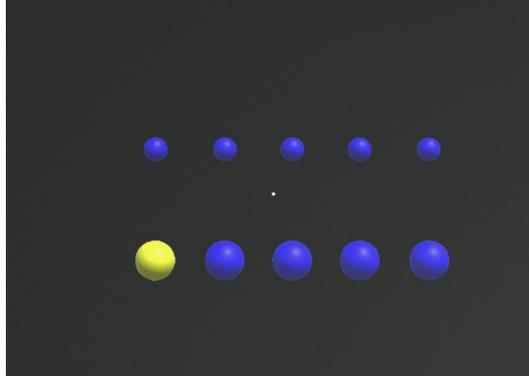


Abbildung 4.5: Die Trainingsumgebung bestand aus zwei Reihen von jeweils fünf Kugeln mit den Durchmessern von 2,33 und 3,36 cm.

Beim Training wurde der Prototyp über WLAN gestartet. Dies erlaubte der Versuchsleiterin das Training zu beobachten, um sicher zu gehen, dass die Aufgabe korrekt durchgeführt wurde. Die Übertragung auf eine lokale Maschine hatte jedoch den Nachteil, dass eine deutlich geringere Framerate angezeigt wurde, wodurch der Prototyp etwas „ruckelte“. Des Weiteren funktionierte die Sprachinteraktion dementsprechend schlechter und die Selektion über alle Modalitäten wurde erschwert. Trotz der verursachten Schwierigkeiten reichte es für die Trainingszwecke. Die im Folgenden beschriebenen Durchläufe spiegeln nicht die Reihenfolge innerhalb der Versuche wieder.

Durchlauf Fußinteraktion

Die Probanden mussten bei diesem Durchlauf ihre Schuhe ausziehen und die am besten passende OP-Schuhe anziehen. Der rechte Schuh wurde mit einem Inertialsensor versehen. Die Versuchsleiterin erklärte und demonstrierte die Toe tap Geste und der Trainingsdurchlauf aus [Abbildung 4.4](#) startete. Während des Trainings konnte die Versuchsperson ein Gefühl für den Toe tap und seine Reaktivität entwickeln. Sobald sich die Versuchsperson sicher fühlte, wurde das Training beendet und es folgten die zwei gemessenen Durchläufe.

Durchlauf Handgesteninteraktion

Hierbei wurde von der Versuchsleiterin die Air tap Geste noch einmal präsentiert und die Versuchsperson wurde zusätzlich auf den sichtbaren Bereich der HoloLens hingewiesen. Danach erfolgte der Trainingsdurchlauf und im Anschluss die zwei gemessenen Durchläufe ([Abbildung 4.6](#)).

Durchlauf Sprachinteraktion

In diesem Durchlauf wurden der Versuchsperson die verschiedenen Sprachkommandos erklärt. Während des Trainingsdurchlaufs wurde ihr ein Blatt mit den Sprachkommandos als Gedankenstütze gegeben. Innerhalb des Trainings durfte sie sämtliche Sprachkommandos ausprobieren. Am Ende des Trainingsdurchlaufs wurde das am besten funktionierende

Sprachkommando ausgewählt. Dieses wurde dann für die gemessenen Durchläufe verwendet.



Abbildung 4.6: Versuchsperson beim Durchlauf mit Handgesteninteraktion im Multitasking-Kontext.

Nach jedem Durchgang wurde von der Versuchsperson ein NASA TLX Fragebogen ausgefüllt (siehe Anhang [Unterabschnitt A.2](#)). Dieser spiegelt die subjektive Beanspruchung des Nutzers wieder. Hierfür werden die Gesichtspunkte geistige, körperliche und zeitliche Anforderung, Ausführung der Aufgabe, Anstrengung und Frustration betrachtet. Es können Antworten auf einer Skala von 0 bis 20 abgegeben werden, wobei 0 als gering und 20 als hoch gewertet wird. Bei *Ausführung der Aufgabe* wird 0 als gut und 20 als schlecht gewertet.

Sobald eine Modalität abgeschlossen wurde, nach Durchführung beider Bedingungen, wurde ein Fragebogen in Anlehnung an die Publikation von Mansfield & Marshall [61] ausgefüllt (siehe Anhang [Unterabschnitt A.2](#)). Hierbei wird eine menschliche Silhouette von der Vorder- und Rückseite mit abgegrenzten Körperregionen abgebildet. Angestrenzte oder ermüdete Regionen werden orange markiert, schmerzende Körperregionen rot. Nach Abschluss des Blocks wurde ein kurzes Interview durchgeführt. Hierbei äußerte die Versuchsperson ihre Eindrücke und ordnete die Interaktionstechnologien nach persönlicher Präferenz. Die gestellten Fragen sind im Anhang [Unterabschnitt A.2](#) zu sehen.

4.1.2 Manipulationsblock

Im Manipulationsblock sollten die Probanden die Aufgabe des in [Unterabschnitt 3.3.4](#) beschriebenen Manipulationsprototypen durchführen. Hierbei wurde der Fokus auf die kontinuierliche Veränderung eines Wertes gelegt, in diesem Fall auf das Scrollen von Bildern. Wie auch im Selektionsblock beschrieben, gab es auch bei dieser Aufgabe zwei Bedingungen (mit und ohne Multitasking), die nach dem gleichen Prinzip durchgeführt wurden. Die Bedingungen sind in [Abbildung 3.11](#) dargestellt. Innerhalb der gesamten Aufgabe mussten die Probanden 6, 8, 10, 12 und 14 Bilder sowohl nach rechts als auch nach links

scrollen. Dabei wurde die Reihenfolge der aufleuchtenden Zahlen randomisiert. Vor jedem Durchlauf wurde die Versuchsperson darum gebeten, so schnell wie möglich die Aufgabe durchzuführen und gleichzeitig darauf zu achten, nicht über die angezeigte Zahl hinaus zu scrollen. Bei den Durchläufen mit Multitasking wurde die Versuchsperson erneut darum gebeten, parallel darauf zu achten, mit der Markierung innerhalb des grünen Bereichs zu bleiben.

Zwischen den zwei Blöcken wurde eine kurze Pause eingelegt. Danach startete der Kreislauf aus [Abbildung 4.4](#) mit einer Einführung in die Manipulationsaufgabe mit Hilfe von einem Schaubild. Auch hier wurden insgesamt sechs gemessene Durchgänge und drei Trainingsdurchgänge durchgeführt. Die Trainingsumgebung war identisch zu der der Manipulationsaufgabe, nur ohne vorgegebene Scrollanzahl unter den Richtungspfeilen. Der Trainingsdurchlauf wurde wieder über WLAN auf dem Laptop der Versuchsleiterin beobachtet.

Durchlauf Fußinteraktion

Die Probanden haben erneut die OP-Schuhe angezogen und der rechte Schuh wurde mit einem Inertialsensor versehen. Die Versuchsleiterin erklärte die Funktionsweise des Scrollens bei diesem Durchlauf und demonstrierte die Heel rotation Geste. Während des Trainingsdurchlaufs wurde die Versuchsperson darauf hingewiesen, wann die Fußgeste nicht registriert wurde. Nachdem die Versuchsperson sich sicher fühlte, starteten die beiden gemessenen Durchläufe.

Durchlauf Handgesteninteraktion

Zuerst wurde das Prinzip und die genaue Ausführung der Drag Handgeste erklärt und demonstriert. Die Versuchsperson wurde darauf hingewiesen, dass die Drag Geste beim Verlassen des sichtbaren Bereiches abgebrochen wird. Während des Trainings konnte sie sich mit der Funktionsweise vertraut machen. Danach erfolgten die zwei gemessenen Durchläufe.

Durchlauf Sprachinteraktion

Für diesen Durchlauf wurden acht verschiedene Sprachkommandos zur Verfügung gestellt. Die Versuchsperson erhielt erneut ein Blatt mit den Sprachkommandos als Gedankenstütze. Innerhalb des Trainings wurden zwei Sprachkommandos (eins pro Funktion) ausgewählt, die am besten erkannt wurden. Zu Beginn des Trainingsdurchlaufes sollte die Versuchsperson in den Scroll-Modus gelangen und ihn dann sofort wieder verlassen. Sobald ihr dies gelang, durfte sie über die Kopfrotation das Scrollen üben. Nach Abschluss der Trainingsphase erfolgten die beiden gemessenen Durchläufe.

In diesem Block wurde ebenfalls nach jedem Durchgang ein NASA TLX Fragebogen und nach jeder Modalität ein Fragebogen in Anlehnung an die Publikation von Mansfield & Marshall [61] ausgefüllt. Am Ende des Manipulationsblocks wurde erneut ein kurzes Interview, in dem die Versuchsperson den Interaktionstechniken Platzierungen zuweisen sollten, durchgeführt.

4.2 Ergebnisse und Diskussion

Im Folgenden werden die einzelnen Ergebnisse aus dem Selektions- und Manipulationsblock präsentiert und diskutiert. In [Abbildung 4.4](#) sind die Stationen im Ablauf blau hervorgehoben, in denen Daten erhoben wurden. Es wurden in beiden Blöcken folgende quantitative und qualitative Daten erfasst:

- **Task Completion Time**

Die gesamte Bearbeitungszeit der Aufgabe vom Start bis zum Ausblenden des Prototyps.

- **Overshoot**

- Selektionsprototyp

Wird die hervorgehobene Kugel anvisiert (Kugel wird grün) und während dieses Zustandes wieder deselektiert (Kugel wird wieder gelb), wird dies als Overshoot gewertet.

- Manipulationsprototyp

Jedes Bild, welches in eine nicht vorgegebene Richtung gescrollt wird, wird als Overshoot gewertet.

- **Primäraufgabenprototyp Leistung**

Es wird die Dauer gemessen, in der sich der weiße Marker innerhalb des grünen Bereiches befindet. Diese wird in Prozentzahl von der Task Completion Time angegeben.

- Baseline (Primäraufgabe alleine)

- Durchgänge Selektion + Primäraufgabe und Manipulation + Primäraufgabe

- **Subjektive Nutzerbewertung**

- subjektive Arbeitsbelastung (NASA TLX Fragebogen)

- subjektive körperliche Belastung (menschliche Silhouette Fragebogen)

- subjektives Ranking der Interaktionstechniken (Interview)

4.2.1 Ergebnisse: Selektionsblock

Task Completion Time

Bei der durchschnittlichen Bearbeitungszeit der Aufgabe ohne Multitasking, hatte die Fußinteraktion einen Median und Durchschnitt von 93s ([Abbildung 4.7](#)). In dem Bereich von 62 bis 128s lagen die Daten symmetrisch verstreut und es gab keine Ausreißer. Die Interaktion über Handgeste hatte eine durchschnittliche Bearbeitungszeit von 84s. Es gab eine Versuchsperson, die für die Aufgabe 144s brauchte. Dies stellte einen Ausreißerwert dar. Die restlichen Probanden befanden sich in einem Bereich zwischen 63 und 94s. Mit einem Median von 68s war die Verteilung rechtsschief. Dies bedeutet, dass die Hälfte aller Zeiten sich im unteren Bereich ansiedelten. Diese waren ebenfalls sehr gebündelt. Die Sprachinteraktion hatte zwei Ausreißer mit den Werten von jeweils 130 und 153s, wodurch der

Durchschnitt auf 96s anstieg. Bis auf den Ausreißer, lagen die Bearbeitungszeiten in einem Bereich zwischen 70 und 107s mit einem Median von 89s.

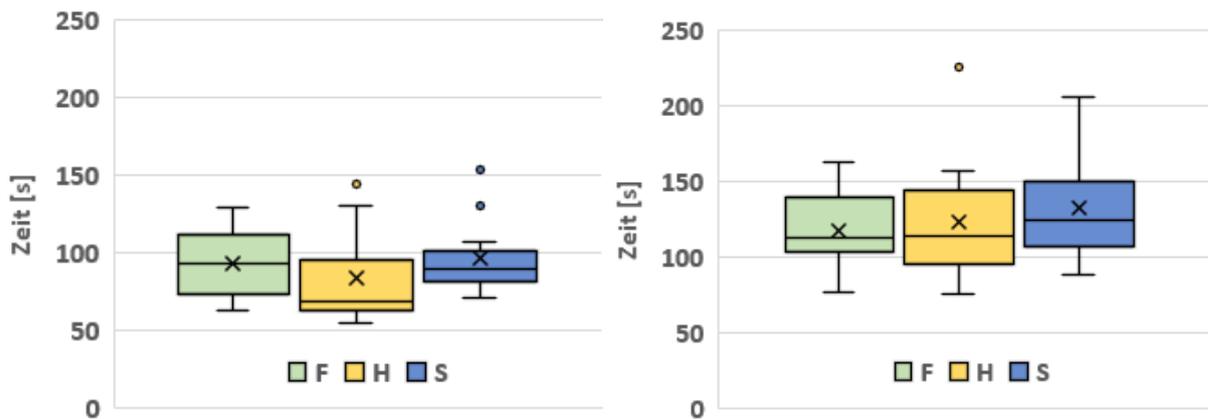


Abbildung 4.7: Die Task Completion Times von den Interaktionen über Fußgeste (F), Handgeste (H) und Sprache (S). Links das Ergebnis ohne und rechts mit Multitasking.

Der Vergleich zeigt, dass die Handgesteninteraktion die geringste durchschnittliche Bearbeitungszeit hatte. Über die Hälfte aller Probanden waren bei dieser Interaktionstechnik am schnellsten. Die Fußinteraktion hatte bezüglich des Durchschnitts und des Medians einen ähnlichen durchschnittlichen Wert zur Sprachinteraktion. Jedoch hatte die Sprachinteraktion den kleineren Interquartilabstand, welcher 50% der Werte beinhaltet, und hatte somit eine geringere Streuung innerhalb der Verteilung.

Bei der durchschnittlichen Bearbeitungszeit der Aufgabe mit Multitasking, hatte die Fußinteraktion einen Durchschnitt von 117s. Die Zeiten lagen in einem Bereich zwischen 76 und 162s. Die Verteilung war mit dem Median von 112s rechtsschief. Die Interaktion über Handgeste hatte einen Durchschnitt von 126s, jedoch befand sich unter den Daten ein Ausreißer mit einem Task Completion Time Wert von 255s. Die durchschnittlichen Zeiten lagen zwischen 75 und 157s. Die Sprachinteraktion hatte einen Durchschnitt von 132s. Die Bearbeitungszeiten befanden sich in einem Bereich von 88 bis 205s.

Im Vergleich hatten Fuß- und Handgesteninteraktion eine ähnliche durchschnittliche Bearbeitungszeit, wobei die Fußinteraktion knapp schneller war. Während die Fußinteraktion die geringste Streuung innerhalb des Interquartilabstands hatte, hatte die Handgesteninteraktion die größte. Verglichen zu den anderen Interaktionstechniken, schnitt Sprachinteraktion bezüglich der Task Completion Time am langsamsten ab.

Es ist zu beobachten, dass im Vergleich zu den Durchläufen ohne Primäraufgabe, die durchschnittlichen Bearbeitungszeiten, in den Durchläufen mit Primäraufgabe, deutlich anstiegen. Dabei brauchten die Probanden bei der Interaktion über Handgeste 42s länger für die Durchführung der Aufgabe. Die Handgesteninteraktion hatte damit den größten Anstieg, gefolgt von Sprachinteraktion mit 36s. Bei der Fußinteraktion wurde im Durchschnitt 24s länger für die Durchführung gebraucht. Der Ausreißerwert bei der Sprachinteraktion lag bei dem Durchlauf mit Multitasking deutlich höher. Die Sprachinteraktion hatte einen deutlich größeren Interquartilabstand im Durchlauf mit Multitasking als im Durchlauf ohne.

Nutzerbewertung NASA TLX

Bei der subjektiven Arbeitsbelastung für die Bedingung ohne Multitasking ist in [Abbildung 4.8](#) zu sehen, dass die Sprachinteraktion nur unter dem Gesichtspunkt *Ausführung der Aufgabe* schlechter als die Handgesteninteraktion abschnitt und ansonsten am besten bewertet wurde. Hingegen fiel die Bewertung für die Fußinteraktion, bis auf die Punkte *Anstrengung* und *geistige Anforderung*, am schlechtesten aus. Die Durchschnitte aller Gesichtspunkte betragen für Handgesteninteraktion 6,07, für die Sprachinteraktion 4,53 und für die Fußinteraktion 7,44.

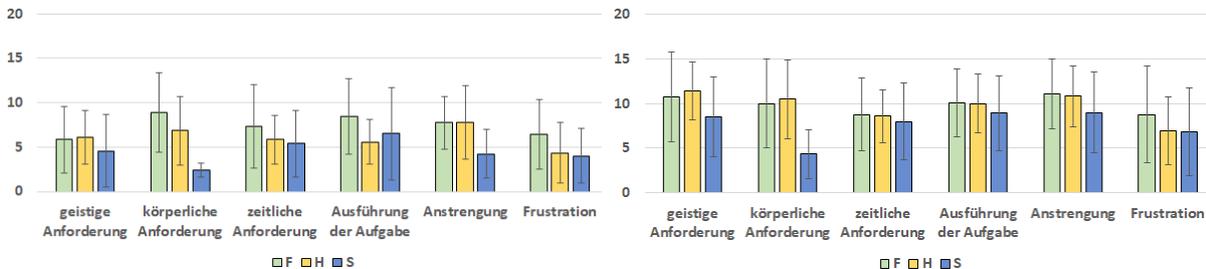


Abbildung 4.8: Die durchschnittliche Bewertung der Arbeitsbelastung von den Interaktionen über Fußgeste (F), Handgeste (H) und Sprache (S). Links die Durchgänge ohne und rechts mit Primäraufgabe.

Bei zusätzlicher Primäraufgabe stieg die Beanspruchung deutlich an. Die jeweils durchschnittlichen Werte aller Punkte betragen für Handgesteninteraktion 9,71, für die Sprachinteraktion 7,6 und für die Fußinteraktion 9,9. Die Sprachinteraktion wurde am besten bewertet, nur unter *Frustration* fiel die Bewertung mit der Handgesteninteraktion gleich aus. Zu bemerken ist, dass die Handgeste unter *geistige Anforderung* und *körperliche Anforderung* am schlechtesten ausfiel und auch unter anderen Gesichtspunkten eine ähnlich hohe Bewertung wie die Fußinteraktion hatte. Lediglich bei der Frustration war die Beurteilung der Fußinteraktion deutlich höher.

Selektionszeit

[Abbildung 4.9](#) zeigt, dass bei den Durchläufen ohne Multitasking die Selektion über Fußgeste eine durchschnittliche Selektionszeit von 2,79s und einen Median von 2,74s hatte. Die durchschnittlichen Selektionszeiten befanden sich in einem Bereich von 1,79 bis 4,03s und die Verteilung war symmetrisch. Bei der Selektion über Handgeste lag der Durchschnitt bei 2,18s. Es gab drei Ausreißer mit den Werten 3,94s, 4,24s und 4,77s. Der Median lag bei 1,55s und die Verteilung war rechtsschief. Die Sprachinteraktion hatte eine durchschnittliche Selektionszeit von 3,04s und einen Ausreißer mit dem Wert von 5,17s. Der Median lag bei 2,76s und die Verteilung war ebenfalls rechtsschief.

Insgesamt hatte die Interaktion über Handgeste zwar die meisten Ausreißer, aber auch die geringste durchschnittliche Selektionszeit. Über die Hälfte der Probanden hatten bei der Handgesteninteraktion die schnellste Selektionszeit. Die Fuß- und Sprachinteraktion hatten fast den gleichen Median, wobei die durchschnittliche Selektionszeit bei der Fußinteraktion etwas schneller war. Der Interquartilabstand war bei der Sprachinteraktion am geringsten.

Die Fußinteraktion hatte in dem Durchlauf mit Primäraufgabe eine durchschnittliche Selektionszeit von 2,96s. Die Zeiten befanden sich in einem Bereich zwischen 1,94 und 4,70s.

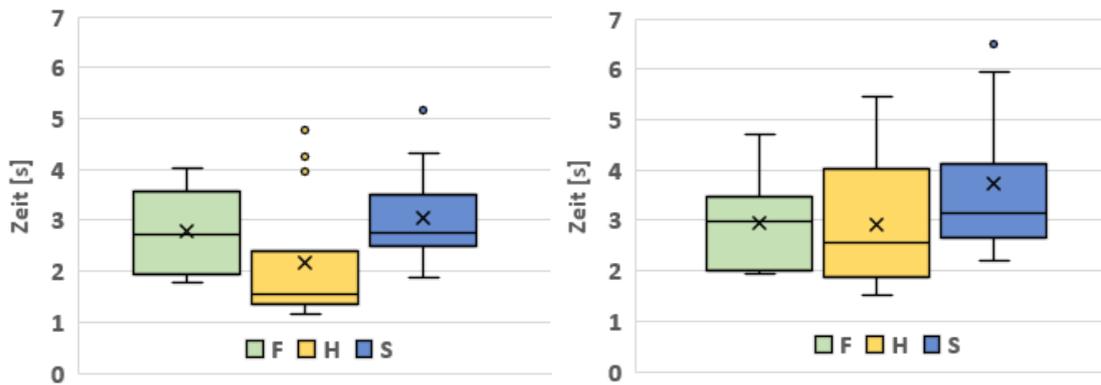


Abbildung 4.9: Überblick über die durchschnittlichen Selektionszeiten der Probanden mit den Interaktionen über Fußgeste (F), Handgeste (H) und Sprache (S). Links die Zeiten ohne und rechts mit Multitasking.

Der Median lag bei 2,99s, womit die Verteilung linksschief war. Die Handgesteninteraktion hatte eine durchschnittliche Selektionszeit von 2,92s und einen Median von 2,57s, womit die Verteilung rechtsschief war. Die durchschnittliche Selektionszeit verteilte sich in einem Bereich zwischen 1,52 und 5,47s. Der Durchschnitt bei der Sprachinteraktion lag bei 3,72s und der Median bei 3,13s. Die Verteilung war rechtsschief. Die Zeiten befanden sich in einem Bereich von 2,20 und 6,49s.

Im Vergleich waren die durchschnittlichen Selektionszeiten von Fuß- und Handgesteninteraktion fast gleich. Die Verteilung der Zeiten waren bei der Interaktion über Handgeste größter, jedoch hatten die Hälfte der Probanden eine durchschnittliche Selektionszeit unter 2,57s hatten, während es bei der Selektion über Fußgeste 2,99s waren. Zwar ist der Interquartilabstand bei der Selektion über Sprache weiter nach oben versetzt, jedoch lag dies an den höheren durchschnittlichen Selektionszeiten von drei Probanden, deren Werte hier nicht als Ausreißer gezählt wurden. Der Median der Sprachinteraktion war nur gering höher als von der Fußinteraktion.

Insgesamt ist ein deutlicher Unterschied bezüglich der durchschnittlichen Selektionszeiten von den Durchläufen mit und ohne Primäraufgabe zu erkennen. Vor allem bei der Selektion über Handgeste, die durchschnittlich 0,74s länger dauerte. Die Sprachinteraktion dauerte durchschnittlich 0,78s länger und die Fußinteraktion hatte den geringsten Unterschied mit 0,17s längerer Selektionszeit. Die Streuung innerhalb des Interquartilabstands stieg bei der Selektion über Handgeste und Sprache stark an, lediglich bei der Selektion über Fußgeste ist sie etwas gesunken.

Overshoot

Abbildung 4.10 zeigt, dass unter beiden Bedingungen (mit und ohne Multitasking) die Interaktion über Handgeste die wenigsten Overshoots pro Selektion hatte (2,58 und 3,83). Die Fußinteraktion hatte hingegen in beiden Durchläufen die höchste Anzahl an Overshoots (3,98 und 4,39). Die Sprachinteraktion befand sich dazwischen (2,94 und 4).

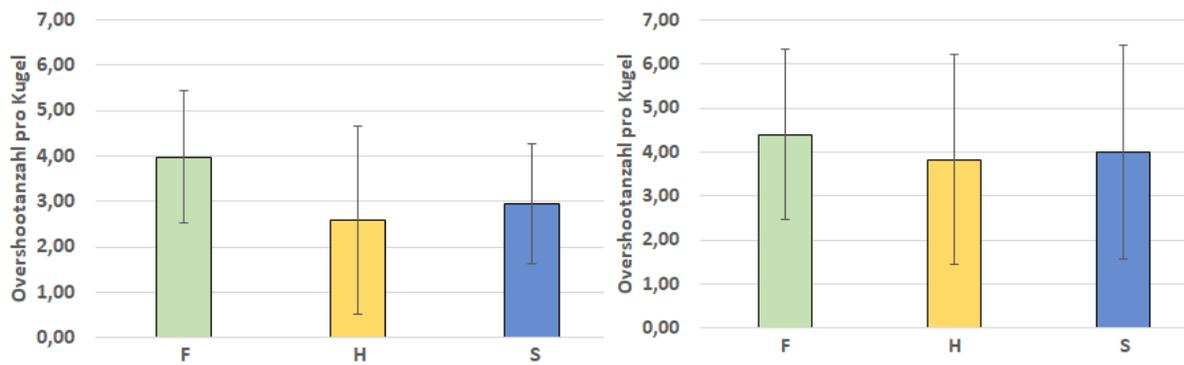


Abbildung 4.10: Die durchschnittliche Overshoot Anzahl pro Selektion der Probanden über Fußgeste (F), Handgeste (H) und Sprache (S). Links die Anzahl ohne und rechts mit Multitasking.

Im Durchgang mit Multitasking stiegen die Anzahl der Overshoots um durchschnittlich +1 pro Selektion an. Den größten Anstieg hatte die Handgesteninteraktion mit 1,25 mehr Overshoots pro Selektion, während Fußinteraktion mit 0,41 den geringsten Anstieg aufwies.

Leistung innerhalb der Primäraufgabe bei den Selektion + Primäraufgabe Durchläufen

13 von 14 Probanden erzielten innerhalb der Primäraufgabe durchschnittlich über 90% (Abbildung 4.11).

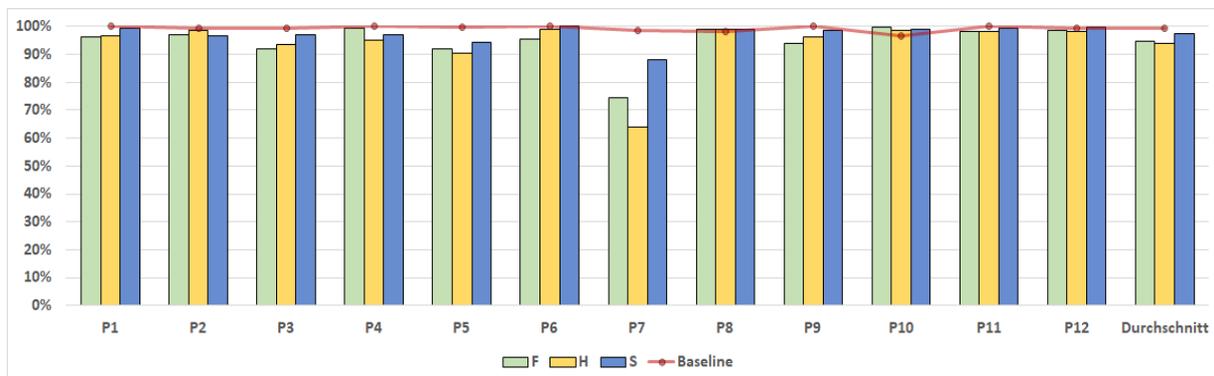


Abbildung 4.11: Die Leistungen der einzelnen Probanden innerhalb der Primäraufgabe. Die Prozentzahlen richten sich nach dem Anteil des im grünen Bereich gehaltenen Markers. Die rote Linie stellt die Baseline dar.

Das Resultat war bei fast allen schlechter als die Baseline. Nur eine Versuchsperson war mit 74,41% (Fuß), 64,12% (Hand) und 87,93% (Sprache) deutlich schlechter als ihr Baselinewert (98,42%). Eine Versuchsperson war mit 99,81% (Fuß), 98,55% (Hand) und 99,08% (Sprache) besser als ihr Baselinewert mit 96,59%. Im Durchschnitt schnitt die Sprachinteraktion mit 97,34% gegenüber der Fußinteraktion mit 94,66% und der Handgesteninteraktion mit 93,95% am besten ab.

Körperliche Belastung

Abbildung 4.12 zeigt die Auswertung für die körperliche Belastung. Die Zahl steht für die Anzahl der Probanden, die eine bestimmte Körperregion mit einer Farbe markiert haben.

Haben ≥ 3 Probanden eine Region markiert, wurde diese in der Farbe dargestellt, in der sie am häufigsten markiert wurde. In fast allen Fällen war dies orange. Sobald mindestens eine Versuchsperson eine Region rot markiert hat, wurde die jeweilige Region zusätzlich mit rot schraffiert.

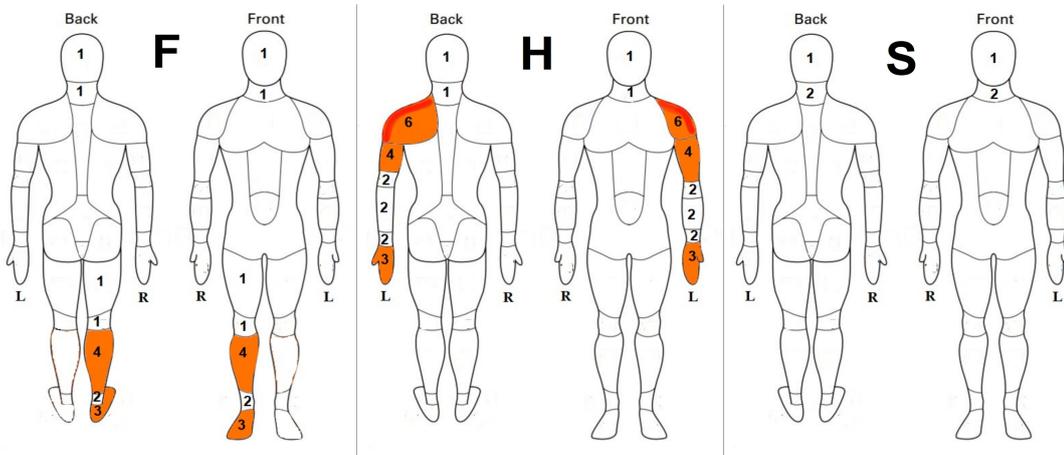


Abbildung 4.12: Auswertung der körperlichen Belastung bei den Interaktionen über Fußgeste (F), Handgeste (H) und Sprache (S). Orange steht für angestrengte oder ermüdete Körperregionen, rot für schmerzende Regionen.

Bei der Fußinteraktion wurde der Bereich der Wade, des Schienbeins und des Fußes als ermüdet oder angestrengt angegeben. Die anderen Bereiche des rechten Beins (der rechte Fuß wurde für die Selektion verwendet) wurden ebenfalls angegeben, haben jedoch nicht die Anzahl von ≥ 3 überschritten. Eine Versuchsperson gab bei der Fußinteraktion und Handgesteninteraktion einen angestrenkten Kopf- und Nackenbereich an, bei der Sprachinteraktion waren es zwei Probanden im Nackenbereich. Die Hälfte der Probanden gaben bei der Handgesteninteraktion (diese wurden von allen mit links ausgeführt) eine angestrengte linke Schulterpartie sowie einen angestrenkten Bizeps und Trizeps an. Eine Versuchsperson hatte eine schmerzende Schulter und drei Probanden hatten eine ermüdete oder angestrengte Hand.

Ranking

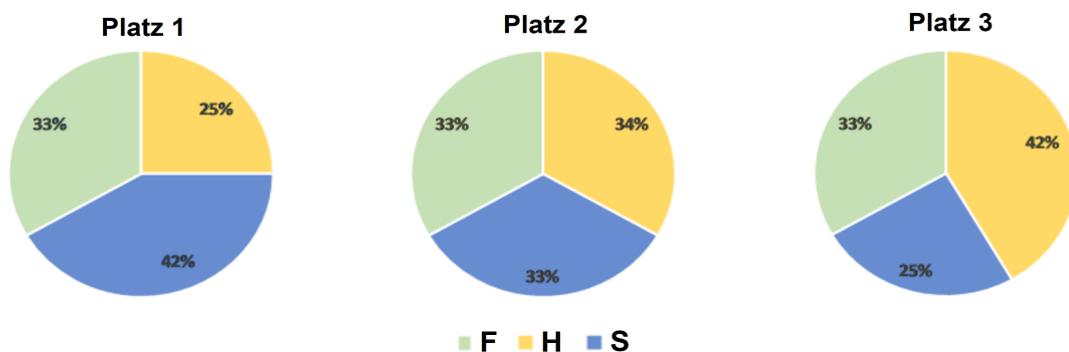


Abbildung 4.13: Die Platzierungen der Interaktionen über Fußgeste (F), Handgeste (H) und Sprache (S) ohne Primäraufgabe.

Abbildung 4.13 zeigt die Platzierungen der Interaktionstechniken der Probanden. Bei Platz 1 und 3 wurden die Anteile bei der Sprachinteraktion und der Interaktion über Handgeste getauscht, wodurch die Sprachinteraktion von 42% der Probanden die favorisierte Interaktionstechnik darstellte und die Handgesteninteraktion mit 42% am wenigsten favorisiert wurde.

Bei den Durchläufen mit Multitasking ist in Abbildung 4.14 eine deutliche Veränderung zu erkennen. Keiner der Probanden favorisierte die Interaktion über Handgeste und acht der 12 Probanden bevorzugten sie am wenigsten. Während die Interaktion über Fußgeste gleich oft favorisiert wurde, nahm der Anteil von Probanden, denen Sprachinteraktion am meisten gefiel, deutlich zu. Über die Hälfte der Probanden setzten die Fußinteraktion auf Platz 2.

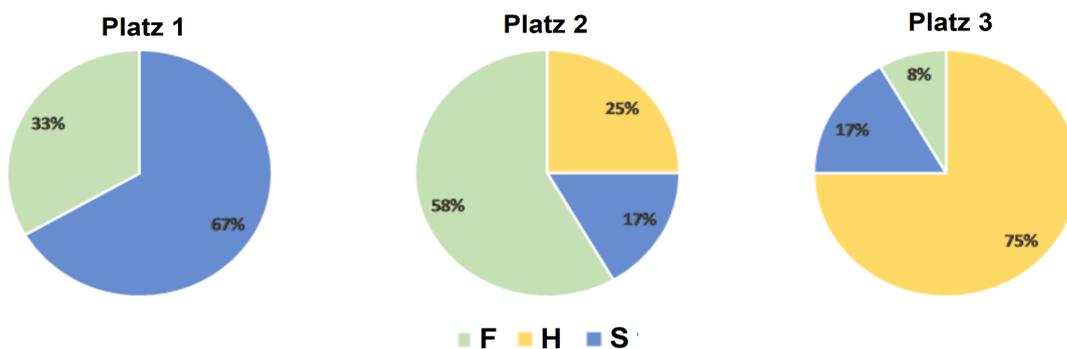


Abbildung 4.14: Die Platzierungen der Interaktionen über Fußgeste (F), Handgeste (H) und Sprache (S) mit Primäraufgabe

4.2.2 Diskussion: Selektionsblock

Wird die Task Completion Time in den Durchläufen ohne Multitasking betrachtet (Abbildung 4.7), wurde bei der Ausführung der Aufgabe mit der Handgesteninteraktion die schnellste Zeit erzielt. Allerdings lag der Abstand zu der nächstbesten Interaktionstechnik unter 20s. Die bessere Bearbeitungszeit spiegelte sich auch in der durchschnittlichen Selektionszeit und der Anzahl an Overshoots pro Selektion wieder, welche im Vergleich am niedrigsten waren. Bei der Nutzerbewertung wurde die Handgesteninteraktion meist zwischen Fuß- und Sprachinteraktion ohne ausschlaggebende Ausreißer bewertet. Dennoch wurden die Methoden von den Probanden nicht präferiert, wie Abbildung 4.13 zeigt. Ein Grund hierfür zeigt Abbildung 4.12. Da die Aufgabe bei 11 von 12 Probanden über eine Minute dauerten, traten im linken Schulter-, Arm- und Handbereich Anstrengungen und Müdigkeit auf. Eine Versuchsperson gab an, Schmerzen in der Schulter gehabt zu haben. Des Weiteren wurde beobachtet, wie mehrere Probanden zwischenzeitig ihren Arm ausschüttelten, bevor sie die Handgeste weiter ausführten.

Die Sprachinteraktion wurde von mehr Probanden als die beste Technik angegeben, welches sich in der Nutzerbewertung widerspiegelte (Abbildung 4.8). Vor allem unter den Gesichtspunkten *körperliche Anforderung* und *Anstrengung*, wurde die Sprachinteraktion sichtlich besser bewertet. Dies ist plausibel, da hierbei lediglich der Kopf bewegt werden musste, um die Kugel zu selektieren. Werden bei der Task Completion Time

und der durchschnittlichen Selektionszeit die Ausreißerwerte bei der Sprachinteraktion nicht mitbetrachtet, so fiel bei den Durchläufen ohne Multitasking die Fußinteraktion am schlechtesten aus. Auch die Zahl der Overshoots waren bei der Fußinteraktion am höchsten.

An dieser Stelle ist zu beachten, dass der Schwierigkeitsgrad der Aufgabe mit 3,4 bit relativ hoch war. Der Tipptest nach ISO 9241-420:2011 setzte voraus, dass alle Kugeln sichtbar sein müssen. Da der Sichtbereich der HoloLens sehr schmal ist, mussten die Zielobjekte kleiner gemacht werden, wodurch der hohe Schwierigkeitsgrad zustande kam. Dadurch wurde das Anvisieren über Head-Tracking zwar erschwert, aber dies hielt sich noch im Rahmen. Der höhere Schwierigkeitsgrad wirkte sich deutlich mehr auf die Selektion aus und stellte damit die größte Problematik dar.

Bei der Ausführung des Toe taps bewegte sich der Kopf durch die Erschütterung unweigerlich mit. Bei den kleinen Zielobjekten führte dies häufig dazu, dass der Cursor kurzzeitig aus der Kugel wanderte und es zu keiner Selektion kam, obwohl die Geste erkannt wurde. Dieses Phänomen führte dementsprechend zu einer höheren Overshootanzahl, zu einer längeren Selektionszeit und somit zu einer höheren Frustration der Probanden. Diesbezüglich hing es stark davon ab, mit wie viel „Gefühl“ die Probanden den Toe tap ausführten. Bei Probanden, die ihren Fuß sanfter bewegten, erfolgte eine geringere Erschütterung, wodurch die Selektion besser durchgeführt wurde. Hierbei wurde meist das Gewicht auf das andere Bein verlagert, um den Tap sanfter auszuführen. Probanden, die ihr Gewicht weniger verlagerten und die Fußbewegung stärker ausführten, hatten demnach öfter Schmerzen im Schienbein- und Wadenbereich. Die Sprachinteraktion wurde ebenfalls dadurch beeinflusst. Einige Probanden berichteten davon, dass sich ihr Kopf aufgrund des Sprechens mitbewegte und dadurch Kugeln häufig deselektiert wurden. Da bei der Air tap Armbewegung meist keine Kopfbewegung verursacht wurden, war das Problem bei der Handgesteninteraktion nicht präsent.

Im Gegensatz zum Ranking ohne Multitasking, fiel das Ergebnis bei den Durchläufen mit Multitasking deutlich bestimmender aus. Wie in [Abbildung 4.14](#) zu sehen ist, präferierten die meisten Probanden Sprachinteraktion und die wenigsten Handgesteninteraktion. Dies spiegelte sich auch in den Ergebnissen des NASA TLX wieder ([Abbildung 4.8](#)), in denen die Bewertung für Handgesteninteraktion in Bezug auf *geistige Anforderung*, *körperliche Anforderung* und *Anstrengung* deutlich anstieg. Auch die Ergebnisse innerhalb der Task Completion Time und der Selektionszeit hatten einen deutlich geringeren bis gar keinen Abstand zu den anderen Interaktionstechniken.

Da bei der parallelen Ausführung von Selektion und Steuern des Controllers in beiden Fällen die Hände benötigt wurden, wurde hierbei der gleiche motorische Kanal beansprucht. Die Verwendung von beiden Händen bei verschiedenen Tätigkeiten steigerte die geistige Anforderung, was vielen Probanden missfiel. Eine Versuchsperson äußerte das Gefühl, bei der Handgesteninteraktion mehr tun zu müssen, da sie noch zusätzlich darauf achten musste, ihre Hand im sichtbaren Bereich der HoloLens zu halten. Zwar wurde bei der Fußinteraktion ebenfalls der motorische Kanal doppelt beansprucht, jedoch ließen sich hierbei die Bewegungen besser separieren.

Trotz der etwas längeren Task Completion Time der Sprachinteraktion, ließ sie sich in einem Multitasking-Kontext deutlich besser vereinen. Ein Nachteil der Sprachinteraktion, der aber in diesem Aufgabenkontext zum Vorteil wurde, war die verzögerte Reaktionszeit des Spracherkennungssystems. Wurde das Sprachkommando geäußert, wur-

de die Selektion verzögert ausgeführt. So passierte es oft, dass Probanden Dank der unbeabsichtigten Kopfbewegung die Kugel kurzzeitig deselektierten, aber diese bei der eigentlichen Durchführung der Selektion bereits wieder anvisierten. Auch die Auswertung der Primäraufgabe zeigte (Abbildung 4.11), dass dort die Aufgabe am besten erfüllt wurden und die Probanden im Durchschnitt nur 1,94% schlechter waren als ihre Baseline.

4.2.3 Ergebnisse: Manipulationsblock

Task Completion Time

Die Bearbeitungszeit für den Manipulationsblock verlief bei diesem Aufgabentyp schneller als im ersten Block. Abbildung 4.15 zeigt, dass die Fußinteraktion einen Durchschnitt von 90s hatte.

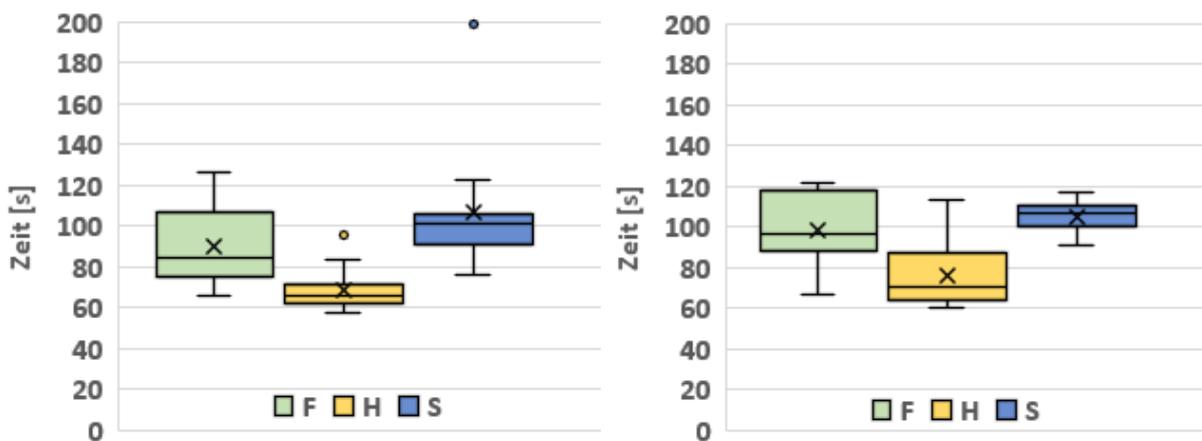


Abbildung 4.15: Die Task Completion Times von den Interaktionen über Fuß (F), Hand (H) und Sprache (S). Links das Ergebnis ohne und rechts mit Multitasking.

Die Probanden hatten eine Bearbeitungszeit zwischen 66 und 126s, wobei die Hälfte der Probanden unter 105s benötigten. Es gab keine Ausreißer. Bei der Interaktion über Handgeste lag die durchschnittliche Bearbeitungszeit bei 68s. Ein Proband benötigte 95s für die Bearbeitung, dessen Zeit als Ausreißer gekennzeichnet wurde. Alle anderen Bearbeitungszeiten befanden sich in einem Bereich von 57 bis 83s. Die Sprachinteraktion hatte einen Durchschnitt von 107s. Der höhere Durchschnitt bei der Sprachinteraktion wurde durch die Versuchsperson, die für die Bearbeitung 199s brauchte, beeinflusst. Die restlichen Probanden befanden sich zwischen 76 und 122s. Die Hälfte der Probanden hatte länger als 100s für die Bearbeitung der Aufgabe benötigt.

Die Handgesteninteraktion war in diesem Durchlauf bezüglich der Task Completion Time am schnellsten. Die Fußinteraktion war schneller als die Sprachinteraktion, jedoch war die Streuung der Werte innerhalb des Interquartilabstands am größten.

Bei den Durchläufen mit Primäraufgabe hatte die Fußinteraktion eine durchschnittliche Bearbeitungszeit von 99s. Die Verteilung der Zeiten lag zwischen 67 und 121s und war rechtsschief. Die Hälfte der Probanden hatten eine Bearbeitungszeit von unter 96s. Die Interaktion über Handgeste hatte einen Durchschnitt von 76s und einen Median von

71s. Die durchschnittlichen Bearbeitungszeiten befanden sich in einem Bereich von 60 bis 113s. Die Verteilung war rechtsschief, die Hälfte der Probanden hatten eine Task Completion Time unter 71s. Die Sprachinteraktion hatte einen Durchschnitt von 105s und einen Median von 106s. Die Verteilung war hier symmetrisch. Die Zeiten befanden sich in einem Bereich zwischen 91 und 117s. Keiner der drei Interaktionstechniken hatte Ausreißerwerte.

Die Interaktion über Handgeste hatte auch in diesem Durchlauf die schnellste Bearbeitungszeit, jedoch auch die größte Streuung. Die Fußinteraktion hatte die größte Streuung innerhalb des Interquartilabstands, dennoch waren über die Hälfte der Probanden schneller als bei der Sprachinteraktion. Die Sprachinteraktion hatte die geringste Streuung, jedoch die schlechteste Task Completion Time.

Ein Vergleich der Durchläufe ohne und mit Primäraufgabe zeigt, dass die durchschnittlichen Bearbeitungszeiten von Fuß- und Handgesteninteraktion im Multitasking-Kontext etwas anstieg. Die Probanden brauchten bei der Fußinteraktion durchschnittlich 9s länger für die Aufgabe, bei der Handgesteninteraktion war es mit 7s etwas weniger. Lediglich bei der Sprachinteraktion brauchten die Probanden im Durchschnitt 2s weniger. Die Unterschied zwischen der Streuung der Daten beider Durchläufe, war bei der Handgesteninteraktion am größten.

Nutzerbewertung NASA TLX

Bei der subjektiven Arbeitsbelastung der Interaktionstechniken ohne Primäraufgabe wurde die Fußinteraktion, außer unter den Gesichtspunkten *körperliche Anstrengung* und *Frustration*, am besten bewertet [Abbildung 4.16](#). Die Probanden schätzten ihre Leistung mit Sprachinteraktion am schlechtesten ein und waren dort am meisten frustriert. Der Durchschnitt aller Kategorien betrug bei der Fußinteraktion 4,32, bei der Sprachinteraktion 5,46 und bei der Handgesteninteraktion 5,65.

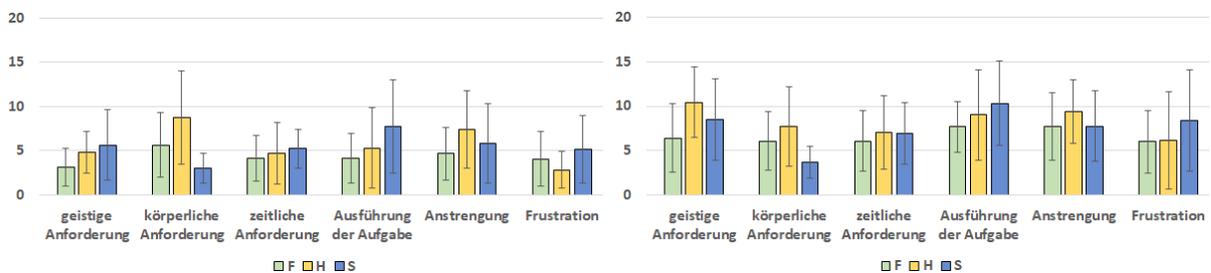


Abbildung 4.16: Die durchschnittliche Bewertung der Arbeitsbelastung der Interaktionen über Fuß (F), Hand (H) und Sprache (S). Links die Durchgänge ohne und rechts mit Primäraufgabe.

Das Resultat zeigt, dass sämtliche Bewertungen bei den Durchläufen mit Primäraufgabe anstiegen, bis auf den Wert bei *körperliche Anforderungen* bei der Handgesteninteraktion. Wie auch bei den Durchläufen ohne Primäraufgabe, war die Fußinteraktion unter allen Gesichtspunkten, bis auf *körperliche Anstrengung* und *Frustration*, am besten.

Overshoot

Unter beiden Bedingungen hatte die Sprachinteraktion mit 1,10 (ohne Multitasking) und 0,97 (mit Multitasking) falsch gescrollten Bildern pro Scrollvorgang die meisten Overshoots (Abbildung 4.17).

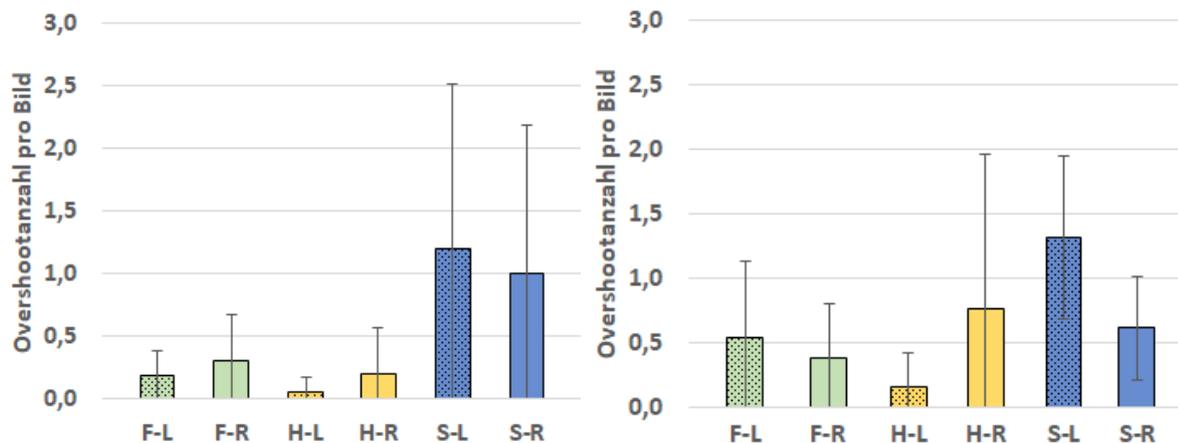


Abbildung 4.17: Die durchschnittliche Overshoot Anzahl pro Scrollvorgang der Probanden für jede Modalität. Links befindet sich die Anzahl ohne und rechts mit Multitasking. Die Overshoots wurden nach links und rechts Overshoots unterteilt.

Während in den Durchläufen ohne Primäraufgabe die Handgesteninteraktion insgesamt mit 0,13 weniger Overshoots als die Fußinteraktion mit 0,24 hatte, waren die beiden Interaktionstechniken in den Durchläufen mit Primäraufgabe mit 0,47 gleichauf.

Die gesamte durchschnittliche Overshootanzahl im Multitasking-Kontext von 0,49 auf 0,63. Erneut hatte die Handgesteninteraktion mit 0,34 mehr Bilder pro Scrollvorgang den größten Anstieg, während die Sprachinteraktion mit 0,13 den geringsten Anstieg hatte.

Auffällig ist, dass die Overshootanzahl bei der Handgesteninteraktion unter beiden Bedingungen zum einen bei rechts deutlich höher war als bei links, zum anderen der Anstieg bei links auch deutlich geringer war. Bei der Sprachinteraktion hingegen war die Overshootanzahl bei links höher als bei rechts, jedoch verringerte sich bei rechts die Overshootanzahl.

Leistung innerhalb der Primäraufgabe bei den Manipulation + Primäraufgabe Durchläufen

Abbildung 4.18 zeigt die Leistungen aller Probanden innerhalb der Primäraufgabe. Ganz rechts befindet sich die durchschnittliche Leistung. Im Vergleich zu den Leistungen im Selektionsblock war die durchschnittliche Leistung aller Probanden bei der Fuß- und Handgesteninteraktion besser. Lediglich die Sprachinteraktion schnitt schlechter ab. Es hielten alle Probanden bei allen Interaktionstechniken den Marker $\geq 90\%$ der Zeit innerhalb des grünen Bereichs. Zwar wurde mit der Sprachinteraktion mit 98,49% wieder die beste Leistung erbracht, jedoch war der Abstand zu der Fußinteraktion mit 96,88% und der Handgesteninteraktion mit 97,83% geringer.

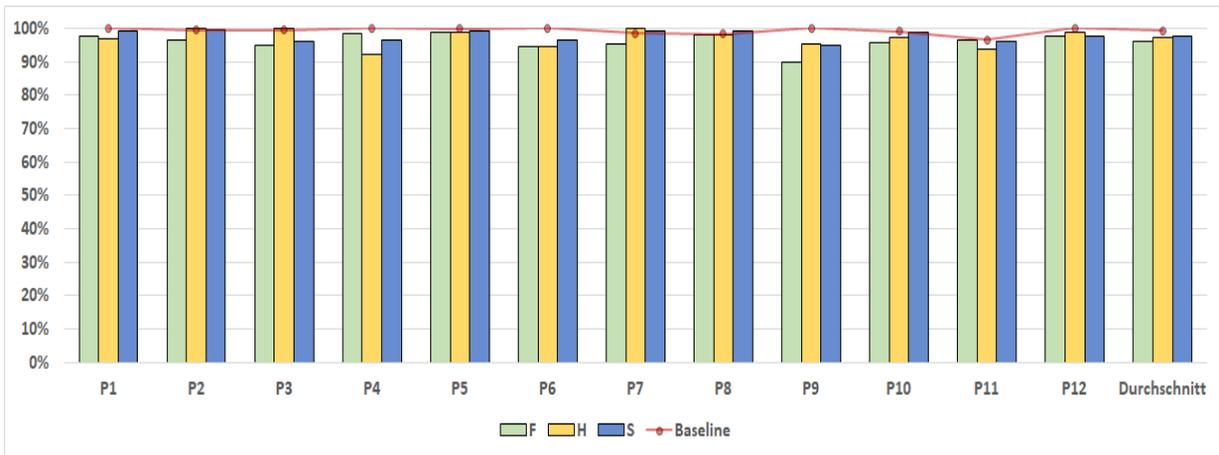


Abbildung 4.18: Die Leistungen der einzelnen Probanden innerhalb der Primäraufgabe. Die Prozentzahlen richten sich nach dem Anteil des im grünen Bereich gehaltenem Marker. Die rote Linie stellt die Baseline dar.

Körperliche Belastung

Als ermüdete oder angestrengte Region wurde bei der Sprachinteraktion der Nacken- und Halsbereich angegeben (Abbildung 4.19).

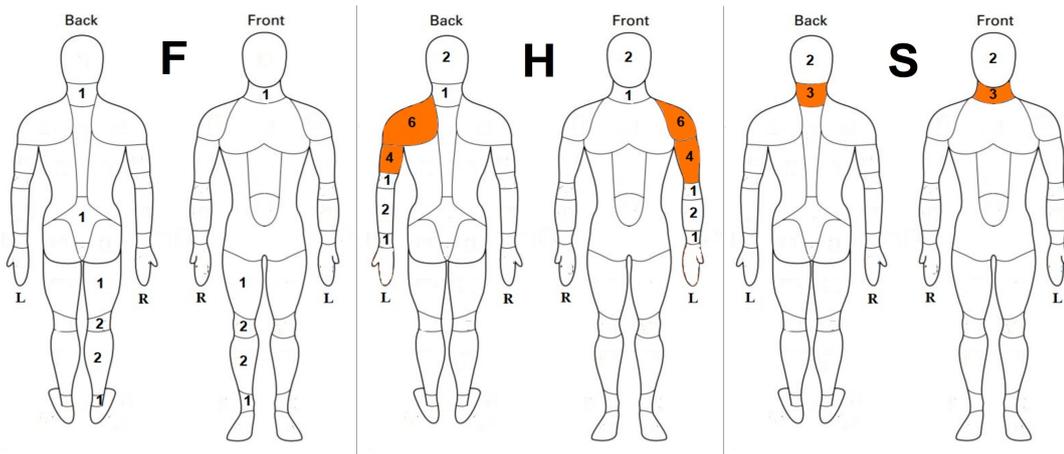


Abbildung 4.19: Auswertung der körperlichen Belastung. Orange steht für angestrengte oder ermüdete Körperregionen. Rot für schmerzende Regionen.

Zwei Probanden gaben an, dass ihre Augen bei der Sprach- und Handgesteninteraktion angestrengt waren. Bei der Interaktion über Handgeste wurden fast die gleichen Körperpartien als ermüdet oder angestrengt angegeben wie im Selektionsblock. In diesem Block hatte aber keiner der Probanden Schmerzen in der Schulter und niemand gab die Region der Hand an. Zwar wurde bei der Fußinteraktion der Bereich des Beins von einigen Probanden angegeben, jedoch wurde nicht die Grenze von ≥ 3 überschritten. Eine Versuchsperson hatte beim Scrollen durch die Bilder einen angestrenzten unteren Rücken.

Ranking

Bei der Platzierung der Interaktionstechniken ohne Multitasking (Abbildung 4.20) setzte keiner der Probanden die Sprachinteraktion auf Platz 1 und mehr als die Hälfte gaben die

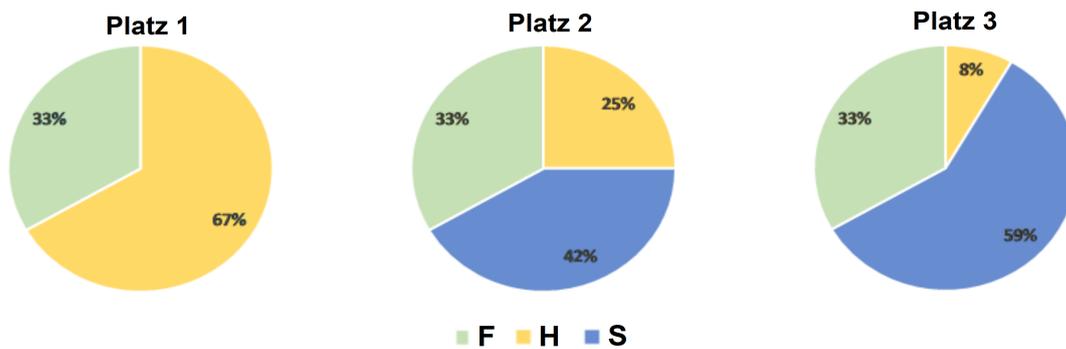


Abbildung 4.20: Die Platzierungen der Interaktionstechniken ohne Primäraufgabe.

Sprachinteraktion als die am wenigsten favorisierte Technik für diesen Aufgabentypen an. Die Fußinteraktion wurde unter allen drei Platzierungen von einem Drittel der Probanden angegeben. Mit 67% wurde die Interaktion über Handgeste von den Probanden am meisten favorisiert.

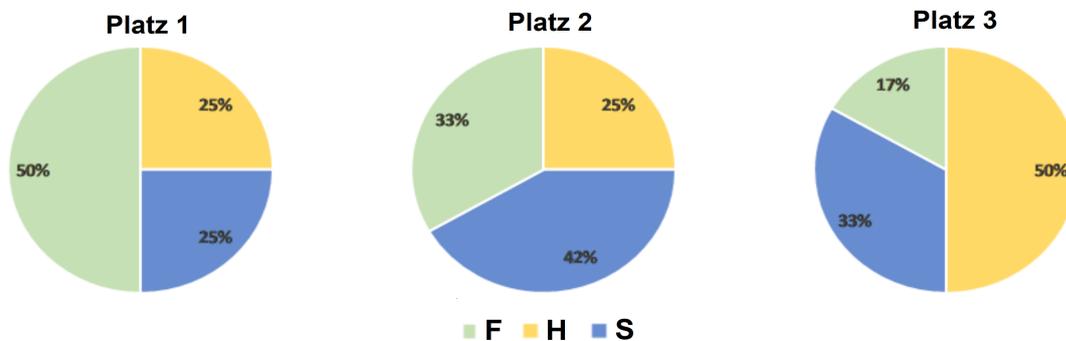


Abbildung 4.21: Die Platzierungen der Interaktionstechniken mit Primäraufgabe.

Die Platzierung mit Multitasking (Abbildung 4.21) zeigt, dass die Hälfte der Probanden die Fußinteraktion am meisten und die Handgesteninteraktion am wenigsten favorisierten. Im Gegensatz zu den Durchläufen ohne Primäraufgabe, wurde von 25% der Probanden die Sprachinteraktion favorisiert. Die Verteilung innerhalb der zweiten Platzierung blieb gleich.

4.2.4 Diskussion: Manipulationsblock

Die meisten Probanden berichteten, dass die Aufgabe aus dem Manipulationsblock ihnen deutlich leichter fiel und weniger anstrengend war. Dies ist mit dem Schwierigkeitsgrad der Aufgabe und der längeren Dauer begründet. Des Weiteren fanden einige Probanden es einfacher, sich bei den Durchläufen mit Multitasking auf die Primäraufgabe zu konzentrieren.

Die Betrachtung der quantitativen Messdaten unter der Bedingung ohne Primäraufgabe zeigt, dass die Handgesteninteraktion in Bezug auf die Task Completion Time (Abbildung 4.15) und Overshootanzahl (Abbildung 4.17) die besten Resultate lieferte. Bei der

durchschnittlichen Bearbeitungszeit hatte die Handgesteninteraktion mehr als 20s Abstand zu den anderen Interaktionstechniken. Bei der subjektiven Arbeitsbelastung ([Abbildung 4.16](#)) bewerteten die Probanden die Handgesteninteraktion in Bezug auf *körperliche Anforderungen* und *Anstrengung* schlechter, was jedoch bei der Handgeste plausibel ist. Zwar bewerteten die Probanden sich selbst unter *zeitliche Anforderung* und *Ausführung der Aufgabe* schlechter als bei der Fußinteraktion, dennoch gaben 75% der Probanden beim Ranking ([Abbildung 4.20](#)) die Handgeste als Favoriten an. Die etwas schlechtere Bewertung war, insbesondere unter dem ersten Punkt, so gering, dass sie nicht signifikant ist.

Ein Drittel der Probanden entschieden sich für die Fußinteraktion ([Abbildung 4.20](#)), die innerhalb der quantitativen Daten zwischen den beiden Interaktionstechniken angesiedelt war. Die Entscheidung für die Fuß- und gegen die Handgesteninteraktion beruhte zum einen auf den persönlichen Vorlieben der Probanden, zum anderen darauf, wie anstrengend die Handgeste empfunden wurde. Das längere Halten der Hand innerhalb des Sichtbereichs der HoloLens war für die Hälfte der Probanden in der Schulterpartie anstrengend und ermüdend, wie [Abbildung 4.19](#) zeigt. Bei der Fußinteraktion wurde im Gegensatz zum Selektionsblock der Schwellwert von ≥ 3 nicht überschritten. In der gleichen [Abbildung](#) wurde bei der Sprachinteraktion zum ersten Mal der Nacken- und Halsbereich markiert. Bei genauerer Betrachtung konnte festgestellt werden, dass die zwei zusätzlichen Probanden, die diesen Bereich markierten, nach dem Studiendesign als letztes die Sprachinteraktion hatten. Dies bedeutet, dass die Probanden bereits eine längere Zeit die HoloLens trugen, wodurch sie zu diesem Zeitpunkt eine Anstrengung empfanden. Dementsprechend sollte ein angestrenzter oder ermüdeter Nacken- sowie Augenbereich keiner der drei Interaktionstechnik zuordnet werden. Diese Angabe sollte als ein Hinweis auf die Ergonomie der HoloLens aufgefasst werden.

Die Sprachinteraktion war für diese Art der Aufgabe am wenigsten geeignet. Sowohl quantitative als auch qualitative Messdaten zeigten, dass die Sprachinteraktion in diesen Durchläufen am schlechtesten ausfiel. Keiner der Probanden hatte die Sprachinteraktion favorisiert und über die Hälfte empfanden Fuß- und Handgesteninteraktion für besser geeignet ([Abbildung 4.20](#)). Bei der Manipulation eines kontinuierlichen Wertes wird eine schnelle Reaktion des Systems erwartet. Der Vorteil der etwas verzögerten Verarbeitung der Spracherkennung aus dem Selektionsblock wurde hier zum Nachteil. Durch die verzögerte Verarbeitung scrollten die Probanden im Schnitt bis zu 1,1 Bilder mehr als erwünscht. Zu erwähnen ist, dass einige Probanden hierbei bereits das Sprachkommando zwei Bilder vorher aussprachen, um im richtigen Moment anzuhalten.

In den Durchläufen mit Multitasking wurden die Interaktionstechniken unter einer ganz anderen Verteilung favorisiert. Das Ranking ([Abbildung 4.21](#)) zeigte, dass die Hälfte der Probanden, also 25% mehr, die Interaktion über die Füße bevorzugten. Die Sprachinteraktion wurde von ein Viertel der Probanden bevorzugt, während die Hälfte der Probanden Handgesteninteraktion in diesem Durchlauf am wenigsten bevorzugten. Die Nutzerbewertung innerhalb des NASA TLX zeigte, dass die geistige Anforderung bei der Handgesteninteraktion sehr stark anstieg. Wie auch in der Diskussion im Selektionsblock ist dies damit zu erklären, dass die beiden Interaktionsaufgaben den gleichen Kanal beanspruchten. Zwei Probanden beschrieben die Problematik, dass es sie stark angestrengt hat z.B. mit der linken Hand nach links zu ziehen um zu Scrollen, aber gleichzeitig mit der rechten Hand den Thumbstick nach rechts drücken mussten, um innerhalb der grünen Markierung

zu bleiben. Ein ähnliches Problem empfanden sie bei der Fußinteraktion, wenn auch nicht so stark ausgeprägt. Trotz des Rankings war die Handgesteninteraktion in Bezug auf die Task Completion Time weiterhin am schnellsten.

Die Auswertung der Primäraufgabe zeigte, dass im Gegensatz zu den Ergebnissen aus dem Selektionsblock, Sprachinteraktion einen sehr geringen Vorsprung gegenüber den anderen Interaktionstechniken hatte. Die bessere Leistung innerhalb der Primäraufgabe deutet darauf hin, dass die Aufgabe den Probanden einfacher fiel als in dem ersten Block.

Auffällig sind die Verteilungen, in welche Scrollrichtung mehr Overshoots entstanden sind (Abbildung 4.17). Bei der Handgesten- und Sprachinteraktion wurden beim rechts bzw. links Scrollen unter beiden Bedingungen mehr Overshoots verursacht. Bei der Handgesteninteraktion könnte es am sichtbaren Bereich liegen. Während der Ausführung der Drag Geste mit der linken Hand wird der sichtbare Bereich der HoloLens beim links Scrollen deutlich schneller verlassen als beim rechts Scrollen. Beim Verlassen des sichtbaren Bereichs wird die Geste abgebrochen und der Scrollvorgang gestoppt. Vermutlich mussten die Probanden hierbei öfters nachziehen, wodurch die Aufmerksamkeit mehr auf den Bildern lag. Da beim rechts Scrollen meist nicht nachgezogen werden musste, war hier vermutlich die Gefahr eines Overshoots höher gewesen. Der Grund bei der Sprachinteraktion ist hierbei nicht ersichtlich. Allerdings hatte ein Proband beim ersten Durchgang Probleme mit der Erkennung der Sprachkommandos. Erst als er seine Aussprache von [slaid] zu [flaid] änderte, wurde das Sprachkommando „slide“ einwandfrei erkannt.

Die gesamte Anzahl der Overshoots bei der Fußinteraktion liegen vermutlich daran, dass die Sensoren nicht immer die richtige Richtung bei der Fußinteraktion registrierten, oder die Sendung der Richtung gestoppt wurde. Dadurch mussten die Probanden die Fußgeste zum Teil mehrfach erneut ansetzen, wodurch sie eventuell, sobald das System dann doch den Scrollvorgang fortsetzte, nicht schnell genug waren, um das Scrollen zu stoppen.

Es muss erwähnt werden, dass zwar einige Probanden Erfahrungen mit Fußinteraktion hatten, sich dies jedoch bei fast allen auf einen Sensorfußboden bezogen hat. Die verwendete Fußinteraktion verhielt sich allerdings anders und erforderte mehr Gefühl von den Probanden im Gegensatz zu z.B. der Handgesteninteraktion, die zuerst intuitiver für die Probanden erscheint. Deshalb stellt sich die Frage, ob mit mehr Training bessere Ergebnisse erzielt werden könnten.

5 Abschluss

Innerhalb dieser Masterarbeit wurde zum einen untersucht, wie sich die Primäraufgabe und die Interaktionstechniken gegenseitig beeinflussen, zum anderen, welche Interaktionstechniken sich in diesem Zusammenhang am besten mit der HoloLens vereinen lassen. Das folgende Kapitel bildet den Abschluss dieser Arbeit. Zunächst werden die wesentlichen Bestandteile der Arbeit zusammengefasst. Im Anschluss werden die aus den Ergebnissen und Diskussionen gewonnenen Erkenntnisse gebündelt betrachtet und, abschließend, zukünftige Verbesserungen und Arbeitsschritte diskutiert.

5.1 Zusammenfassung

Zu Beginn wurden für die Untersuchungen innerhalb der Arbeit die nötigen Grundlagen erarbeitet. Dafür wurde auf den Bereich der **MCI** eingegangen. Es wurde speziell die Interaktion innerhalb des Operationssaals beleuchtet und gezeigt, dass Interaktionen mit Computersystemen während einer Operation eine besondere Herausforderung darstellen. Um Lösungen für die Problematik herauszuarbeiten, wurde eine Literaturrecherche bezüglich möglicher berührungsloser Interaktionstechnologien durchgeführt. Dabei wurde der Stand der Technik und die verwandten Forschungsergebnisse in Bezug auf die intraoperative Anwendung untersucht. Anschließend wurde der Bereich der AR beleuchtet und dessen Anwendungsgebiete vorgestellt. In diesem Zusammenhang wurden die Vorteile eines intraoperativen Einsatzes von AR beschrieben, speziell die der HoloLens. Schließlich wurden Studien präsentiert, die untersucht haben, wie sich zusätzliche Belastungen in Form von Zeitdruck und Multitasking auf die Interaktionen auswirken.

Basierend auf den Kenntnissen des Grundlagenkapitels, wurden die Interaktionstechnologien abgewogen und ausgewählt. Daraufhin wurden geeignete Interaktionsaufgaben konzeptioniert, die auf eine abstrakte Weise Aufgaben aus dem Operationssaal widerspiegeln. Diese wurden anschließend zu drei Prototypen zusammengebaut. Im Zuge dessen wurden die dazugehörigen Interaktionstechniken entworfen und in einer Studie untersucht.

Innerhalb der Studie wurden die Interaktionsaufgaben der Selektion und Manipulation einmal für sich untersucht und einmal in einem Multitasking-Kontext mit einer Primäraufgabe. Hierfür wurden die zwei Aufgaben in separaten Blocks durchgeführt. Nach der Erklärung des Ablaufs wurden zuerst die Ergebnisse der Blocks präsentiert und anschließend diskutiert.

Die Untersuchungen haben gezeigt, dass sich das Ausüben einer zusätzlichen Primäraufgabe deutlich auf die subjektive Bewertung der Interaktionstechniken auswirkt. Innerhalb der Interaktionsaufgaben wirkt sich der Multitasking-Kontext vor allem negativ auf die

Bewertung der Handgesteninteraktion aus. In den Durchläufen ohne Primäraufgabe fiel hingegen die Handgesteninteraktion sowohl in den quantitativen als auch in den qualitativen Daten zufriedenstellend aus. Die Sprachinteraktion hatte in der Selektionsaufgabe eine der besten Bewertungen, in der Manipulationsaufgabe jedoch eine der schlechtesten. Die Leistungen in der Primäraufgabe waren in beiden Fällen mit der Sprachinteraktion am besten, wobei innerhalb der Selektionsaufgabe der Abstand zu den beiden anderen Interaktionstechniken höher war, als bei der Manipulationsaufgabe.

5.2 Fazit und Ausblick

Die Ergebnisse der Studie weisen Ähnlichkeiten zu den Studien von van Galen & Huygevoort [92] sowie Moorthy et al. [68] auf. Beide Studien zeigen, dass sich zusätzliche Belastungen negativ auf die Ausführung der Primäraufgabe auswirken. Eine erhöhte Anzahl von Fehlern bei Moorthy et al. wurde neben dem gleichzeitigen Einsatz aller untersuchten Stressfaktoren vor allem bei der Multitasking Rechenaufgabe beobachtet. Die negative Auswirkung der Interaktionstechnik auf die Primäraufgabe konnte in dieser Arbeit ebenfalls beobachtet werden, aber auch die Auswirkung der Primäraufgabe auf die Interaktionstechniken konnte festgestellt werden. Bei beiden Aufgaben mit allen drei Interaktionstechniken war die Task Completion Time, die Overshootanzahl sowie die subjektive Arbeitsbelastung bei den Durchläufen mit zusätzlicher Primäraufgabe erhöht.

Das Ergebnis zeigt, dass Handgesteninteraktion eine intuitive und schnell reaktive Interaktionstechnik darstellt, sie jedoch bei längerer Verwendung anstrengend für den Benutzer wird. Des Weiteren wirkt sie sich am meisten auf die Primäraufgabe aus, indem sie eine erhöhte kognitive Belastung verursacht und somit mehr Konzentration und Aufmerksamkeit vom Benutzer fordert. Dies wirkt sich negativ auf die Primäraufgabe aus, da sie in manchen Fällen schlechter parallel ausgeführt werden kann. Der Vorteil der schnellen Reaktivität der Interaktionstechnik geht durch die erhöhte kognitive Belastung verloren.

Während die Fußinteraktion bei der Selektion ohne Multitasking keinen in den Daten messbaren Vorteil gegenüber den anderen Interaktionstechniken zeigt und sowohl quantitativ als auch qualitativ am schlechtesten ausfällt, ist im Multitasking-Kontext ein positiver Unterschied zu der Handgesteninteraktion zu erkennen. Mit der Fußinteraktion kann auf die besser Primäraufgabe geachtet werden und es entsteht eine niedrigere kognitive Belastung, wodurch die Selektionszeit ähnlich zu der der Handgesteninteraktion ist. Mit der Ausbesserung der Schwachstellen stellt sie, vor allem bei der Manipulation von kontinuierlichen Werten, eine gute Alternative zu der Handgesteninteraktion dar.

Bezüglich der kognitiven Belastung innerhalb eines Multitasking-Kontexts verursacht Sprachinteraktion die wenigsten Probleme, sodass sie sich sehr gut parallel zu der Primäraufgabe verwenden lässt. Bei der Interaktion ohne Primäraufgabe ist die langsamere Reaktivität durch die Verarbeitung innerhalb des Spracherkennungssystems bemerkbarer, was die Gesamtbearbeitungszeit verlängert.

Anhand der ermittelten Ergebnisse lassen sich folgende Handlungsempfehlungen bei dem Design von Anwendungen und der Integrierung von Interaktionstechniken in Verbindung mit der HoloLens ableiten:

- Die zu selektierenden Objekte müssen groß genug gewählt sein.
- Bei einer Selektion ohne Ausführung einer weiteren Aufgabe eignen sich Handgesten und Sprachkommandos sehr gut. Hierbei sollte beachtet werden, wie oft und lange die Interaktion ausgeführt werden soll, ob der Benutzer seine Hände frei hat und wie laut die Hintergrundgeräusche sind.
- Bei der Manipulation von kontinuierlichen Werten ohne zusätzliche Aufgabe sollte die Manipulation selbst möglichst nicht über Sprachkommandos durchgeführt werden. Hierfür eignen sich Handgesten und Fußgesten besser.
- Die Selektion mit gleichzeitiger Ausführung einer Aufgabe, welche die Hände erfordert, sollte am besten durch Sprachinteraktion oder Fußinteraktion durchgeführt werden.
- Bei der Manipulation mit einer gleichzeitigen Ausführung einer Aufgabe, welche die Hände erfordert, sollte auf Handgesten verzichtet werden. Hierbei eignet sich der Einsatz von Fußgesten am besten. Diese kann durch Sprachkommandos unterstützt werden, um in den jeweiligen Manipulationsmodus zu gelangen.

Im Folgenden werden die Einschränkungen innerhalb dieser Arbeit aufgezählt: Die Studienergebnisse wurden nicht statistisch ausgewertet und deswegen können keine Aussagen über die Signifikanz der Ergebnisse getätigt werden. Da es sich bei dieser Studie um eine Vorstudie handelt, um die technische Reife des Systems zu beurteilen, wurde eine geringere Stichprobe gewählt. Die nächstgrößere Stichprobe in diesem Studiendesign besteht aus 24 Probanden und soll in der Hauptstudie vorgenommen werden. Wie schon zuvor in der Diskussion im [Unterabschnitt 4.2.2](#) erwähnt, ist der Schwierigkeitsgrad der Selektionsaufgabe relativ hoch. Der dazugehörige Toe tap bedarf einer Verbesserung, da sich die erschwerte Interaktion deutlich auf die quantitativen und qualitativen Resultate auswirkte. Der beschriebene Effekt hat sich ebenfalls auf die Sprachinteraktion ausgewirkt. Bei der Manipulationsaufgabe zeigte auch die Fußinteraktion Verbesserungsbedarf bezüglich der Erkennung der Rotationsrichtung, da diese nicht immer einwandfrei erkannt wurde. Solange das Gebiet der Sprachinteraktion ein fortlaufender Forschungsbereich bleibt, ist die Spracherkennung stark von der technischen Umsetzung abhängig. Im Fall der HoloLens ist diese noch ausbaufähig, vgl. Diskussion im [Unterabschnitt 4.2.4](#). Die Erkennung der Sprachkommandos war ähnlich gut zu den Ergebnissen von Liu et al. [58], die bei 74,47% lag. Wobei die Erkennung in dieser Studie stark von der Aussprache der Versuchsperson abhängig war. Zusätzlich wurde die HoloLens Spracherkennung in keiner lauten Umgebung evaluiert.

In diesem Zusammenhang besteht zuallererst die weiterführende Arbeit darin, die Selektion über Fußgeste und Sprachkommandos zu verbessern. Die in der Diskussion des Selektionsblocks beschriebenen kleinen Bewegungen sollen sich möglichst nicht auf den Cursor auswirken. Hier könnte ein „sticky“ Cursor dieses Problem beheben. Ein solcher Cursor würde ab dem Anvisieren am Zielobjekt haften bleiben und nur durch eine stärkere Bewegung vom Zielobjekt gelöst werden. Des Weiteren muss die Heel rotation Geste

bezüglich der Gestenerkennung verbessert werden, sodass beim Scrollvorgang die gleichzeitige Rotation des Kopfes und der Füße in die gleiche Richtung vermieden werden kann. Der erste Ansatz hier ist, eventuelle fehlerhafte Überlegungen in der jetzigen Verarbeitung der Sensordaten zu finden oder die Daten über eine andere Kernidee auszuwerten.

Im Februar 2019 wurde die neue Generation der HoloLens (Microsoft HoloLens 2) vorgestellt. Diese verspricht eine komfortablere Trageform, wodurch längeres Arbeiten möglich sein soll. Des Weiteren soll der Sichtbereich deutlich größer und mit Eye-Tracking ausgestattet sein. Die Sprachbefehle sollen auch bei lauten Hintergrundgeräuschen funktionieren. Eine erneute Untersuchung zusammen mit Eye-Tracking oder in Kombination mit Head-Tracking wäre hierbei angemessen. Dabei sollte in der Zukunft eine Studie mit einer konkreten medizinischen Aufgabe, einer spezifischeren Probandengruppe und einer größerer Stichprobe in einem operativen Umfeld von großem Interesse sein. Hierzu sollten Ergebnisse aus Workflow-Analysen im Operationssaal verwendet werden, z.B. die von Franke et al. [34], um dadurch die verschiedenen multimodalen Interaktionstechniken besser integrieren zu können.

Abschließend lässt sich durch die Untersuchung schlussfolgern, dass die Wahl der Interaktionstechnik stark vom Aufgabenkontext abhängt. Des Weiteren müssen Interaktionstechniken, die später in einem Multitasking-Kontext eingesetzt werden sollen, unbedingt zusammen mit der jeweiligen Primäraufgabe untersucht werden.

Anhang

A. Fragebögen

A.1 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

A.2 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?

A.3 NASA TLX

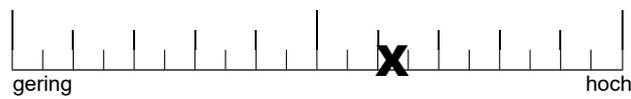
Code:



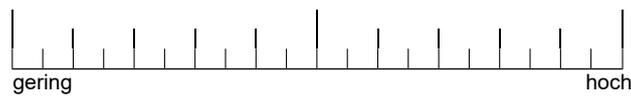
NASA – TLX

Geben Sie bitte an, wie hoch Ihre Beanspruchung in den folgenden Dimensionen war. Die Erklärung dafür finden Sie auf einem separaten Blatt. Markieren Sie dazu auf den folgenden Skalen, in welchem Maße Sie sich in den 6 Dimensionen von der Aufgabe beansprucht oder gefordert gesehen haben:

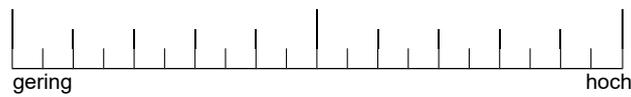
Beispiel:



Geistige Anforderungen



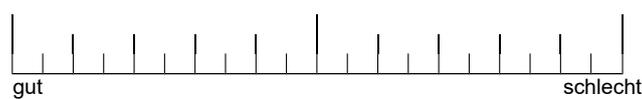
Körperliche Anforderungen



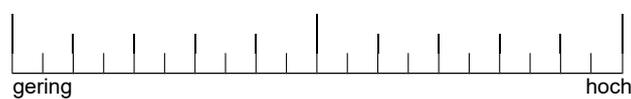
Zeitliche Anforderungen



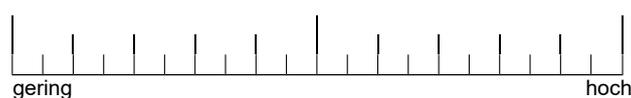
Ausführung der Aufgaben



Anstrengung

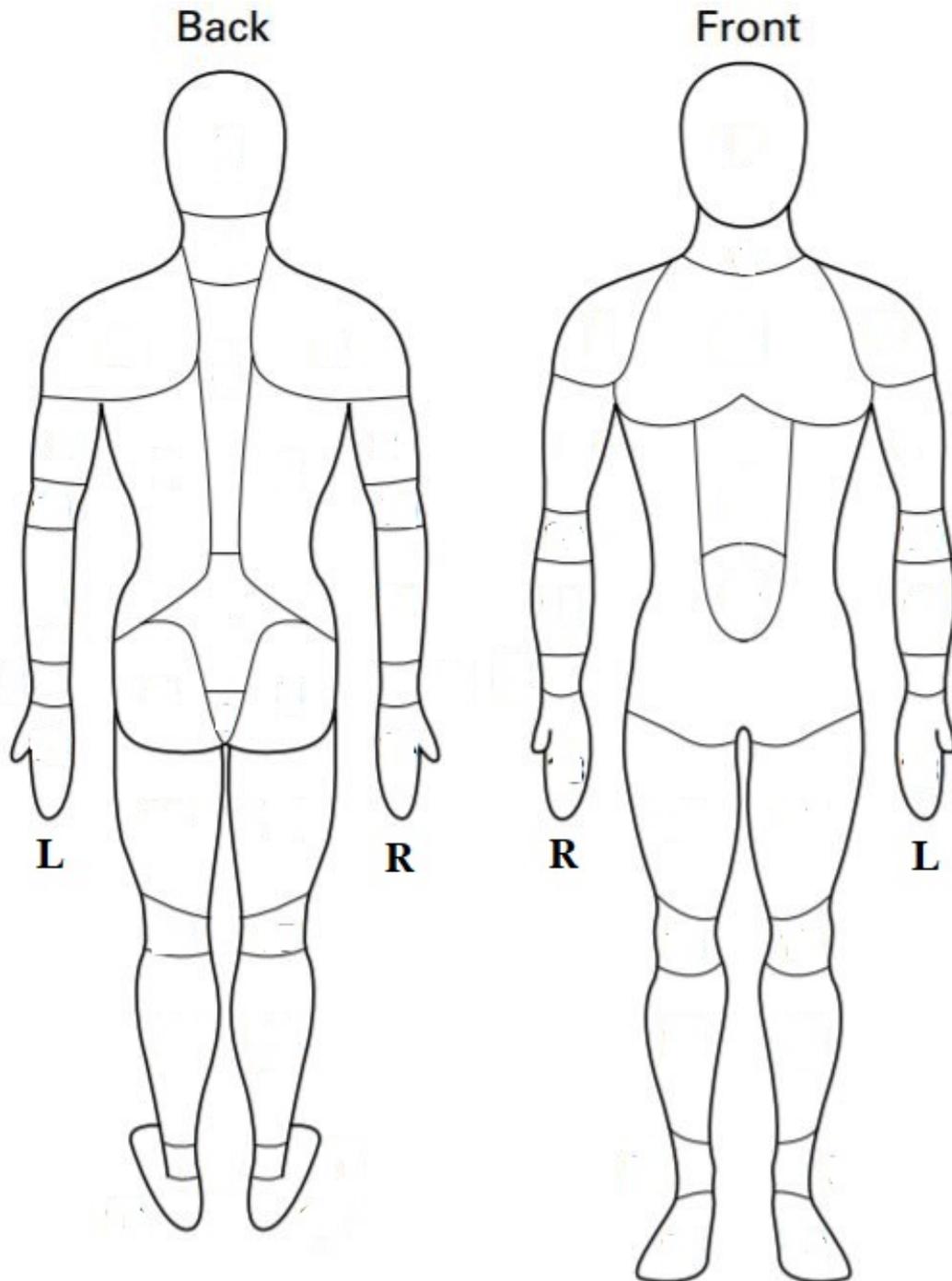


Frustration



NASA TLX

A.4 Menschliche Silhouette Fragebogen



B. Quantitative Ergebnisse

B.1 Selektionsblock

	Task Completion					
	ohne Primäraufgabe			mit Primäraufgabe		
	Fuß	Hand	Sprache	Fuß	Hand	Sprache
P1	97	126	79	108	225	136
P2	73	66	70	123	91	91
P3	97	144	89	150	150	113
P4	89	84	98	116	106	143
P5	115	65	89	102	119	110
P6	62	54	74	76	75	88
P7	89	71	81	103	142	94
P8	115	63	95	162	87	130
P9	72	58	107	104	96	168
P10	110	62	153	144	157	205
P11	128	80	130	137	122	193
P12	67	130	82	82	107	118
Durchschnitt	92,83	83,58	95,58	117,25	123,08	132,42

	Selektionszeit ohne Primäraufgabe		
	Fuß 1	Hand 1	Sprache 1
P1	2,68	3,97	2,27
P2	1,94	1,16	1,87
P3	3,06	4,77	2,75
P4	2,63	1,86	3,22
P5	3,61	1,37	2,76
P6	1,79	1,16	2,47
P7	2,79	1,87	2,52
P8	3,60	1,27	2,86
P9	1,93	1,40	3,59
P10	3,55	1,38	5,17
P11	4,03	1,70	4,33
P12	1,89	4,24	2,64
Durchschnitt	2,79	2,18	3,04

	Selektionszeit ohne Primäraufgabe		
	Fuß 2	Hand 2	Sprache 2
P1	1,97	5,47	3,16
P2	3,42	1,93	2,20
P3	4,43	4,00	3,09
P4	2,90	1,62	3,64
P5	2,27	2,57	3,10
P6	1,94	1,76	2,64
P7	3,07	4,08	2,61
P8	4,70	1,52	2,64
P9	2,01	2,25	5,55
P10	3,59	4,36	6,49
P11	3,28	2,87	5,95
P12	1,95	2,57	3,56
Durchschnitt	2,96	2,92	3,72

mit Primäraufgabe			
Fuß	Hand	Sprache	Baseline
96,26 %	96,52 %	99,29 %	100,00 %
97,19 %	98,71 %	96,61 %	99,45 %
92,15 %	93,69 %	96,90 %	99,20 %
99,39 %	94,98 %	97,06 %	100,00 %
91,93 %	90,55 %	94,50 %	99,90 %
95,32 %	99,07 %	100,00 %	100,00 %
74,41 %	64,12 %	87,93 %	98,42 %
98,85 %	98,84 %	98,87 %	98,32 %
93,80 %	96,14 %	98,66 %	100,00 %
99,82 %	98,55 %	99,08 %	96,59 %
98,15 %	98,05 %	99,37 %	100,00 %
98,60 %	98,18 %	99,77 %	99,45 %
94,66 %	93,95 %	97,34 %	99,28 %

	Overshoot					
	ohne Primäraufgabe			mit Primäraufgabe		
	Fuß	Hand	Sprache	Fuß	Hand	Sprache
P1	3,58	3,64	1,65	2,21	9,76	3,24
P2	2,13	1,00	2,33	5,67	1,84	2,83
P3	5,38	8,32	3,38	8,38	5,04	3,21
P4	5,67	2,36	2,16	4,67	2,04	2,96
P5	5,21	1,32	2,38	2,52	3,04	2,04
P6	2,79	1,12	1,96	3,21	2,36	2,50
P7	4,88	2,64	2,71	4,79	5,60	2,75
P8	2,96	1,20	2,16	6,58	1,12	2,73
P9	2,42	1,48	3,42	2,71	2,16	4,67
P10	4,88	1,72	6,46	5,63	5,64	11,00
P11	5,79	1,96	4,21	4,04	3,68	5,58
P12	2,08	4,20	2,50	2,30	3,68	4,50
Durchschnitt	3,98	2,58	2,94	4,39	3,83	4,00

B.2 Manipulationsblock

	Task Completion					
	ohne Primäraufgabe			mit Primäraufgabe		
	Fuß	Hand	Sprache	Fuß	Hand	Sprache
P1	72	71	199	89	113	110
P2	110	69	102	118	71	102
P3	82	64	91	103	65	117
P4	79	59	103	86	87	102
P5	66	65	89	67	62	93
P6	76	59	76	89	70	91
P7	67	63	100	120	60	112
P8	116	73	101	118	90	107
P9	95	57	122	79	64	106
P10	106	83	113	101	62	109
P11	87	95	98	92	79	114
P12	126	67	84	121	88	95
Durchschnitt	90,17	68,75	106,50	98,58	75,92	104,83

	Overshoot L					
	ohne Primäraufgabe			mit Primäraufgabe		
	Fuß	Hand	Sprache	Fuß	Hand	Sprache
P1	0,00	0,20	0,00	0,80	0,40	1,40
P2	0,40	0,40	3,40	0,80	0,00	1,00
P3	0,40	0,00	0,40	0,00	0,00	2,20
P4	0,00	0,00	0,20	0,20	0,00	2,60
P5	0,40	0,00	3,00	1,00	0,80	1,60
P6	0,40	0,00	0,20	2,00	0,20	1,00
P7	0,00	0,00	0,60	0,00	0,00	0,20
P8	0,00	0,00	0,40	0,00	0,00	1,00
P9	0,20	0,00	3,20	0,20	0,40	1,00
P10	0,40	0,00	2,00	0,80	0,20	1,60
P11	0,00	0,00	0,60	0,20	0,00	1,20
P12	0,00	0,00	0,40	0,60	0,00	1,00
Durchschnitt	0,18	0,05	1,20	0,55	0,17	1,32

Overshoot R						
ohne Primäraufgabe			mit Primäraufgabe			
	Fuß	Hand	Sprache	Fuß	Hand	Sprache
P1	1,00	0,20	0,00	0,80	4,00	0,00
P2	0,40	0,00	0,60	0,00	0,20	0,80
P3	0,00	0,00	0,20	0,00	0,20	0,20
P4	0,00	0,00	0,20	0,00	4,00	0,80
P5	0,60	2,40	0,80	0,40	0,20	1,00
P6	0,00	0,60	0,20	0,20	1,80	0,80
P7	0,20	0,60	2,40	0,00	2,00	0,00
P8	0,00	0,00	0,00	0,60	0,00	0,40
P9	0,20	0,00	3,00	0,20	0,00	0,80
P10	1,00	0,40	3,00	1,00	0,20	0,40
P11	0,20	0,00	1,60	1,20	1,00	1,20
P12	0,00	0,00	0,00	0,20	0,00	1,00
Durchschnitt	0,30	0,35	1,00	0,38	1,13	0,62

Primäraufgabe			
Fuß	Hand	Sprache	Baseline
97,64 %	96,70 %	99,32 %	100,00 %
96,55 %	100,00 %	99,49 %	99,45 %
95,02 %	100,00 %	96,27 %	99,45 %
98,46 %	92,29 %	96,34 %	100,00 %
98,65 %	98,84 %	99,30 %	99,90 %
94,60 %	94,56 %	96,41 %	100,00 %
95,42 %	100,00 %	99,17 %	98,42 %
98,01 %	98,06 %	99,37 %	98,32 %
89,79 %	95,18 %	94,86 %	100,00 %
95,79 %	97,18 %	98,94 %	99,04 %
96,45 %	93,67 %	96,27 %	96,59 %
97,64 %	98,85 %	97,45 %	100,00 %
96,17 %	97,11 %	97,77 %	99,26 %

Abkürzungsverzeichnis

ACM Association for Computing Machinery

AR Augmented Reality

CT Computertomographie

EMG Elektromyografie

HCI Human-computer-interaction

HMD Head-Mounted Display

HoloLens Microsoft HoloLens

IMU Inertial Measurement Unit

MCI Mensch-Computer-Interaktion

MR Mixed Reality

OP Operation

SIGCHI Special Interest Group on Computer-Human Interaction

UDP User Data Protocol

Abbildungsverzeichnis

2.1	Skizze der Zonen um den Operationstisch. Der rote Bereich zeigt die Sterilizzone. Entnommen aus [1].	10
2.2	Skizze eines Angiographie Operationssaals mit den beweglichen Geräten in blau. Entnommen aus [51].	11
2.3	Im Operationssaal verwendeter Fußschalter. Entnommen aus [13].	12
2.4	Eine Übersicht der in der Literatur verwendeten Fußgesten. Entnommen aus [95].	14
2.5	Links: Ein Inertialsensor (orange) mit einem Bluetooth Empfänger. Rechts: Angebrachte Sensoren zur Erfassung von Bewegungen. Entnommen aus [87] und [8].	17
2.6	Links: Eine Datenbrille mit Eye-Tracking. Rechts: Die Orientierungsachsen des Head-Trackings. Entnommen aus [9] und [10].	18
2.7	Das Realitäts-Virtualitäts-Kontinuum von Milgram & Koshino. Entnommen aus [7]	22
2.8	Komponenten der HoloLens. 1) Rechneinheit. 2) Umgebungserkennungskameras. 3) Infrarot Laser Projektor. 4) Tiefenkamera. 5) HD Videokamera. 6) Umgebungslichtsensor. 7) Holographischer Projektor. 8) See-Through Waveguide Linsen. 9) Mikrophon. 10) Eingebaute Lautsprecher. 11) Batterie. Entnommen aus [58].	24
2.9	Eingeblendete virtuelle Ansicht im Sichtbereich des Chirurgen. Entnommen aus [29].	25
2.10	HoloLens Bloom Geste. Entnommen aus [4].	26
3.1	Beispiel eines Hologramms. Links ist das komplette Hologramm zu sehen, rechts wurde der Kopf etwas nach unten gesenkt. Entnommen aus [5]. . . .	32
3.2	Modul 1. Roter Slider mit einem beweglichen grünen Bereich. Die Abbildung zeigt das Modul vor Beginn der Aufgabe.	36
3.3	Ablauf des Tipptests in verschiedene Richtungen nach EN ISO 9241-420:2011. Entnommen aus der deutschen Fassung von Oktober 2011.	37
3.4	Modul 2. Links: Die zu selektierenden Kugel in gelb. Rechts: Die zu selektierende Kugel wird über die Kopfposition anvisiert und färbt sich grün. . .	37
3.5	Modul 3. Die Bildergalerie befindet sich im nicht aktiven Scroll-Modus. . .	38
3.6	OP-Schuh mit angebrachtem Inertialsensor.	40
3.7	Ausschnitt aus der C# Windows Forms-Anwendung der MT Software Suite [6].	41
3.8	Selektionsprototyp. Links: Selektion, rechts: Selektion + Primäraufgabe . .	42
3.9	HoloLens Air tap Geste: zuerst werden die Finger in die vorbereitende Position gebracht (links), anschließend folgt der Tap, indem die Finger zusammengespreizt werden. Entnommen aus [14].	42
3.10	Der sichtbare Bereich der HoloLens in welchem Gesten erkannt werden. Entnommen aus [12].	43

3.11	Prototyp 3. Links: Manipulation, rechts: Manipulation + Primäraufgabe	45
3.12	Scrollen durch die Gesteninteraktion. Der grüne Pfeil zeigt, dass kontinuierlich nach rechts gescrollt wird.	45
4.1	Links: Schematische Skizze des Aufbaus mit den Positionen der Probanden beim Testdurchlauf und beim gemessenem Durchlauf. Rechts: Bereich des gemessenen Durchlaufs.	47
4.2	Der Verbindungsaufbau der verwendeten Geräte. Im Trainingsmodus wird die Verbindung zwischen Sensor Anwendung und HoloLens durch die gestrichelten Verbindungen ersetzt.	48
4.3	Das Ergebnis der Befragung bezüglich der Erfahrungen. Oben sind die Ergebnisse aus dem Selektionsblock und unten aus dem Manipulationsblock zu sehen.	49
4.4	Der schematische Ablauf der Studie. Es wurden zwei Blöcke durchgeführt mit jeweils drei Modalitäten. Blaumarkierte Kästen zeigen Zeitpunkte an, in denen Daten aufgezeichnet wurden.	49
4.5	Die Trainingsumgebung bestand aus zwei Reihen von jeweils fünf Kugeln mit den Durchmessern von 2,33 und 3,36 cm.	51
4.6	Versuchsperson beim Durchlauf mit Handgesteninteraktion im Multitasking-Kontext.	52
4.7	Die Task Completion Times von den Interaktionen über Fußgeste (F), Handgeste (H) und Sprache (S). Links das Ergebnis ohne und rechts mit Multitasking.	55
4.8	Die durchschnittliche Bewertung der Arbeitsbelastung von den Interaktionen über Fußgeste (F), Handgeste (H) und Sprache (S). Links die Durchgänge ohne und rechts mit Primäraufgabe.	56
4.9	Überblick über die durchschnittlichen Selektionszeiten der Probanden mit den Interaktionen über Fußgeste (F), Handgeste (H) und Sprache (S). Links die Zeiten ohne und rechts mit Multitasking.	57
4.10	Die durchschnittliche Overshoot Anzahl pro Selektion der Probanden über Fußgeste (F), Handgeste (H) und Sprache (S). Links die Anzahl ohne und rechts mit Multitasking.	58
4.11	Die Leistungen der einzelnen Probanden innerhalb der Primäraufgabe. Die Prozentzahlen richten sich nach dem Anteil des im grünen Bereich gehaltenen Markers. Die rote Linie stellt die Baseline dar.	58
4.12	Auswertung der körperlichen Belastung bei den Interaktionen über Fußgeste (F), Handgeste (H) und Sprache (S). Orange steht für angestrengte oder ermüdete Körperregionen, rot für schmerzende Regionen.	59
4.13	Die Platzierungen der Interaktionen über Fußgeste (F), Handgeste (H) und Sprache (S) ohne Primäraufgabe.	59
4.14	Die Platzierungen der Interaktionen über Fußgeste (F), Handgeste (H) und Sprache (S) mit Primäraufgabe	60
4.15	Die Task Completion Times von den Interaktionen über Fuß (F), Hand (H) und Sprache (S). Links das Ergebnis ohne und rechts mit Multitasking.	62
4.16	Die durchschnittliche Bewertung der Arbeitsbelastung der Interaktionen über Fuß (F), Hand (H) und Sprache (S). Links die Durchgänge ohne und rechts mit Primäraufgabe.	63

4.17	Die durchschnittliche Overshoot Anzahl pro Scrollvorgang der Probanden für jede Modalität. Links befindet sich die Anzahl ohne und rechts mit Multitasking. Die Overshoots wurden nach links und rechts Overshoots unterteilt.	64
4.18	Die Leistungen der einzelnen Probanden innerhalb der Primäraufgabe. Die Prozentzahlen richten sich nach dem Anteil des im grünen Bereich gehaltenem Marker. Die rote Linie stellt die Baseline dar.	65
4.19	Auswertung der körperlichen Belastung. Orange steht für angestrengte oder ermüdete Körperregionen. Rot für schmerzende Regionen.	65
4.20	Die Platzierungen der Interaktionstechniken ohne Primäraufgabe.	66
4.21	Die Platzierungen der Interaktionstechniken mit Primäraufgabe.	66

Literaturverzeichnis

- [1] Der OP - Bereich. <https://www.sana.de/unternehmen/blaubuch-verborgene-orte/operationssaal/scharfe-grenzen-strenge-regeln/>. [Online; Zugriff 30.Januar.2019].
- [2] Development - Mixed Reality. <https://docs.microsoft.com/en-us/windows/mixed-reality/development>. [Online; Zugriff 10.April.2019].
- [3] Hallway Usability Testing. <https://www.techopedia.com/definition/30678/hallway-usability-testing>. [Online; Zugriff 10.April.2019].
- [4] Hololens Gestures. https://www.synconfusion.com/ebooks/hololens_succinctly/gestures. [Online; Zugriff 12.April.2019].
- [5] The Hololens might have a small field of view. <https://hololens.reality.news/news/hololens-might-have-small-field-view-but-thats-actually-good-thing-0171349/>. [Online; Zugriff 12.April.2019].
- [6] Mt Software Suite MTw Awinda. <https://www.xsens.com/mt-software-suite-mtw-awinda/>. [Online; Zugriff 11.März.2019].
- [7] Was ist Mixed Reality? <https://www.mixed-reality-training.com/uebersicht-microsoft-mixed-reality/>. [Online; Zugriff 31.Januar.2019].
- [8] Xsens MTw Development Kit Lite. <https://autonomoustuff.com/product/xsens-mtw-development-kit-lite/>. [Online; Zugriff 31.Januar.2019].
- [9] Head-Tracking. <http://blog.dsky.co/tag/head-tracking/>, 2015. [Online; Zugriff 31.Januar.2019].
- [10] Tobii Pro Glasses 2 wearable eye tracker. <https://www.tobii.com/product-listing/tobii-pro-glasses-2/>, 2015. [Online; Zugriff 31.Januar.2019].
- [11] Xbox Controller Input for HoloLens. <https://assetstore.unity.com/packages/tools/input-management/xbox-controller-input-for-hololens-70068>, 2016. [Online; Zugriff 25.Februar.2019].
- [12] Hololens gesture frame. <https://support.microsoft.com/en-us/help/12644/hololens-use-gestures/>, 2017. [Online; Zugriff 25.Februar.2019].
- [13] Fußschalter mit User Interfaces für den vernetzen Operationssaal. <https://medizin-und-technik.industrie.de/allgemein/fusschalter-mit-user-interfaces-fuer-den-vernetzen-operationssaal/>, 2018. [Online; Zugriff 01.Februar.2019].
- [14] Gestures - Mixed Reality. <https://docs.microsoft.com/en-us/windows/mixed-reality/gestures>, 2018. [Online; Zugriff 16.Februar.2019].
- [15] MR Input 212 - Voice - Mixed Reality. <https://docs.microsoft.com/en-us/windows/mixed-reality/holograms-212>, 2018. [Online; Zugriff 31.Januar.2019].

- [16] H. Afkari, D. G. de Gómez Pérez, and R. Bednarik. Command selection in gaze-based see-through virtual image-guided environments. In *Proceedings of the 2018 AH*, page 8. ACM, 2018.
- [17] A. Alapetite. Impact of noise and other factors on speech recognition in anaesthesia. *Int. J. Med. Inform.*, 77(1):68–77, 2008.
- [18] D. Andersen, C. Lin, V. Popescu, E. R. Munoz, M. E. Cabrera, B. Mullis, B. Zarzaur, S. Marley, and J. Wachs. Augmented visual instruction for surgical practice and training. In *2018 IEEE Workshop on VAR4Good*, pages 1–5. IEEE, 2018.
- [19] S. Arora, N. Sevdalis, D. Nestel, T. Tierney, M. Woloshynowych, and R. Kneeboone. Managing intraoperative stress: What do surgeons want from a crisis training program? *Am. J. Surg*, 197(4):537–543, 2009.
- [20] W. Barfield. *Fundamentals of Wearable Computers and Augmented*, volume 2016. Apple Academic Press, 2016.
- [21] S. Bernhardt, S. A. Nicolau, L. Soler, and C. Doignon. The status of augmented reality in laparoscopic surgery as of 2016. *Med Image Anal*, 37:66–90, 2017.
- [22] A. Bigdelou, L. Schwarz, and N. Navab. An adaptive solution for intra-operative gesture-based human-machine interaction. In *Proceedings of the 2012 IUI*, pages 75–84. ACM, 2012.
- [23] P. Biswas and P. Langdon. Multimodal intelligent eye-gaze tracking system. *Int. J. Hum-Comput. Int.*, 31(4):277–294, 2015.
- [24] J. Blattgerste, P. Renner, and T. Pfeiffer. Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. In *Proceedings of the 2018 COGAIN*, page 1. ACM, 2018.
- [25] B. Böhm, N. Rötting, W. Schwenk, S. Grebe, and U. Mansmann. A prospective randomized trial on heart rate variability of the surgical team during laparoscopic and conventional sigmoid resection. *Arch Surg.*, 136(3):305–310, 2001.
- [26] R. A. Bolt. “Put-that-there”: *Voice and gesture at the graphics interface*, volume 14. ACM, 1980.
- [27] 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. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, 3(2):113–120, 2016.
- [28] C. Cometti, C. Paizis, A. Casteleira, G. Pons, and N. Babault. Effects of mixed reality head-mounted glasses during 90 minutes of mental and manual tasks on cognitive and physiological functions. *PeerJ*, 6:e5847, 2018.
- [29] G. Deib, A. Johnson, M. Unberath, K. Yu, S. Address, L. Qian, G. Osgood, N. Navab, F. Hui, and P. Gailloud. Image guided percutaneous spine procedures using an optical see-through head mounted display: proof of concept and rationale. *JNIS*, 10(12):1187–1191, 2018.

- [30] R. Dörner, W. Broll, P. Grimm, and B. Jung. *Virtual und Augmented Reality (VR/AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität*. Springer-Verlag, 2013. aufl. edition, 2014.
- [31] M. Dorr, M. Böhme, T. Martinetz, and E. Barth. Gaze beats mouse: A case study. *Proceedings of the 2007 COGAIN*, pages 16–19, 2007.
- [32] D. B. Douglas, C. A. Wilke, D. Gibson, E. F. Petricoin, and L. Liota. Virtual reality and augmented reality: Advances in surgery. *Biol*, 2(5):1–8, 2017.
- [33] B. Fida, F. Cutolo, G. di Franco, M. Ferrari, and V. Ferrari. Augmented reality in open surgery. *Updates in surgery*, pages 1–12, 2018.
- [34] S. Franke, M. Rockstroh, M. Hofer, and T. Neumuth. The intelligent OR: Design and validation of a context-aware surgical working environment. *J. Comput. Assist. Radiol. Surg.*, 13(8):1301–1308, 2018.
- [35] D. Gasques Rodrigues, A. Jain, S. R. Rick, L. Shangley, P. Suresh, and N. Weibel. Exploring mixed reality in specialized surgical environments. In *Proceedings of the 2017 CHI EA*, pages 2591–2598. ACM, 2017.
- [36] M. A. Grasso, D. S. Ebert, and T. W. Finin. The integrality of speech in multimodal interfaces. *TOCHI*, 5(4):303–325, 1998.
- [37] A. Grinshpoon, S. Sadri, G. J. Loeb, C. Elvezio, and S. K. Feiner. Hands-free interaction for augmented reality in vascular interventions. In *Proceedings of the 2019 IEEE VR*, pages 751–752. IEEE, 2018.
- [38] J. Hall, C. Ellis, and J. Hamdorf. Surgeons and cognitive processes. *Br. J. Surg.*, 90(1):10–16, 2003.
- [39] D. Han, X. Li, and T. Zhao. The application of augmented reality technology on museum exhibition—a museum display project in mawangdui han dynasty tombs. In *Proceedings of the 2017 VAMR*, pages 394–403. Springer, 2017.
- [40] B. Hatscher and C. Hansen. Hand, foot or voice: Alternative input modalities for touchless interaction in the medical domain. In *Proceedings of the 2018 ICMI*, pages 145–153. ACM, 2018.
- [41] B. Hatscher, M. Luz, and C. Hansen. Foot interaction concepts to support radiological interventions. *i-com*, 17(1):3–13, 2018.
- [42] 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 2017 ICMI*, pages 243–251. ACM, 2017.
- [43] J. Herron. Augmented reality in medical education and training. *J. Electron. Resour. Med. Libr.*, 13(2):51–55, 2016.
- [44] 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 the Eurographics Workshop on VCBM*, pages 177–185. Eurographics Association, 2015.

- [45] A. M. Hötker, M. B. Pitton, P. Mildenerger, and C. Düber. Speech and motion control for interventional radiology: Requirements and feasibility. *J. Comput. Assist. Radiol. Surg.*, 8(6):997–1002, 2013.
- [46] T. Issa and P. Isaias. Usability and Human Computer Interaction (HCI). In *Proceedings of the 2015 Sustainable Design*, pages 19–36. Springer, 2015.
- [47] M. G. Jacob and J. P. Wachs. Context-based hand gesture recognition for the operating room. *Pattern Recognit. Lett.*, 36:196–203, 2014.
- [48] R. J. Jacob. What you look at is what you get: Eye movement-based interaction techniques. In *Proceedings of the 1990 CHI*, pages 11–18. ACM, 1990.
- [49] S. Jalaliniya, D. Mardanbegi, and T. Pederson. Magic pointing for eyewear computers. In *Proceedings of the 2015 ISWC*, pages 155–158. ACM, 2015.
- [50] S. Jalaliniya, J. Smith, M. Sousa, L. Bütke, and T. Pederson. Touch-less interaction with medical images using hand & foot gestures. In *Proceedings of the 2013 Ubicomp*, pages 1265–1274. ACM, 2013.
- [51] 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 2011 CHI*, pages 3323–3332. ACM, 2011.
- [52] V. Kėpuska and G. Bohouta. Comparing speech recognition systems (microsoft api, google api and cmu sphinx). *Int. J. Eng. Res. Appl.*, 7:20–24, 2017.
- [53] M. Kersten-Oertel, I. Gerard, S. Drouin, K. Mok, D. Sirhan, D. S. Sinclair, and D. L. Collins. Augmented reality in neurovascular surgery: Feasibility and first uses in the operating room. *J. Comput. Assist. Radiol. Surg.*, 10(11):1823–1836, 2015.
- [54] T. Kim, H. Ju, and J. R. Cooperstock. Pressure or movement? In *Proceedings of the 2018 DIS*, pages 1219–1227. ACM, 2018.
- [55] C. Kirschbaum, O. T. Wolf, M. May, W. Wippich, and D. H. Hellhammer. Stress-and treatment-induced elevations of cortisol levels associated with impaired declarative memory in healthy adults. *Life Sci*, 58(17):1475–1483, 1996.
- [56] M. Liehn, H. Richter, and L. Kasakov. *OTA-Lehrbuch: Ausbildung zur Operationstechnischen Assistenz*. Springer-Verlag, 2014. Aufl. edition, 2014.
- [57] H.-C. K. Lin, M.-C. Chen, and C.-K. Chang. Assessing the effectiveness of learning solid geometry by using an augmented reality-assisted learning system. *Interactive Learning Environments*, 23(6):799–810, 2015.
- [58] Y. Liu, H. Dong, L. Zhang, and A. El Saddik. Technical evaluation of HoloLens for multimedia: A first look. *IEEE MultiMedia*, 25(4):8–18, 2018.
- [59] W. O. C. López, P. A. Navarro, and S. Crispin. Intraoperative clinical application of augmented reality in neurosurgery: A systematic review. *Clin. Neurol. Neurosurg.*, 177.
- [60] I. S. MacKenzie. *Human-computer interaction: An empirical research perspective*. Newnes, 2012.

- [61] N. Mansfield and J. Marshall. Symptoms of musculoskeletal disorders in stage rally drivers and co-drivers. *BJSM*, 35(5):314–320, 2001.
- [62] D. McNeill. *Hand and mind: What gestures reveal about thought*. University of Chicago press, 1992.
- [63] C. Menk, E. Jundt, and R. Koch. Visualisation techniques for using spatial augmented reality in the design process of a car. In *Proceedings of the 2011 Computer Graphics Forum*, volume 30, pages 2354–2366. Wiley Online Library, 2011.
- [64] 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 *Proceedings of the 2015 CHI EA*, pages 773–780. ACM, 2015.
- [65] J. W. Meulstee, J. Nijsink, R. Schreurs, L. M. Verhamme, T. Xi, H. H. Delye, W. A. Borstlap, and T. J. Maal. Toward holographic-guided surgery. *Surgical innovation*, 26(1):86–94, 2019.
- [66] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *Transactions on Information Systems*, 77(12):1321–1329, 1994.
- [67] C. M. M. Mojica, N. V. Navkar, N. V. Tsekos, D. Tsagkaris, A. Webb, T. Birbilis, and I. Seimenis. Holographic interface for three-dimensional visualization of MRI on Hololens: A prototype platform for MRI guided neurosurgeries. In *Proceedings of the 2017 BIBE*, pages 21–27. IEEE, 2017.
- [68] K. Moorthy, Y. Munz, A. Dosis, S. Bann, and A. Darzi. The effect of stress-inducing conditions on the performance of a laparoscopic task. *Surg. Endosc.*, 17(9):1481–1484, 2003.
- [69] S. Moosburner, C. Remde, P. Tang, M. Queisner, N. Haep, J. Pratschke, and I. M. Sauer. Real world usability analysis of two augmented reality headsets in visceral surgery. *Artificial organs*, 2018.
- [70] K. O’Hara, G. Gonzalez, G. Penney, A. Sellen, R. Corish, H. Mentis, A. Varnavas, A. Criminisi, M. Rouncefield, N. Dastur, et al. Interactional order and constructed ways of seeing with touchless imaging systems in surgery. *CSCW*, 23(3):299–337, 2014.
- [71] 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. *CACM*, 57(1):70–77, 2014.
- [72] S. Oviatt. Ten myths of multimodal interaction. *CACM*, 42(11):74–81, 1999.
- [73] S. Oviatt. Multimodal interfaces. In *The human-computer interaction handbook*, pages 286–304. L. Erlbaum Associates Inc., 2002.
- [74] S. Oviatt, A. DeAngeli, and K. Kuhn. Integration and synchronization of input modes during multimodal human-computer interaction. In *In Proceedings of 1997 Referring Phenomena in a Multimedia Context and their Computational Treatment*, pages 1–13. ACL, 1997.
- [75] V. Paelke. Augmented reality in the smart factory: Supporting workers in an industry 4.0. environment. In *Proceedings of the 2014 ETFA*, pages 1–4. IEEE, 2014.

- [76] A. Perrakis, W. Hohenberger, and T. Horbach. Integrated operation systems and voice recognition in minimally invasive surgery: Comparison of two systems. *Surg Endosc.*, 27(2):575–579, 2013.
- [77] T. Phienthrakul. Armband gesture recognition on electromyography signal for virtual control. In *Proceedings of the 2018 KST*, pages 149–153. IEEE, 2018.
- [78] B. Preim and C. P. Botha. *Visual computing for medicine: Theory, algorithms, and applications*. Newnes, 2013.
- [79] B. Preim and R. Dachsel. *Interaktive Systeme: Band 2: User Interface Engineering, 3D-Interaktion, Natural User Interfaces*. Springer-Verlag, 2. Aufl. edition, 2015.
- [80] L. Qian, M. Unberath, K. Yu, B. Fuerst, A. Johnson, N. Navab, and G. Osgood. Towards virtual monitors for image guided interventions - real-time streaming to optical see-through head-mounted displays, 2017.
- [81] Y. Y. Qian and R. J. Teather. The eyes don’t have it: An empirical comparison of head-based and eye-based selection in virtual reality. In *Proceedings of the 2017 SUI*, pages 91–98. ACM, 2017.
- [82] V. Rajanna and T. Hammond. Gawschi: gaze-augmented, wearable-supplemented computer-human interaction. In *Proceedings of the 2016 ETRA*, pages 233–236. ACM, 2016.
- [83] V. Rajanna and T. Hammond. A fitts’ law evaluation of gaze input on large displays compared to touch and mouse inputs. In *Proceedings of the 2018 COGAIN*, page 8. ACM, 2018.
- [84] L. M. Reeves, J. Lai, J. A. Larson, S. Oviatt, T. Balaji, S. Buisine, P. Collings, P. Cohen, B. Kraal, J.-C. Martin, et al. Guidelines for multimodal user interface design. *CACM*, 47(1):57–59, 2004.
- [85] N. Rossol, I. Cheng, R. Shen, and A. Basu. Touchfree medical interfaces. In *Proceedings of the 2014 EMBC*, pages 6597–6600. IEEE, 2014.
- [86] I. M. Sauer, M. Queisner, P. Tang, S. Moosburner, O. Hoepfner, R. Horner, R. Lohmann, and J. Pratschke. Mixed reality in visceral surgery: development of a suitable workflow and evaluation of intraoperative use-cases. *Ann. Surg.*, 266(5):706–712, 2017.
- [87] M. Schepers, M. Giuberti, and G. Bellusci. Xsens mvn: Consistent tracking of human motion using inertial sensing. *Technical Report*, 2018.
- [88] S. L. Shishkin, B. M. Velichkovsky, E. V. Melnichuk, I. A. Dubynin, D. G. Zhao, and A. V. Isachenko. The pursuing gaze beats mouse in non-pop-out target selection. In *SMC*, pages 3518–3523. IEEE, 2018.
- [89] A. Solovjova, D. Labsch, B. Hatscher, M. Fritzsche, and C. Hansen. Plantar pressure-based gestures for medical image manipulation. In R. Dachsel and G. Weber, editors, *Mensch und Computer 2018 - Tagungsband*. Gesellschaft für Informatik e.V., 2018.

- [90] J. Sutherland, J. Belec, A. Sheikh, L. Chepelev, W. Althobaity, B. J. Chow, D. Mitsouras, A. Christensen, F. J. Rybicki, and D. J. La Russa. Applying modern virtual and augmented reality technologies to medical images and models. *JDI*, 32(1):1–16, 2018.
- [91] O. M. Tepper, H. L. Rudy, A. Lefkowitz, K. A. Weimer, S. M. Marks, C. S. Stern, and E. S. Garfein. Mixed reality with hololens: Where virtual reality meets augmented reality in the operating room. *Plast. Reconstr. Surg.*, 140(5):1066–1070, 2017.
- [92] G. P. van Galen and M. van Huygevoort. Error, stress and the role of neuromotor noise in space oriented behaviour. *Biol. Psychol.*, 51(2-3):151–171, 2000.
- [93] 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. *J. Healthc. Eng.*, 2017:1–9.
- [94] 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 *Proceedings of the 2015 INTERACT*, pages 384–401. Springer, 2015.
- [95] E. Velloso, D. Schmidt, J. Alexander, H. Gellersen, and A. Bulling. The feet in human–computer interaction: A survey of foot-based interaction. *CSUR*, 48(2):21, 2015.
- [96] S. Wang, M. Parsons, J. Stone-McLean, P. Rogers, S. Boyd, K. Hoover, O. Meruvia-Pastor, M. Gong, and A. Smith. Augmented reality as a telemedicine platform for remote procedural training. *Sensors*, 17(10):2294, 2017.
- [97] C. M. Wetzel, R. L. Kneebone, M. Woloshynowych, D. Nestel, K. Moorthy, J. Kidd, and A. Darzi. The effects of stress on surgical performance. *Am. J. Surg.*, 191(1):5–10, 2006.
- [98] A. Wheelock, A. Suliman, R. Wharton, E. Babu, L. Hull, C. Vincent, N. Sevdalis, and S. Arora. The impact of operating room distractions on stress, workload, and teamwork. *Ann. Surg.*, 261(6):1079–1084, 2015.
- [99] T. S. Young, R. J. Teather, and I. S. MacKenzie. An arm-mounted inertial controller for 6DOF input: Design and evaluation. In *Proceedings of the 2017 3DUI*, pages 26–35. IEEE, 2017.
- [100] 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 *Proceedings of the 2018 CHI EA*. ACM, 2018.
- [101] T. O. Zander, M. Gaertner, C. Kothe, and R. Vilimek. Combining eye gaze input with a brain–computer interface for touchless human–computer interaction. *Int. J. Hum-Comput. Int.*, 27(1):38–51, 2010.